



Universitat Autònoma de Barcelona

ADVERTIMENT. L'accés als continguts d'aquesta tesi queda condicionat a l'acceptació de les condicions d'ús establertes per la següent llicència Creative Commons:  http://cat.creativecommons.org/?page_id=184

ADVERTENCIA. El acceso a los contenidos de esta tesis queda condicionado a la aceptación de las condiciones de uso establecidas por la siguiente licencia Creative Commons:  <http://es.creativecommons.org/blog/licencias/>

WARNING. The access to the contents of this doctoral thesis it is limited to the acceptance of the use conditions set by the following Creative Commons license:  <https://creativecommons.org/licenses/?lang=en>

Universitat Autònoma de Barcelona

Non-classical correlations in quantum mechanics and beyond

by

Ludovico Lami

under supervision of

Prof. Andreas Johannes Winter

A thesis submitted in partial fulfillment for the
degree of Doctor of Philosophy

in

Unitat de Física Teòrica: Informació i Fenòmens Quàntics
Departament de Física
Facultat de Ciències

Bellaterra, September, 2017

“Noi ci allegrammo, e tosto tornò in pianto,
ché della nova terra un turbo nacque
e percosse del legno il primo canto.

Tre volte il fé girar con tutte l'acque;
alla quarta levar la poppa in suso
e la prora ire in giù, com' altrui piacque,
infin che 'l mar fu sovra noi richiuso”.

Ulisse

La Divina Commedia, Dante Alighieri

Abstract

This thesis is concerned with a seemingly naive question: what happens when you separate two physical systems that were previously together? One of the greatest discovery of the last century is that systems that obey quantum mechanical instead of classical laws remain inextricably linked even after they are physically separated, a phenomenon known as *entanglement*. This leads immediately to another, deep question: is entanglement an exclusive feature of quantum systems, or is it common to all non-classical theories? And if this is the case, how strong is quantum mechanical entanglement as compared to that exhibited by other theories?

The first part of the thesis deals with these questions by considering quantum theory as part of a wider landscape of physical theories, collectively called *general probabilistic theories* (GPTs). Chapter 1 reviews the compelling motivations behind the GPT formalism, preparing the ground for Chapter 2, where we translate the above questions into precise conjectures, and present our progress toward a full solution. In Chapter 3 we consider entanglement at the level of measurements instead of states, which leads us to the investigation of one of its main implications, *data hiding*. In this context, we determine the maximal data hiding strength that a quantum mechanical system can exhibit, and also the maximum value among all GPTs, finding that the former scales as the square root of the latter.

In the second part of this manuscript we explore some problems connected with quantum entanglement. In Chapter 4 we discuss its resistance to white noise, as modelled by channels acting either locally or globally. Due to the limited number of parameters on which these channels depend, we are able to answer all the basic questions concerning various entanglement transformation properties. The following Chapter 5 presents our view on the topic of Gaussian entanglement, with particular emphasis on the role of the celebrated ‘positive partial transposition criterion’ in this context. Extensively employing matrix analysis tools such as Schur complements and matrix means, we present unified proofs of classic results, further extending them and closing some open problems in the field along the way.

The third part of this thesis concerns more general forms of non-classical correlations in bipartite continuous variable systems. In Chapter 6 we look into Gaussian steering and problems related to its quantification, moreover devising a general scheme that allows to consistently classify correlations of bipartite Gaussian states into ‘classical’ and ‘quantum’ ones. Finally, Chapter 7 explores some problems connected with a ‘strong subadditivity’ matrix inequality that plays a crucial role in our analysis of correlations in bipartite Gaussian states. Among other things, the theory we develop allows us to conclude that a Rényi-2 Gaussian version of the elusive *squashed entanglement* coincides with the corresponding entanglement of formation when evaluated on Gaussian states.

Resumen

Esta tesis versa sobre una cuestión aparentemente naïf: ¿qué ocurre cuando se separan dos sistemas físicos que estaban juntos previamente? Uno de los mayores descubrimientos del siglo pasado es que los sistemas que obedecen leyes mecano-cuánticas, en lugar de clásicas, permanecen ligados inextricablemente incluso tras haber sido separados físicamente, un fenómeno conocido como *entrelazamiento*. Aquí nos preguntamos algo más profundo si cabe: ¿es el entrelazamiento una característica exclusiva de los sistemas cuánticos o es común a todas las teorías no-clásicas? Y, si es este el caso, ¿cuán fuerte es el entrelazamiento mecano-cuántico comparado con aquel exhibido por otras teorías?

La primera parte de esta tesis trata estas cuestiones considerando la teoría cuántica como parte de un conjunto más amplio de teorías físicas, colectivamente llamadas *teorías probabilísticas generales* (TPG). En el Capítulo 1 revisamos la sólida motivación que subyace al formalismo TPG, preparando el terreno para el Capítulo 2, donde traducimos las anteriores cuestiones a conjeturas precisas, y donde presentamos nuestro progreso hacia una solución completa. En el Capítulo 3 consideramos el entrelazamiento a nivel de medidas en vez de estados, lo cual conduce a la investigación de una de sus implicaciones principales, la *ocultación de información*. En este contexto, determinamos el máximo poder de ocultación de información que puede exhibir un sistema mecano-cuántico, así como el mayor valor entre todas las TPG, hallando que el primero crece como la raíz cuadrada del segundo.

En la segunda parte de este manuscrito exploramos algunos de los problemas relacionados con el entrelazamiento cuántico. En el Capítulo 4 discutimos su resistencia al ruido blanco modelado por canales que actúan bien local o bien globalmente. Debido al número limitado de parámetros de los que dependen estos canales, somos capaces de responder todas las preguntas básicas que conciernen a diversas propiedades de la transformación del entrelazamiento. En el siguiente Capítulo 5 presentamos nuestra perspectiva sobre el tema del entrelazamiento gaussiano, con un énfasis particular sobre el papel del célebre “criterio de la transposición parcial positiva” en este contexto. Empleando extensivamente herramientas del análisis matricial como los complementos de Schur y las medias matriciales, presentamos pruebas unificadas de resultados clásicos, extendiéndolos y cerrando algunos de los problemas abiertos en el campo.

La tercera parte de esta tesis se ocupa de formas más generales de correlaciones no-clásicas en sistemas bipartitos de variable continua. En el Capítulo 6 estudiamos el “steering” gaussiano y problemas relacionados con su cuantificación, y diseñamos un esquema general que permite clasificar consistentemente correlaciones de estados gaussianos bipartitos en “clásicas” y “cuánticas”. Finalmente, en el Capítulo 7 exploramos algunos problemas vinculados a una desigualdad matricial de “subaditividad fuerte” que desempeña un papel crucial en nuestro análisis de las correlaciones en los estados gaussianos bipartitos. Entre otras cosas, la teoría que desarrollamos nos permite concluir que una versión Rényi-2 gaussiana del escurridizo *squashed entanglement* coincide en estados gaussianos con el correspondiente entrelazamiento de formación.

Resum

Aquesta tesis parteix d'una pregunta aparentment ingènua: Què passa si es separen dos sistemes físics que estaven en contacte? Un dels descobriments més rellevants del segle passat és que els sistemes que obeeixen les lleis de la mecànica quàntica, en comptes de les lleis clàssiques, romanen intrínsecament connectats fins i tot quan estan separats físicament. Aquest fenomen és conegut com entrellaçament o *entanglement*. Aquí, ens preguntem quelcom més profund: pertany l'entrellaçament exclusivament als sistemes quàntics o és comú a totes les teories no clàssiques? I, donat el cas, com es pot comparar l'entrellaçament quàntic amb l'entrellaçament que pertany a d'altres teories?

La primera part de la tesis tracta amb aquestes qüestions considerant la teoria quàntica com a part d'un grup més ampli de teories físiques anomenat *general probabilistic theories* (GPTs). El Capítol 1 repassa les motivacions que hi ha darrera el formalisme GPT, contextualitzant el Capítol 2, on plantegem les preguntes mencionades en conjectures formals adjuntant-ne la nostre contribució cap a una solució completa. Al Capítol 3, considerem l'entrellaçament a nivell de mesures i no d'estats, la qual cosa ens porta cap a la investigació d'una de les seves principals implicacions, *data hiding*. En aquest marc, determinem la màxima eficiència de el *data hiding* que un sistema quàntic pot exhibir i també el màxim valor entre tots els GPTs, trobant que els primers escalen amb l'arrel quadrada dels darrers.

En la segona part d'aquest manuscrit estudiem alguns problemes relacionats amb l'entrellaçament quàntic. Al Capítol 4, discutim la seva resistència al soroll blanc, modelitzat amb canals que actuen tant local com globalment. Aquests canals depenen d'un nombre limitat de paràmetres, això fa que siguem capaços de respondre totes les preguntes bàsiques relacionades amb les propietats de transformació de l'entrellaçament. El Capítol 5 presenta la nostra visió sobre l'entrellaçament gaussià, amb especial focus en el rol del anomenat '*positive partial transposition criterion*' en aquest context. Extensament, fent servir tècniques d'anàlisi de matrius com ara *Schur complements* i *matrix means*, presentem demostracions de resultats clàssics generalitzant-los i resolent algun dels problemes oberts existents en la matèria.

La tercera part de la tesis es basa en formes més generals de correlacions no clàssiques en sistemes bipartits i de variable contínua. Al Capítol 6 investiguem el *Gaussian steering* i problemes relacionats en la seva quantificació, així com presentem un esquema general que permeti consistentment classificar correlacions de sistemes bipartits gaussians en 'clàssiques' i 'quàntiques'. Finalment, el Capítol 7 explora alguns dels problemes relacionats amb *strong subadditivity* en desigualtats de matrius que juga un paper clau en el nostre anàlisi de correlacions en estats gaussians bipartits. Entre d'altres coses, la teoria que desenvolupem ens serveix per concloure que una Rényi-2 versió gaussiana del difús *squashed entanglement* coincideix amb el corresponent entrellaçament de formació quan s'avalua en estats gaussians.

Acknowledgements

If in these three years in Barcelona I have learnt anything, then my advisor, Andreas Winter, deserves large part of the credit. He has been of example to me, both scientifically and personally, and these few lines will never do justice to his guide. I thank him for having taught me that our personal taste should not determine the science we do more than its significance for real-world situations, and that I should never give in to the temptation of judging any scientist on the basis of the number of papers published or talks given. I am indebted to him for all the opportunities he has given me, all the conferences he has enthusiastically sent me to, and in general for the unconditional freedom I enjoyed during my PhD.

Having Howard Barnum as a referee of this manuscript is an honour I can hardly express in words. Let me just say that the first paper in quantum information I read, studied and loved is his ‘Largest separable balls around the maximally mixed bipartite quantum state’ (in collaboration with L. Gurvits). I am grateful to Antonio Acín and John Calsamiglia for taking on the thankless task of refereeing this overly long manuscript. In particular, I thank Toni for driving me to Setcases safe and sound, meanwhile holding a conversation in flawless Italian. The countless discussions I had with John in front of a blackboard have taught me a great many things, so I thank him wholeheartedly.

I am beholden to Vittorio Giovannetti, my master’s thesis advisor, for educating me in quantum information and setting me on this path. Part of the present thesis was completed while I was in Pisa, and I thank him and his group for their kind hospitality. Let me take the chance to express my gratitude to the Scuola Normale Superiore, my alma mater. I owe to the professors and students I met there much of my love for physics and mathematics, so they deserve a big shout-out. The quantum optics course held there by Alessio Serafini calls for a special mention, as some of the work I will present here has its roots in those lectures.

I thank Gerardo Adesso for all the support he has given to me since the first time we have met. I am glad and honoured to start a new adventure with him in Nottingham.

I collaborated with many colleagues and friends over the last years. Apart from the already mentioned Andreas, Alessio and Gerardo, let me thank Carlos, Christoph, Fabien, Gláucia, Kaushik, Marco, Marcus, Mark, and Siddhartha. This thesis would not be the same without your contribution. I am also grateful

to Abel and María, who helped me with the translation of the abstract of this thesis into Catalan and Spanish.

Ultimately, the success of a PhD is largely decided by the people you meet. And in my case, I have been as lucky as a person can be. A big shout-out to all GIQers, past and present, for never letting me forget that we are first people and then scientists. Abel, Cécilia, Gianni, Krishna, Marcus, María, Milan, Moha and Sara deserve a special thought for the care they put into making me feel at home during my stay in Barcelona. Abel, as someone said, you are the most Catalan thing in the whole GIQ. Céci, how many cakes you baked are physically part of my body right now? Gianni, your psychological terrorism was decisive in making me defeat the dreadful Spanish bureaucratic machine. Krishna, good try, but there is no way you will look like a 'brown David' before a substantial amount of exercise. Marcus, I have learnt more on Italian cuisine from you than from anybody else in the world. María, we will manage to attend an opera together, I promise. Moha, I hope you will like being mentioned here, even though this thesis pretends to be written in British English. Milan, I am very sorry to tell you that my Hungarian vocabulary has not gone much beyond a laughable 'köszönöm szépen'. Last but not least, Sara, thanks for listening to all my problems and complaints and for hugging me whenever I needed it. I specially acknowledge support from emblematic institutions such as the late lamented LIQUID, the still burgeoning MAFIA, and the invaluable Graciosos group.

Among the many people I have met in Barcelona outside academia, let me thank Valeria and Alfredo for the countless vermouths we downed together. Valeria, you have been a precious help in my vain attempt to extend my Sicilian vocabulary.

Going back to your hometown is always refreshing, but much more when an adorable gang of friends is ready to welcome you. Let me thank Afro, France, Luca, Mola and Ska for their unshakable trust in my capability of completing the PhD. The often legendary discussions I had with Mola and Afro on a variety of absurd topics have been a constant source of inspiration. The friendship with the much more charming Arianna, Camilla and Diletta has been invaluable to me during these years abroad. I thank you wholeheartedly for all the strolls we took together.

I am grateful to Alvise, Enza and Franceschina for never leaving me alone when I felt disheartened. Alvise, trying to keep up with your moral integrity has been all but easy, but I tried my best. Finally, I hope to find the right words to thank Francesca and Benedetta for being so close to me while I was in Barcelona, despite the long distance. Francesca, I enjoyed all the debates we set up on classical music, even if I still firmly believe Mozart's music to be better than anything else. Benedetta, if I had to properly express my gratitude for all you have done for me during these years, our afternoon tea would last forever.

Without the everlasting love and care of my family this thesis would have never been completed, so I thank my parents and my brother Guglielmo from the bottom of my heart. Gug, I hope this will encourage you to pursue your

dreams in spite of all difficulties. My grandparents nonno Manrico and Grazia believed in me since when I was little and have been a constant source of inspiration for me.

My final thought goes to Margherita, who made this whole endeavour possible through her sincere love, which I cherish beyond words.

Contents

Contents	11
Introduction	1
I Non-classicality in general probabilistic theories	7
1 Foundations of General Probabilistic Theories	9
1.1 Introduction	9
1.1.1 Brief history of General Probabilistic Theories	10
1.1.2 Preparing and measuring procedures	11
1.1.3 The path to Ludwig’s embedding theorem	13
1.2 Ordered vector spaces	15
1.3 Topological vector spaces	17
1.4 Banach spaces	20
1.4.1 Some immediate consequences of completeness	21
1.4.2 Hahn-Banach separation theorem	21
1.4.3 Some classic results in Banach space theory	24
1.5 Ordered Banach spaces	27
1.5.1 Duality of cones	28
1.5.2 Order dual vs Banach dual	32
1.6 Order unit and base norm spaces	34
1.6.1 Order unit spaces	34
1.6.2 Base norm spaces	37
1.6.3 Duality of order unit and base norm spaces	42
1.7 Ludwig’s embedding theorem	46
1.7.1 Constructing the basic vector spaces	47
1.7.2 Introducing a Banach space structure	48
1.7.3 Inducing new norms on V	50
1.7.4 Constructing new Banach spaces	51
1.7.5 Identifying symmetric subspaces	52
1.7.6 Constructing the smallest symmetric subspace	53
1.7.7 Order structure: meeting requirement (c)	54
1.7.8 Weak*-denseness: meeting requirement (d)	55
1.7.9 Verifying uniqueness up to isomorphisms	58

2	Composite systems in GPTs	61
2.1	Introduction	61
2.1.1	Some fundamental questions	62
2.1.2	Original contributions	64
2.2	Basic theory	66
2.2.1	Some preliminaries on Banach spaces and tensor norms	66
2.2.2	Single systems	68
2.2.3	Bipartite systems	74
2.2.4	Classes of measurements in bipartite GPTs	77
2.3	Examples	80
2.3.1	Classical probability theory	80
2.3.2	Quantum mechanics	81
2.3.3	Centrally symmetric models	83
2.3.4	Generalised non-signalling theories	86
2.4	Separability problem	88
2.4.1	Woronowicz and complete extendibility criteria	89
2.4.2	Separability under symmetry	91
2.4.3	Tensor rank and separability	94
2.4.4	Separability problem for centrally symmetric models	99
2.5	Minimal and maximal tensor product	101
2.5.1	Generalities	101
2.5.2	Lattices and vector lattices	103
2.5.3	Decomposition and interpolation properties	104
2.5.4	Decomposition lemma and converses thereof	106
2.5.5	Namioka-Phelps theorem	110
2.6	Beyond Namioka-Phelps	114
2.6.1	Nonlocality in maximal tensor products	114
2.6.2	Universal entangleability of centrally symmetric models	119
2.7	Conclusions and outlook	127
3	Ultimate data hiding in quantum mechanics and beyond	131
3.1	Introduction	131
3.1.1	Data hiding	132
3.1.2	Our contributions	133
3.2	Data hiding in general probabilistic theories	134
3.2.1	State discrimination	134
3.2.2	Data hiding and statement of the problem	137
3.3	Data hiding in quantum mechanics	140
3.3.1	Quantum data hiding ratios	141
3.3.2	Data hiding in W -theory	145
3.3.3	Further implications of the teleportation argument	148
3.4	Ultimate bound on data hiding effectiveness	151
3.4.1	Data hiding in spherical models	151
3.4.2	A result on tensor norms	155
3.4.3	Ultimate data hiding ratios	156

3.5	Data hiding in special classes of GPTs	160
3.5.1	Centrally symmetric models	161
3.5.2	Werner construction for symmetric models	168
3.6	Conclusions	177
II Some aspects of quantum entanglement		179
4	Bipartite depolarising maps	181
4.1	Introduction	181
4.1.1	The problem	181
4.1.2	Our contributions	183
4.2	Preliminaries	184
4.2.1	Classes of maps in quantum information	184
4.2.2	Depolarising maps	187
4.2.3	Bipartite depolarising maps	188
4.2.4	Entanglement properties of the Hadamard product	189
4.3	Positivity region	191
4.3.1	Direct approach	192
4.3.2	An alternative proof of Theorem 4.4	194
4.4	Complete Positivity region	197
4.5	Entanglement-breaking region	199
4.5.1	Computation of the entanglement-breaking region	199
4.5.2	A more direct verification of Theorem 4.9	204
4.6	Entanglement-annihilating region	206
4.6.1	Main result	207
4.6.2	Applications	210
4.7	Conclusion	212
5	Gaussian entanglement revisited	215
5.1	Introduction	215
5.1.1	Gaussian separability problem	216
5.1.2	Our contributions	217
5.2	The toolbox: Schur complements, matrix means and Gaussian states	219
5.2.1	Schur complements	219
5.2.2	Matrix means	221
5.2.3	Gaussian formalism	224
5.3	Simplified separability criterion for Gaussian states	229
5.4	Sufficiency of the PPT condition – Revisited	231
5.4.1	PPT equals separability for 1 vs n -mode Gaussian states – Revisited	231
5.4.2	PPT equals separability for multimode isotropic Gaussian states – Revisited	233
5.5	Novel results on Gaussian entanglement	237

5.5.1	Gaussian states that are invariant under partial transpose are separable	237
5.5.2	PPT equals separability for multimode mono-symmetric Gaussian states	238
5.5.3	Entangling Gaussian states via passive optical operations	241
5.6	Summary and outlook	244

III Quantum correlations in continuous variable systems 247

6	Schur complement inequalities and monogamy of quantum correlations	249
6.1	Introduction	249
6.1.1	Quantum correlations in Gaussian states	250
6.1.2	Our contributions	251
6.2	Quantum steering	253
6.2.1	Generalities	253
6.2.2	Resource theory of steering	255
6.2.3	Gaussian steerability measure	258
6.3	Schur complement inequalities and Gaussian steerability	259
6.3.1	Schur complement inequalities	259
6.3.2	Mathematical detour: properties of the functions g_{\pm}	263
6.3.3	Monotonicity of Gaussian steerability	265
6.3.4	Monogamy of Gaussian steerability	272
6.4	Correlations hierarchy in Gaussian states	274
6.4.1	Correlations in bipartite quantum systems	274
6.4.2	Rényi-2 Gaussian entanglement of formation	275
6.4.3	Gaussian entanglement and correlations hierarchy	277
6.4.4	Applications	279
6.5	Conclusions	281
7	From log-det inequalities to Gaussian entanglement via recoverability theory	283
7.1	Introduction	283
7.1.1	Building bridges between matrix analysis, probability theory and Gaussian quantum information	284
7.1.2	Our contributions	285
7.2	Mathematical preliminaries	286
7.2.1	Information theory with Gaussian random variables	286
7.2.2	Log-det entropy	288
7.3	SSA saturation and exact recovery	290
7.4	Improvements of log-det SSA	294
7.4.1	Gaussian recoverability	294
7.4.2	A lower bound on $I_M(A : B C)_V$	298
7.5	SSA for quantum covariance matrices and Rényi-2 Gaussian squashed entanglement	301

<i>CONTENTS</i>	15
7.5.1 More about Gaussian states	301
7.5.2 SSA for quantum covariance matrices	303
7.5.3 Rényi-2 Gaussian squashed entanglement	304
7.6 Conclusions	309
A Computation rules for polars	311
B Complements on the proof of Ludwig’s embedding theorem	315
C Tensor product rule for GPTs	321
D Upper bound on the inseparability ratio $S(2)$	325
E Quantum data hiding with Werner states	329
F Comparing $\ \cdot\ _W$ and $\ \cdot\ _2$	331
G Additional remarks on some results of Chapter 6	333
G.1 Gaussian steerability is not monogamous with respect to the steering party	333
G.2 Rényi extension of Theorem 6.14	334
Bibliography	337

Introduction

When I started the PhD in 2014, the discovery of nonlocality as an intrinsic feature of quantum mechanical experimental predictions was exactly 50 years old. Half a century, half an ideal human life [1], and we are still adding footnotes to the solution of the Einstein-Podolsky-Rosen paradox [2] given by John Stewart Bell in 1964 [3]. And for how obscure and counterintuitive the laws of Nature may look, Bell's reasoning continues to show us the right way to grasp their deepest meanings.

Before we elaborate more on the contents of this thesis and on the way they reflect our standpoint on the topic, let us go back a bit and start with a brief historical note. The famous 1935 paper by Einstein, Podolsky and Rosen (EPR) [2] triggered an animated debate in the burgeoning community of quantum physicists. Bohr [4] tried to answer some of the questions raised there, however not fully appreciating their foundational status. Schrödinger [5] was struck by some of the implications of the EPR paper, but found it hard to believe that these corresponded to real physics. It took almost 30 years to focus the problem, but finally, in 1964, John Stewart Bell [3] published his stunning answer. This answer, nowadays known under the name of *Bell's theorem*, is one of the most profound discoveries in the history of science, and reveals us a fundamental feature of Nature that had eluded our previous investigations: *if the experimental predictions of quantum mechanics are correct, no local hidden variable model can account for the behaviour of bipartite systems*. The point we want to stress here is that this is not a statement about the *theory* of quantum mechanics, but rather concerns only its *predictions*, which have proven to be correct in all situations where we have tested them, and in particular in the setting described by EPR [6]. In other words, Bell's theorem forces us to rethink not our mathematical theory of reality, but rather our ontological view of reality itself.

This last observation motivates our point of view on the investigation of the phenomena connected with bipartite physical systems, and in particular of *entanglement*. A question that we find worth asking is as follows: *to what extent do said phenomena depend on the mathematical description of the systems under examination?* In order to formulate more precisely the question, we need a framework that categorises physical theories in a systematic way. Theories belonging to such framework – an example of which must be, naturally, quantum mechanics – will encompass only the most basic physical requirements and

allow only for the most basic predictions. Ideally, our formalism should not include anything that is not strictly necessary to make sense of the questions we want to ask. What we just described is the realm of *general probabilistic theories* (GPTs). As we shall see, even in its most basic form, this formalism is rich enough to allow for the investigation of a wide variety of phenomena.

The title of the thesis should be a bit less obscure by now. Our main subject of study is in fact the appearance, particularly in bipartite systems, of correlations that evade any ‘classical’ explanation. While quantum systems and quantum phenomena still play a prominent role in our work, particular emphasis is laid on the ‘post-quantum’ case, implying that all physical theories are treated on (a priori) equal footing, and a relevant part of our efforts are devoted to identifying their *universal* behaviours. The manuscript is organised as follows.

The first part deals with general probabilistic theories, introducing the formalism and discussing within this framework some specific problems connected with the question raised above. More in detail, in Chapter 1 we lay the foundations of GPTs, reviewing some compelling derivations of their formalism from first principles. Our main aim there is to provide a fully self-contained proof of one of the main foundational results in the field, the embedding theorem by Ludwig [7]. The purpose of this theorem is to start from elementary and intuitive assumptions concerning the physical system under examination and our ability to access it through measurements, and to translate them into a convenient mathematical picture, which will constitute the arena where our questions find an adequate formulation.

This mathematical picture features a so-called *base norm Banach space*, a particular type of ordered Banach space whose norm is closely related to its order relation. The main difficulty in the proof is that in general there is no reason to assume said space to be finite-dimensional. We review Ludwig’s successful attempt to overcome this technical hurdle, which requires us to discuss some aspects of the theory of Banach spaces, with increasing specialisation to ordered Banach spaces and eventually to base norm and order unit spaces. Along the way, many classic results in the field are rederived. We believe that our review of this topic constitutes a valuable contribution, especially as we strove to give a unified and coherent presentation of the material dispersed over several original papers that are not always easy to read. For certain points in the proof of Ludwig’s theorem we give alternative arguments, which may be helpful to the non-expert on this topic, and may be of independent interest.

In Chapter 2 we review the basics of the GPT formalism in the simplified finite-dimensional setting, extending it as to include bipartite systems, and exploit its full power to ask the following question: *is the appearance of entangled states logically implied when one joins two non-classical systems?* Here, a GPT is called ‘non-classical’ if there is a set of pure states (pure meaning that they can not be obtained by any nontrivial probabilistic mixture) that is linearly dependent, in the sense that the corresponding expectation values on any observable are linearly dependent. A state of a bipartite system is called ‘entangled’ if it can not be prepared with local operations and classical

communication starting from uncorrelated states.

In this context, the question we ask translates to a precise mathematical problem concerning convex cones: *is the maximal tensor product of any two non-simplicial cones always strictly larger than their minimal tensor product?* We conjecture that this is indeed the case, and present substantial evidence that supports this view. In particular, we prove the conjecture in the particular case when both cones have a centrally symmetric section. On a different line, we also consider a reformulation of the above question, where one looks at nonlocality rather than entanglement. Mimicking a classic argument by Namioka and Phelps [8], we show that the combination of any non-classical theory with a ‘gbit’ (a model whose state space is a square) gives rise to states that are not only entangled but that can even violate the simplest Bell inequality, i.e. the CHSH inequality. Along the way, we develop tools to tackle the separability problem in arbitrary GPTs.

In Chapter 3 we shift the focus of our analysis from entangled states to entangled measurements. One of the most striking consequences of the existence of global measurements that can not be performed locally is the phenomenon of *data hiding*, i.e. the possibility of hiding a bit of information in a state discrimination query that two distant parties are unable to resolve unless they have access to joint measurements, even when classical communication is available for free [9, 10]. We introduce a natural figure of merit for this task, called efficiency, and then ask ourselves: *how high can the efficiency of quantum data hiding be, given the two dimensions of the local subsystems?* And more generally, *what is the maximum efficiency an arbitrary GPT can exhibit, for fixed local dimensions?*

The answers to these questions, given here in full generality and up to a small universal constant (multiplicative in the first case, additive in the second), constitute two main results of the present thesis. Perhaps surprisingly, it turns out that the maximal quantum data hiding efficiency scales as the square root of the absolute maximum among all GPTs, which is instead achieved for other kinds of highly symmetric theories, namely those whose state space is a sphere. Before we move on to the next chapter, few words about the proof techniques. The argument we devise for the quantum case is surprisingly elementary, as it rests only on the celebrated teleportation protocol, yet it is so powerful that we are able to deduce a number of interesting corollaries from it, mainly leveraging results of [162]. These include the best known lower bound on the squashed entanglement [11] of a state in terms of its trace distance from the separable set, as well as a new algorithm for the separability problem whose running time scales only polynomially in the size of the larger subsystem. In facing the case of arbitrary GPTs, which is significantly more involved, we uncover an unexpected connection with the branch of functional analysis that studies *tensor norms*. The solution of the problem in that setting rests on the determination of the largest ratio between projective and injective tensor norms that can be achieved by any pair of Banach spaces of fixed dimensions, a result that may be of independent interest.

The second part of the present thesis focuses on some selected aspects of

quantum entanglement, both in finite-dimensional and continuous variable systems. In Chapter 4 we introduce a class of maps that are natural extensions of the well-known depolarising channel to the case of a bipartite input system. The operational significance of these maps, parametrised by three variables, rests on their capability of modelling both local and global white noise acting on a shared quantum state. In this context, we analyse all basic properties connected with entanglement transformation, thus determining the parameter regions for which the maps are: (a) positive; (b) completely positive; (c) entanglement-breaking, meaning that they break any entanglement that was previously shared between the system and the rest of the world; and (d) entanglement-annihilating, in the sense that any input (bipartite) state is transformed into a separable state. This latter characterisation allows us to close some open problems that have been raised in recent literature [12, 13]. Along the way, we develop new tools to tackle the quantum separability problem, harnessing the properties of the Hadamard product in this context.

Chapter 5 is devoted to Gaussian entanglement, i.e. entanglement of Gaussian states in continuous variable quantum systems. Due to the well-known fact that these states are in one-to-one relationship (up to local unitaries) with special positive definite matrices that we dub *quantum covariance matrices*, we advocate the general strategy of using matrix analysis tools to solve problems on Gaussian states. Two tools are found to be specially useful in our approach, namely *Schur complements* and *matrix means*. Extensively exploiting these techniques, we present a unified proof of the equivalence between separability and positivity of the partial transpose (so-called PPT condition) for Gaussian states of 1 vs n modes, for arbitrary n [14, 15]. This constitutes a notable improvement over previous works. Besides rederiving all classic results establishing said equivalence in various other special cases, we extend their validity considerably. For instance, we prove that bipartite Gaussian states that stay PPT under the application of any passive operations are necessarily separable, which closes an old open problem [16].

The third and last part of the manuscript is devoted to the study of other forms of non-classical correlations, especially in the continuous variable setting. In Chapter 6 we look into *steering*, a notion which lies between entanglement and nonlocality, laying emphasis on the case of Gaussian steering, i.e. steering of Gaussian states by Gaussian measurements. We start by showing that the Schur complements of quantum covariance matrices obey a matrix inequality whose structure resembles that of strong subadditivity of quantum entropy. Leveraging this result, we prove that a certain quantifier of Gaussian steerability based on the covariance matrix of the state [17] is an actual monotone of the corresponding resource theory, and thus a fully-fledged Gaussian steerability measure. The focus is then shifted towards the problem of ‘classifying’ correlations of bipartite Gaussian states into classical and quantum. In the approach of [18], this can be done in a consistent way if the quantifier of quantum correlations does not exceed *half* the quantifier of total correlations. Using the theory of matrix means, we prove that the Rényi-2 Gaussian entanglement formation and the corresponding Rényi-2 mutual information do satisfy this

constraint when evaluated on Gaussian states. Monogamy and additivity of the former measure are immediately deduced from this result.

The investigation of these problems is pursued in the final Chapter 7. There, we examine some questions related to the strong subadditivity matrix inequality of Chapter 6. Exploring some connections with probability theory on the one hand and with matrix analysis on the other hand, we provide several improvements of the inequality, determining in particular several remainder terms from which the saturation conditions are easily read. Subsequently, we consider an entanglement measure for Gaussian states that is very similar in spirit to the *squashed entanglement* [11], with the important differences that: (i) the Rényi-2 entropy displaces the standard von Neumann entropy; and (ii) the extensions of the state are required to be Gaussian as well. Surprisingly, it can be shown that – unlike in the finite-dimensional case – this Rényi-2 Gaussian squashed entanglement coincides with the corresponding Rényi-2 Gaussian entanglement of formation on all Gaussian states. Among its many advantages, this remarkable result makes the numerical computation of the former measure possible, since it restricts the size of the extension from unbounded to bounded.

The various chapters of this thesis are based on the following publications or preprints.

Chapter 3. L. Lami, C. Palazuelos, and A. Winter. Ultimate data hiding in quantum mechanics and beyond. *Preprint arXiv:1703.03392*, 2017.

Chapter 4. L. Lami and M. Huber. Bipartite depolarizing maps. *J. Math. Phys.*, 57(9):092201, 2016.

Chapter 5. L. Lami, A. Serafini, and G. Adesso. Gaussian entanglement revisited. *Preprint arXiv:1612.05215*, 2016.

Chapter 6. L. Lami, C. Hirche, G. Adesso, and A. Winter. Schur complement inequalities for covariance matrices and monogamy of quantum correlations. *Phys. Rev. Lett.*, 117(22):220502, 2016.

Chapter 7. L. Lami, C. Hirche, G. Adesso, and A. Winter. From log-determinant inequalities to Gaussian entanglement via recoverability theory. *IEEE Trans. Inf. Theory*, Preprint, 2017.

The first two chapters contain substantially new material, which we intend to publish soon. Finally, the papers that I have contributed to write during my PhD but that are not quite covered in the present thesis are the following.

- M.G. Genoni, L. Lami, and A. Serafini. Conditional and unconditional Gaussian quantum dynamics. *Contemp. Phys.* 57(3):331–349, 2016.
- F. Clivaz, M. Huber, L. Lami, and G. Murta. Genuine-multipartite entanglement criteria based on positive maps. *J. Math. Phys.* 58(8):082201, 2017.

- L. Lami, S. Das, and M.M. Wilde. Approximate reversal of quantum Gaussian dynamics. *Preprint arXiv:1702.04737*, 2017.
- K.P. Seshadreesan, L. Lami, and M.M. Wilde. Rényi relative entropies of quantum Gaussian states. *Preprint arXiv:1706:09885*, 2017.

We conclude with a few remarks on the organisation of the manuscript. The first section of every chapter starts with few sentences that introduce the topic, followed by a quick overview of the structure of the chapter. Then there are two subsections: the first one introduces the problem in broad terms, while the second points the reader to our main original contributions. Chapter 1 presents an additional subsection that aims at sketching the history of general probabilistic theories and giving a concise list of pertinent references.

We are now ready to set out, hoping that the reader will follow us on this journey.

Part I

Non-classicality in general probabilistic theories

Chapter 1

Foundations of General Probabilistic Theories

1.1 Introduction

The glory of quantum mechanics penetrates the whole universe, and shines more brightly when one looks at it from certain perspectives, from others less. Throughout this chapter, we are going to take one of those points of view, from which we hope its brilliance to be more apparent. Our journey to this standpoint will lead us to seeing quantum mechanics as a part of a rich landscape whose landmarks are physical theories on their own, generically called General Probabilistic Theories (GPTs). This is profoundly instructive for at least two reasons. First, it will help us discerning the various peculiarities of quantum mechanics that distinguish it from a classical theory, and establishing a hierarchy among them. In fact, from this point of view it turns out that not all the surprising quantum effects we have been studying and understanding with difficulty throughout the past century should bewilder us to the same extent. Secondly, once this firm foundation has been established, we can turn our attention to the problem of singling out quantum mechanics by means of certain additional assumptions that are believed to be crucial to the construction of a complete physical theory. This latter approach seems to be one of the main reasons that has encouraged many scholars to investigate GPTs, and while it is certainly commendable and conceptually sound, the point of view we are going to adopt here is somewhat closer to the former.

The content of the present chapter is organised as follows. The rest of this section is devoted to sketching the history of the field (Subsection 1.1.1), to introducing the basic framework (Subsection 1.1.2) and to discussing what distinguishes our approach from that taken by others and highlight our modest original contribution (Subsection 1.1.3). Sections 1.2–1.6 are intended to acquaint the reader with the mathematical tools that will be used later on. More in detail, we will first present the notions of ordered vector space (Section 1.2), of topological vector space (Section 1.3), and of Banach space (Section 1.4),

then combine them to study ordered Banach spaces in general (Section 1.5) and two special cases known as order unit and base norm spaces (Section 1.6). The reader who is already familiar with the above notions can jump straight away to Section 1.7, where we apply all this mathematics to present the proof of Ludwig's embedding theorem, which lays the foundation of the so-called abstract state space formalism for GPTs.

1.1.1 Brief history of General Probabilistic Theories

As part of our introduction to GPTs, here we intend to give a brief account of the development of the field. The list of references we provide is by no means complete, since especially in recent years there has been much activity in trying to extend the framework to study all sort of problems. However, we hope that the reader will benefit from our glance into the history of GPTs.

As usual, drawing the starting line is a bit arbitrary, since in modern times every scientist is influenced by many others. That being said, the first author to develop an axiomatic approach to quantum mechanics that ultimately led to the formalisation of GPTs was perhaps Mackey [19] at the beginning of the Sixties. In his book, the prepare-and-measure scheme, the associated probability function and the axioms to be imposed on the latter are introduced and discussed. Seemingly independently, Ludwig [20] worked on a similar framework, further developed in a series of papers published in *Communications in Mathematical Physics* between 1967 and 1972 by Ludwig, Dähn and Stolz [21, 22, 23, 24, 25, 26].

The authorship of the so-called 'abstract state space formalism' that has nowadays become standard for discussing GPTs is hard to attribute. On the one hand, the first proof that the experimentally accessible probability function can be reconstructed by evaluating positive functionals on a subset of some Banach space seems to go back to Ludwig. The content of this crucial statement, which we report here as Theorem 1.43 together with a hopefully more accessible and simplified proof, reached its final form in [27] (see also [7]) after being developed throughout several papers [20, 21, 22]. On the other hand, in the very same year 1970 Davies and Lewis published a paper [28] where they devise such a Banach space formalism in its modern form and use it as a starting point.

This latter paper triggered an animated debate during the next few years, and quick advances were made in improving the basics of the framework [29, 30, 31, 32], expanding it as to encompass state transformations [33, 34], and connecting it to previous approaches [35, 36, 37]. More or less in the same years, Mielnik showed that this class of models is general enough as to support non-linear extensions of quantum mechanics, as could result for instance from a hypothetical inclusion of gravity in the picture [38]. From our point of view, a particularly important progress that was made several years later is the discussion of composite systems [39, 40] (see also [41] for a generalisation to the multipartite case).

All the aforementioned authors made ample use of the mathematical theory of ordered topological vector spaces, whose development in the previous years is summarised in [42, 43, 44]. Two particular classes of ordered topological (in fact, Banach) spaces quickly became central to the formalism, namely order unit spaces and base norm spaces, introduced by Ellis [45, 46] and Edwards [47], respectively.

In more recent times, the framework of general probabilistic theories has been used for instance to investigate non-locality phenomena [48, 49, 50, 51], communication complexity questions [52, 53, 54, 50, 55], the complications arising from introducing post-measurement collapses in the picture [56, 57], and the status of the purification postulate [58]. Several concepts that played a central role in the development of quantum information have been revisited within the GPT realm, such as teleportation [41] and broadcasting [59] protocols (see also [60]), steering [61], entropy [62], and entanglement [63], especially in connection with thermodynamics [64].

On a different line, we already mentioned how the problem of deriving quantum theory from few reasonable axioms has attracted much attention since the very early days of the field. In fact, already von Neumann and co-workers looked into it from the algebraic point of view, and reconstructing quantum theory was one of the main motivations behind the works of Mackey [19] and Ludwig [27, 7, 65] themselves. We will not attempt to sketch a history of the problem here, and for details on early attempts to solve it we refer the reader to [66]. The advent of general probabilistic theories has seen a widespread renewal of interest in this kind of questions, and several new approaches were put forward in recent years. For the sake of the presentation, let us stick to the simplistic idea that these attempts differ from each other in the somewhat decisive axiom that singles out quantum theory at last, separating it from classical probability theory. In [67, 68, 69, 70] this is the existence of (continuous) reversible transformations among pure states, while in [71] an analogous role is played by the assumption that every state can be purified (in an almost unique way). The existence of a ‘symmetric faithful state’ lies at the heart of the treatment given in [72, 73], while an algebraic postulate whose interpretation is not completely transparent does the job in [74]. A somewhat different approach has been pursued in [75], where more exotic theories are excluded based on the non-existence of correlations of ‘order’ higher than two.

1.1.2 Preparing and measuring procedures

The starting point of the reasoning that leads us to the study of general probabilistic theories is a lesson we learnt in a very early stage of the quantum revolution: *our basic requests from a physical theory should not go much beyond a set of rules that allow us to deduce a probabilistic prediction of the outcome of an experiment from the detailed description of its preparation.* We do not claim that this is necessarily all we should expect from a physical theory, but rather that these are the minimal requirements for something to be considered an acceptable physical theory.

Here we do not intend to delve into a complete discussion of the many subtleties incident to such an approach, for which we rather refer the reader to Ludwig's analysis [65]. Let us just stress a couple of points. First, the probabilities themselves are not directly accessible experimentally. Under the hypothesis of repeatability of independent trials, what is accessible are the relative frequencies, that will converge to the abstract probabilities only in the limit. This aspect must be taken into account in the mathematical formulation of the theory. In Ludwig's approach, this is done by means of an additional 'uniform structure of physical imprecisions'. Secondly, any axiomatisation of a physical theory that follows the above scheme requires an additional pre-theory to describe the experimental apparatus. Describing a quantum experiment of Stern-Gerlach type, for instance, requires a prior knowledge of electromagnetism. We now turn to detailing our pragmatic view of a physical experiment and of the associated theory.

In a somewhat simplistic view, an experimental setting consists in being given some physical system, say a laser, and some apparatus to measure it, say a set of lenses, mirrors and photon counters. An experiment will then consist in choosing a way of preparing the system and a configuration of the measuring apparatus. The set of all possible preparation procedures, also called **states**, will be denoted by Ω . As for the measuring apparatus, in the era of digital devices it is clear that we can assume with basically no loss of generality that its output consists in a sequence of binary numbers (yes/no). Measuring the system means recording the yes/no pattern of a particular experiment. Therefore, it makes sense to define an **effect** as a configuration of the measuring apparatus together with a yes/no pattern that belongs to it. We will call Λ the set of all possible effects. From this point of view, and when restricted to this particular setting, a physical theory is nothing but a probability function $\mu : \Lambda \times \Omega \rightarrow [0, 1]$, that associates a probability $\mu(\lambda, \omega)$ to the event of recording the effect λ when measuring the system that has been prepared according to ω .

Naturally, this function μ must be subjected to some nontrivial constraints if we want it to represent a physically realistic scenario. First of all, we can assume that we already made the preliminary step of identifying all preparations that were (even probabilistically!) indistinguishable from each other by all measurements, i.e. such that they always yielded the same probabilities when the same effect was measured. The same is true for the effects: we preliminarily chose to identify all effects that yielded the same probabilities when evaluated on the same states. This allows us to assume the following.

Axiom 1. States that are not distinguishable by any effect are to all intents and purposes the same state, and vice versa: mathematically, one says that μ separates points of both Ω and Λ , i.e. that for all distinct $\omega_1, \omega_2 \in \Omega$ there is $\lambda \in \Lambda$ such that $\mu(\lambda, \omega_1) \neq \mu(\lambda, \omega_2)$, and vice versa that for all distinct $\lambda_1, \lambda_2 \in \Lambda$ there is $\omega \in \Omega$ such that $\mu(\lambda_1, \omega) \neq \mu(\lambda_2, \omega)$.

Next, we can posit that the effects with certain outcomes 'always yes' and 'always no' are among those that we are allowed to consider, and that negating

a valid effect yields another valid effect. This is the content of the following axiom.

Axiom 2. Λ contains the effects ‘always accept’ and ‘never accept’: mathematically, there are $u \in \Lambda$ (also called the **unit effect**) and $0 \in \Lambda$ such that $\mu(u, \omega) = 1$ and $\mu(0, \omega) = 0$ hold for all states $\omega \in \Omega$. Furthermore, for all $\lambda \in \Lambda$ there is $\lambda' \in \Lambda$ such that $\mu(\lambda', \omega) = 1 - \mu(\lambda, \omega)$ for all $\omega \in \Omega$.

Finally, we want to be able to incorporate all pre- and post-randomisations of the preparations and of the outcomes in the picture. In other words, any probabilistic mixture of accessible preparations must be also accessible, and the same for the effects. We state this as follows.

Axiom 3. One can perform probabilistic mixtures of states (or effects) and obtain a valid state (or effect); this means that for all $\omega_1, \omega_2 \in \Omega$ and $p \in [0, 1]$ there exists $\tau \in \Omega$ such that $\mu(\lambda, \tau) = p\mu(\lambda, \omega_1) + (1 - p)\mu(\lambda, \omega_2)$ for all $\lambda \in \Lambda$, and vice versa that for all $\lambda_1, \lambda_2 \in \Lambda$ and $q \in [0, 1]$ there exists $\eta \in \Lambda$ such that $\mu(\eta, \omega) = q\mu(\lambda_1, \omega) + (1 - q)\mu(\lambda_2, \omega)$ for all $\omega \in \Omega$.

The axiomatic picture we just sketched is certainly convincing from the foundational perspective, but it has at least two annoying deficiencies. First, it lacks any geometrical interpretation, and this makes the discussion of even basics concepts a lot more cumbersome. As an example, we invite the reader to formulate the notion of pure state in the above language. Secondly, and perhaps more importantly, the seemingly poor structure with which the sets Ω and Λ are endowed makes the picture totally inconvenient for computations. From a more pragmatic point of view, we could also add to this list the fact that since to this point there is no apparent connection between our framework and any well-studied mathematical theory, we will have a hard time employing standard mathematical tools to tackle our problems.

Fortunately, these deficiencies are really just superficial, and there is an equivalent way of formulating the above framework that fixes all of them at once. This alternative formulation comes in the form of an embedding theorem that allows us to see Ω and Λ as convex subsets of a Banach space and of its dual (respectively), with the probability function μ being given by the canonical bilinear form (for details, see Theorem 1.43). Ultimately, the purpose of this operation is to translate simple and easily understandable physical axioms into a convenient mathematical picture that constitutes the arena of GPTs. To the best of our knowledge, the first to present a complete proof of this remarkable fact – i.e. a proof that does not rely on any further assumption such as the finite dimensionality of the involved linear spaces – was Ludwig [7, IV §§1, 3 and 4] (see also [27]), and for this reason we name it *Ludwig’s embedding theorem*.

1.1.3 The path to Ludwig’s embedding theorem

Ludwig’s original proof [7, IV §§3 and 4] of Theorem 1.43 is deep and elegant, but arguably a bit hard to read from a modern perspective, and even obscure

in certain passages. In my personal view, this difficulty comes primarily from the heterogeneity of the mathematics he employs, and secondly from the sometimes bizarre notation he adopts. In particular, there the reader is assumed to have a strong background in basic topology, and to a lesser extent also in the theory of ordered Banach spaces. The result is a proof that is almost 17 pages long, the last 6 pages being an application (with proofs) of the duality theory of base norm and order unit spaces as developed by Edwards [47] and Ellis [45]. Since Ludwig's theorem is perhaps the foundational result of the field of general probabilistic theories, and its full proof is rather unknown, we believe it important to discuss it here in its entirety. We want to make our small contribution to Ludwig's cause by rewriting his proof the way we understood it, hoping that this will make it more accessible to a broader audience.

In our view, the strength of our presentation lies primarily in the fact that we will not assume any prior knowledge of anything but basic linear algebra. The relevant concepts and results in topology and Banach space theory are discussed and proven in the preliminary Sections 1.2–1.4. We will state without proofs only few standard results that the reader can find in any textbook (Lemmas 1.3, 1.6 and Theorems 1.7, 1.10, 1.14), and anyway appropriate references will be provided. Then, we move on to discussing the theory of ordered Banach spaces (Section 1.5) and in particular of order unit and base norm spaces (Section 1.6). We will spend some time proving from first principle important results like Theorem 1.21 on the duality of cones and Theorem 1.42 on the duality between base norm and order unit spaces. Here we make no claim of originality, since the theory is well-known since the Sixties. Rather, we thought of improving the presentation by collecting all the results that are relevant to the proof of Ludwig's theorem in a single place. A similar operation had already been partially attempted in the series of lecture notes [76], which was of inspiration to us.

With all the mathematical tools already developed in a systematic way in the preceding sections, we can finally delve into the proof of Theorem 1.43 in Section 1.7. Most of the proof is a sequence of ingenious applications of elementary concepts, and we limit ourselves to spelling out some elementary steps that are missing in Ludwig's original presentation, for instance the proof of Proposition 1.44 or the reasoning in Subsection 1.7.9, and to prove more directly some facts like Proposition 1.47. However, there are two places where more advanced mathematical tools come into play, the first one being Subsection 1.7.7, where Theorem 1.42 is applied to shorten Ludwig's reasoning significantly, and the second one being Subsection 1.7.8, where we are forced to invoke several nontrivial results in topology and Banach space theory. We tried to simplify this latter part by proposing a different and perhaps (depending on the personal taste) more elementary proof of Proposition 1.49. Finally, even if not strictly necessary for completing the proof, we provided an alternative and completely elementary reasoning to show the uniqueness part of the claim of Theorem 1.43 in Appendix B.

As the reader might have guessed, the present chapter presents a series of results that are highly interconnected and ultimately work towards proving

Theorem 1.43. We tried to summarise the network of interdependencies by means of the following graph, whose vertices are lemmas, propositions, theorems and corollaries of this chapter, and whose directed edges represent logical connections. The reader can use it to choose what part of the preliminary mathematics he/she wants to gain a deeper knowledge of, if needed.

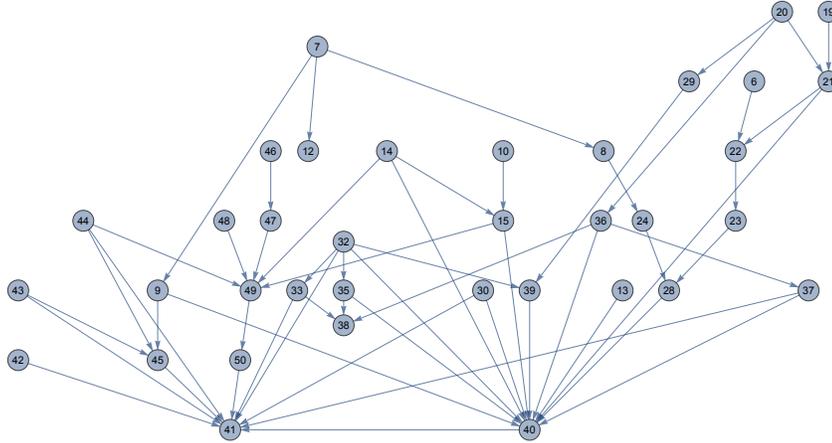


Figure 1.1: Conceptual links among all Lemmas, Propositions, Theorems and Corollaries of the present chapter. An arrow from vertex x to vertex y means that x is needed to prove y .

1.2 Ordered vector spaces

Linear algebra has always been a useful tool in the hand of a physicist and in particular of a mathematical physicist. However, with the advent of quantum mechanics, it has acquired a somewhat different status, becoming rather a conceptually central field for the mathematical formulation of physics itself. Nowadays, the concept of vector space is so ubiquitous in mathematical physics, that it becomes not only interesting but almost unavoidable to study its interplay with some other major mathematical objects. For instance, in particle physics the concept of linear representation of a group becomes relevant, and representations are nothing but vector spaces endowed with some additional structure, in this case a subgroup of the corresponding linear group.

Analogously, a notion that plays a central role in this thesis is that of ordered vector space. We assume that the reader is familiar with vector spaces and their basic properties, but no more knowledge than this is required, in that any other concept we will need throughout the chapter will be introduced and discussed when relevant. Perhaps I should say at this point that in the field

of quantum information vector spaces are often surreptitiously assumed to be finite-dimensional. It goes without saying, this technical assumption makes our life a lot easier, and for the purpose of asking the kind of questions we will be investigating in this thesis is even fundamental. Thus, starting from the next chapter we will always assume all vector spaces to be finite-dimensional. However, since now we want to discuss the foundations of general probabilistic theories, we will *not* make this assumption in this chapter unless explicitly stated.

The material of this section is covered by many excellent textbooks, most notably [42, 43, 44] as well as the excellent set of lecture notes [76]. Let us start by reminding the reader that an **ordering** on a set X is a binary relation \leq on X that is reflexive ($x \leq x$ for all $x \in X$), transitive ($x \leq y, y \leq z$ implies $x \leq z$ for all $x, y, z \in X$) and anti-symmetric ($x \leq y$ and $y \leq x$ is only possible when $x = y$). A crucial fact to keep in mind is that in general \leq is *not required to be total*, i.e. it might happen that for $x, y \in X$ neither $x \leq y$ nor $y \leq x$. We are now ready to introduce one of our main subjects of study throughout this chapter.

Definition 1.1. An *ordered vector space* is a real vector space E equipped with an ordering \leq which is:

- (i) *translationally invariant*, meaning that $x \leq y \Rightarrow x + z \leq y + z, \forall x, y, z \in E$;
- (ii) *positively homogeneous*, i.e. $x \leq y \Rightarrow \lambda x \leq \lambda y, \forall x, y \in E, \lambda \geq 0$.

Let us now discuss some related concepts and constructions. By the end of this discussion, we hope that the reader who is less interested in the case of E having infinite dimension will anyway be aware of the main differences with the finite-dimensional case. To help identifying these subtleties, we will point them out as soon as they arise. For further details, we refer the reader to [76, ‘Orderings of vector spaces’]. Let us start by fixing some terminology.

Definition 1.2. A subset $C \subseteq E$ of a real vector space E is called a **cone** if:

- (i) *it is closed under sums*, i.e. $C + C = C$;
- (ii) *it is closed under multiplication by positive scalars*, i.e. $\lambda C \subseteq C$ for all $\lambda \geq 0$; and
- (iii) *does not contain nontrivial vector subspaces*, i.e. $C \cap -C = \{0\}$.

Note. It is understood that for sets $X, Y \subseteq E$ one defines $X + Y := \{x + y : x \in X, y \in Y\}$, and analogously $\lambda X = \{\lambda x : x \in X\}$ for scalars $\lambda \in \mathbb{R}$.

Note. Some authors prefer to call a set that satisfies requirements (i)-(iii) of Definition 1.2 a ‘salient convex cone’ or also a ‘proper cone’. Usually, when this nomenclature is adopted a cone is simply required to be closed under multiplication by positive scalars. The convention we adopt here follows that in [45].

In every ordered vector space the set $E_+ = \{x \in E : x \geq 0\}$ is easily seen to be a cone. In fact, since the order relation can be rewritten in an equivalent way as $x \leq y \Leftrightarrow y - x \in E_+$, one sees that giving an ordering on a vector space is the same as choosing a cone. As an easy consequence of this reformulation, observe that $x \leq y \Rightarrow -y \leq -x$. With a rather obvious terminology, a cone C is said to be **spanning** if $\text{span}(C) = E$, which can equivalently be written as $C - C = E$. For $x, y \in E$ with $x \leq y$, the corresponding **interval** is defined as $[x, y] := \{z : x \leq z \leq y\}$.

Given any real vector space E , we can consider the associated vector space of linear functionals $\varphi : E \rightarrow \mathbb{R}$, called the **algebraic dual** of E and denoted by E' . For $x \in E$ and $\varphi \in E'$, we will often write $\langle \varphi, x \rangle$ instead of $\varphi(x)$. This defines a bilinear form $\langle \cdot, \cdot \rangle : E' \times E \rightarrow \mathbb{R}$, also called the **canonical bilinear form**.

Note. Most standard textbooks (like [77]) define the canonical bilinear form with the two arguments exchanged, writing $\langle x, \varphi \rangle$ instead of $\langle \varphi, x \rangle$. We prefer instead this latter convention since it is reminiscent of the Dirac notation in quantum mechanics, with dual vectors (bras) on the left and primal vectors (kets) on the right.

When E is also an ordered vector space, also its dual E' is naturally ordered by the **dual cone** E_+^{\otimes} formed by all functionals that are positive on E_+ . The subspace $E^{\otimes} := E_+^{\otimes} - E_+^{\otimes} \subseteq E'$ is again an ordered vector space, called the **order dual** of E . It is not difficult to realise that if E is finite dimensional and E_+ is in addition closed, then the order dual coincides with the dual vector space, i.e. $E^{\otimes} = E'$.

1.3 Topological vector spaces

As we said, throughout this chapter we want to stay as general as possible, without making any assumption on the dimension of the vector spaces we will encounter along the way. The difficulties in dealing with infinite-dimensional spaces emerge especially when one comes to the somewhat risky business of taking duals.¹ For this reason, many vector spaces come with a topological structure that identifies a special class of functionals within the algebraic dual E' , namely the continuous functionals.

In order to identify a concept of continuity of functions, we need some notion of neighbourhood. While finite-dimensional (real) spaces inherit such concepts from the isomorphism with \mathbb{R}^n identified by a basis, in the infinite-dimensional case we must enforce an analogous notion from the outside, or mathematically speaking we must equip the space with a topology. We remind the reader that a **topology** on a set X is a collection $\tau \subseteq \mathcal{P}(X)$ of subsets of X (called **open sets**) that [78, I]:

¹For instance, it is known that the dual of an infinite-dimensional vector space has always a larger dimension – in the sense of cardinalities – than the original space.

- (i) includes \emptyset and X ;
- (ii) is closed under unions; and
- (iii) is closed under finite intersections.

A set X endowed with a topology is called a **topological space**. Our archetypical example of a topological space is the real line \mathbb{R} equipped with the standard notion of open sets. Let us summarise quickly how some elementary constructions we learnt in that context generalise to arbitrary topologies. For more thorough discussions we refer the reader to [78], while a concise yet comprehensive introduction to the basic concepts can also be found in [79, II].

- Topologies τ_1, τ_2 on the same set X can be compared. If $\tau_1 \subseteq \tau_2$ we say that τ_1 is **coarser** (or **weaker**) than τ_2 .
- A **neighbourhood** of $x \in X$ is an open set $U \in \tau$ that contains x . An extremely weak requirement to impose on a topology is that it is **Hausdorff** or **separating**, i.e. such that every two disjoint points admit disjoint neighbourhoods. All the topology we are going to discuss here will be Hausdorff.

For an arbitrary subset $S \subseteq X$ of a topological space X , those points $x \in S$ admitting a neighbourhood $U \subseteq S$ form the **interior** of S , denoted with $\text{int}(S)$ or with $\text{int}_\tau(S)$ when we want to specify the topology τ . Observe that if τ_1 is coarser than τ_2 , then τ_1 -neighbourhoods are a fortiori τ_2 -neighbourhoods, hence $\text{int}_{\tau_1}(S) \subseteq \text{int}_{\tau_2}(S)$.

- A subset K of a topological space (X, τ) is said to be **closed** if its complement K^c in X is open, i.e. if $K^c \in \tau$. Intersections and finite unions of closed sets are again closed.

For a fixed set $S \subseteq X$, the smallest closed superset of S in X (i.e. the intersection of all closed supersets of S) is called the **closure** of S , and denoted with $\text{cl}(S)$ or with $\text{cl}_\tau(S)$ when there is need to specify the topology τ . If τ_1 is coarser than τ_2 , then τ_1 -closed sets are a fortiori τ_2 -closed, and hence $\text{cl}_{\tau_1}(S) \supseteq \text{cl}_{\tau_2}(S)$. A set whose closure is the entire space X is called **dense** in X .

- Another important concept in topology is that of compactness. A set $S \subseteq X$ in a topological space (X, τ) is said to be **compact** if every open cover of S has a finite subcover, i.e. if for all families $\{U_i\}_{i \in I} \subseteq \tau$ such that $\bigcup_{i \in I} U_i \supseteq S$ one can find $n \in \mathbb{N}$ and $i_1, \dots, i_n \in I$ such that $\bigcup_{k=1}^n U_{i_k} \supseteq S$. Evidently, every finite collection of points is compact. It is not difficult to realise that closed subsets of compact sets are again compact. Moreover, if the underlying topology is Hausdorff, compact subsets are always closed.

In a Euclidean space \mathbb{R}^n equipped the standard topology, S is compact iff it is closed and bounded, but no such simple characterisations are

available for more general topological spaces. An intuitive rule of thumb it is useful to remember is that in many circumstances compact sets behave like single points. Now, let us present a lemma that is not so commonly found but will be needed in order to prove Ludwig's embedding theorem (Theorem 1.43).

Lemma 1.3. [80, I §9.4, Corollary 3]. *Let (X, τ) be a compact topological space. A Hausdorff topology on X that is coarser than τ must coincide with the latter.*

- Topological spaces can be multiplied, in the sense that the Cartesian product $X_1 \times X_2$ of two topological spaces X_1, X_2 can be endowed with the **product topology** whose open sets are unions of products of local open sets.
- A function $f : X_1 \rightarrow X_2$ between two topological spaces is said to be **continuous** when the pre-images of open sets are again open, in symbols $f^{-1}(U_2) \in \tau_1$ for all $U_2 \in \tau_2$. Alternatively, one can cast this condition in terms of closed sets by saying that f is continuous when $f^{-1}(K_2)$ is closed for all τ_2 -closed sets K_2 . It does not take much to realise that for real functions $f : \mathbb{R} \rightarrow \mathbb{R}$ this is just a rephrasing of the notion of continuity we learn in a standard course on calculus.

Reverse engineering the process, one can use functions to define a topology. Namely, if X is a set and Y is a topological space, for a family $\mathcal{F} = \{f_i\}_{i \in I}$ of functions $f_i : X \rightarrow Y$ there is a coarsest topology on X such that all functions f_i are continuous. The open sets of this topology are precisely the unions of finite intersections of sets of the form $f_i^{-1}(U)$, where $U \in \tau_Y$. In many cases we will have $Y = \mathbb{R}$ with the standard topology, and we will refer to the above construction as the **initial topology** generated by \mathcal{F} , usually denoted by $\sigma(X, \mathcal{F})$.

- In the case of \mathbb{R} with the standard topology, many useful notions can be conveniently rephrased in terms of sequences. Perhaps surprisingly, in the general case the notion of sequence is not quite sufficient, and we must resort to that of a **net** [78, II]. A net on a set X is a function $x : A \rightarrow X$, usually denoted by $(x_\alpha)_{\alpha \in A}$, whose domain A is a **directed set**, i.e. an ordered set in which every two elements $a, b \in A$ admit a common upper bound $a, b \leq c \in A$. When $A = \mathbb{N}$ is the set of natural numbers, a net becomes just a standard sequence. Nets on topological spaces are useful because one can give a definition of limit. Namely, $\tilde{x} \in X$ is said to be a **limit of a net** $(x_\alpha)_\alpha \subseteq X$ if for all neighbourhoods V of \tilde{x} there is $\alpha \in A$ such that $x_\beta \in V$ holds for all $\beta \in A$ such that $\beta \geq \alpha$. When the underlying topology is Hausdorff, which will be always the case in this thesis, limits of nets (when they exist) are unique. In this case, we write $\lim_\alpha x_\alpha = \tilde{x}$.

How to define subnets, in analogy with subsequences? We say that a net $(y_\beta)_{\beta \in B}$ is a **subnet** of a net $(x_\alpha)_{\alpha \in A}$ when there is a function $\phi : B \rightarrow A$

such that: (i) $y = x \circ \phi$; and (ii) for all $\alpha \in A$ there is $\beta \in B$ such that $\beta' \geq \beta \Rightarrow \phi(\beta') \geq \alpha$, for all $\beta' \in B$. Intuitively, this latter condition ensures that $\phi(\beta)$ becomes large when so does β . It is easy to show that if \tilde{x} is a limit of the net $(x_\alpha)_\alpha$, then it is also a limit of all subnets $(y_\beta)_\beta$ of $(x_\alpha)_\alpha$.

The concept of net is useful because it allows to recast several topological concepts in a form that lends itself to convenient manipulations. For instance, it can be shown that a subset $K \subseteq X$ is closed iff it contains all the limits of its nets, i.e. iff for all nets $(x_\alpha)_\alpha \subseteq K$ converging to $x \in X$ one has $x \in K$. Moreover, K is compact iff every net $(x_\alpha)_\alpha \subseteq K$ admits a subnet that converges to some $x \in K$. Continuity of functions can be characterised in a similar way: a function $f : X \rightarrow Y$ between two topological spaces X, Y is continuous iff for all nets $(x_\alpha)_{\alpha \in A} \subseteq X$ converging to $\tilde{x} \in X$ the corresponding net $(f(x_\alpha))_{\alpha \in A} \subseteq Y$ converges to $f(\tilde{x}) \in Y$. These results are totally analogous to the usual reformulations of closedness and compactness in a Euclidean space.

The following definition is an example of how the concept of topology can be reconciled with that of an underlying algebraic structure. In our case, the most interesting algebraic structure is that of a vector space. For details, we refer the reader to the monographs [42, 81].

Definition 1.4. A *topological vector space* is a (real) vector space E equipped with a topology τ such that the sum $+: E \times E \rightarrow E$ and the scalar multiplication $\cdot : \mathbb{R} \times E \rightarrow E$ are continuous functions with respect to the product topologies on $E \times E$ and $\mathbb{R} \times E$.

1.4 Banach spaces

There is a very natural way to endow a vector space E with a topology, namely through a **norm**, i.e. a function $\|\cdot\| : E \rightarrow \mathbb{R}_+$ with codomain the set of non-negative real numbers and such that [77, §1.2]:

- (i) $\|x\| = 0 \Leftrightarrow x = 0$;
- (ii) $\|\lambda x\| = |\lambda| \|x\|$ for all $\lambda \in \mathbb{R}$ and $x \in E$;
- (iii) $\|x + y\| \leq \|x\| + \|y\|$ for all $x, y \in E$.

Incidentally, if $\|\cdot\|$ satisfies only (ii) and (iii) then it is called a **semi-norm**. When a vector space is equipped with a norm, we can give the usual definition of an open set as a subset $U \subseteq E$ such that for all $x \in U$ there is $\epsilon > 0$ with the property that $B_\epsilon(x) \subseteq U$, where $B_\epsilon(x) := \{y \in E : \|x - y\| \leq \epsilon\}$ is the **ball** centered in x with radius ϵ . Open sets defined in this way form a topology on E , usually called **norm topology**. When dealing with normed spaces, if we refer to some topological concepts without specifying the underlying topology, this is understood to be the norm topology.

For our purposes, the most interesting case is when E is equipped with a norm $\|\cdot\|$ that makes it **complete**. In this context, completeness means that every **Cauchy sequence** $(x_n)_n$ admits a limit, i.e. there is $x \in E$ satisfying $\lim_n \|x - x_n\| = 0$. As in standard analysis, a sequence of vectors $(x_n)_{n \in \mathbb{N}} \subset E$ is a Cauchy sequence if for all $\epsilon > 0$ there exists $N \in \mathbb{N}$ such that for all $n, m \geq N$ one has $\|x_n - x_m\| \leq \epsilon$. One gives the following fundamental definition [77, §1.2].

Definition 1.5. A **Banach space** is a real vector space E endowed with a norm $\|\cdot\|$ that makes it complete.

Useful references for the theory of Banach spaces are [77, 79], while a concise introduction can also be found in [82, IX].

1.4.1 Some immediate consequences of completeness

The requirement of completeness is in some sense just technical. In fact, it turns out that every normed vector space E can be thought of as a norm-dense subspace of a Banach space \bar{E} called its **completion**.² Many elementary results in Banach space theory hold even if one drops the completeness axiom. However, this assumption plays a crucial role in proving some deeper theorems that are widely regarded as cornerstones of the field such as the so-called uniform boundedness principle (Theorem 1.10). Ultimately, many of these results stem from *Baire's category theorem* [77, Theorem 2.2], which holds only in complete metric spaces. We will not state Baire's theorem, but will content ourselves with presenting one of its Banach space consequences, which we will make use of in Section 1.5. For a direct proof that does not make use of Baire's theorem, we refer the reader to [79, pp.22-23]. Recall that a subset $K \subseteq E$ of a real vector space E is said to be **convex** if $x, y \in K$ implies $px + (1-p)y \in K$, for all $p \in [0, 1]$.

Lemma 1.6. [79, Theorem 1.3.14]. *Let $K \subseteq E$ be a (norm-)closed, convex subset of a Banach space E . If $E = \bigcup_{n \in \mathbb{N}} nK$, then K contains a (norm-)neighbourhood of the origin, i.e. there is $\delta > 0$ such that $\|x\| \leq \delta \Rightarrow x \in K$ for all $x \in E$.*

1.4.2 Hahn-Banach separation theorem

If E is a Banach space when equipped with the norm $\|\cdot\|$, its **Banach dual** E^* is by definition the space of all linear functionals $\varphi : E \rightarrow \mathbb{R}$ that are **continuous** with respect to $\|\cdot\|$, i.e. such that $\lim_n \langle \varphi, x_n \rangle = \langle \varphi, x \rangle$ whenever $(x_n)_n \subset E$ is a sequence that converges to x in the norm topology, i.e. $\lim_n \|x - x_n\| = 0$.

²Furthermore, the completion is unique up to isometric isomorphisms, and can be taken as the space of all equivalence classes of Cauchy sequences in E obtained by the relation $(x_n)_n \sim (y_n)_n \Leftrightarrow \lim_n \|x_n - y_n\| = 0$.

Because of linearity, φ is continuous everywhere iff it is continuous in 0, which is in turn equivalent to φ being **bounded**, i.e. such that

$$\|\varphi\|_* := \sup_{0 \neq x \in E} \frac{|\langle \varphi, x \rangle|}{\|x\|} \quad (1.1)$$

is finite. Another useful fact to keep in mind is that φ is continuous iff its kernel $\ker \varphi = \varphi^{-1}(0)$ is closed. The vector space E^* becomes a Banach space itself when equipped with the norm $\|\cdot\|_*$ defined by (1.1). A useful fact we will make use of is the following. If a functional $\varphi : V \rightarrow \mathbb{R}$ is defined only on a dense subspace $V \subseteq E$ of a Banach space E and there is bounded (in the sense that $\sup_{0 \neq x \in V} \frac{|\langle \varphi, x \rangle|}{\|x\|} < \infty$), then it can always be extended to a continuous functional on the whole space, i.e. with a slight abuse of notation we can write $\varphi \in E^*$ [77, Theorem 3.6]. Moreover, this process of **continuous linear extension**³ preserves the norm, in the sense that

$$\sup_{0 \neq x \in V} \frac{|\langle \varphi, x \rangle|}{\|x\|} = \sup_{0 \neq x \in E} \frac{|\langle \varphi, x \rangle|}{\|x\|}.$$

Note. Throughout the rest of this section, the Banach dual of a Banach space E will be denoted with E^* . Instead, we reserve the notation E^\otimes for the order dual of an ordered vector space E , as discussed in Section 1.2. This distinction is necessary, since these two subspaces of the algebraic dual E' may happen to have nothing to do with each other. In fact, they can even exhibit a trivial intersection [43, Example 2.15(b)].

We can of course iterate the construction of the dual and consider the space E^{**} (the dual of the dual). As in the case of finite-dimensional spaces, there is a canonical linear embedding of E into E^{**} , since every vector $x \in E$ can be seen as acting on functionals $\varphi \in E^*$ in a rather obvious way via the formula $x(\varphi) = \langle \varphi, x \rangle$. Since this action is continuous with respect to the norm (1.1) on E^* , x belongs to $(E^*)^*$. When E is finite-dimensional we have $\dim E^{**} = \dim E^* = \dim E$, thus this embedding becomes an (isometric) isomorphism, and we can identify E^{**} with E . However, when E is infinite-dimensional it can happen that the image of E through the canonical embedding is a proper subspace of E^{**} . When this is the case we say that E is **non-reflexive**. While Hilbert spaces are well-known to be reflexive, it is important to bear in mind that non-reflexivity is not to be regarded as a pathological behaviour for Banach spaces. For instance, the standard sequence spaces we will examine in Examples 1.28 and 1.36 turn out to be non-reflexive.

With the concept of dual space at hand, we can state an intuitive yet fundamental result in the theory of Banach spaces.

³Although we stated the continuous linear extension lemma for functionals, a similar result holds for general linear applications $\phi : V \rightarrow F$, where F is any Banach space. In this case, the norm $\|\phi(x)\|$ displaces $|\langle \varphi, x \rangle|$ in expressions like (1.1).

Theorem 1.7 (Hahn-Banach separation theorem). [77, Theorem 3.4(b)]. *Let M, N be disjoint, convex subsets of a Banach space E . Assume M is compact and N is closed. Then there are $\varphi \in E^*$, $\alpha \in \mathbb{R}$, $\epsilon > 0$ such that*

$$\langle \varphi, x \rangle \leq \alpha < \alpha + \epsilon \leq \langle \varphi, y \rangle \quad \forall x \in M, y \in N. \quad (1.2)$$

The interpretation of the above result is quite intuitive: any point which does not belong to a closed convex set can be strictly separated from it by means of a hyperplane. As one realises quickly, there is really nothing special about the numbers α, ϵ we use to separate the two sets. In fact, in many cases one can rescale them as pleased by ‘zooming’ in or out with the probe functional φ . For the application we have in mind, we will need a simplified version of the Hahn-Banach separation theorem that deals with the case of N containing the origin or being invariant under multiplication by positive scalars. Let us state it and prove it now.

Corollary 1.8. *Let $M, N \subseteq E$ satisfy the hypotheses of Theorem 1.7.*

(a) *If additionally $0 \in N$, then in (1.2) we can choose $\alpha = -1$ and $\epsilon < 1$.*

(b) *If N is even invariant under multiplication by positive scalars, i.e. $N = tN$ for all $t > 0$, then in (1.2) we can take $\alpha = -1$ and $\epsilon = 1$, so that*

$$\langle \varphi, x \rangle \leq -1 \quad \text{and} \quad \langle \varphi, y \rangle \geq 0 \quad \forall x \in M, y \in N. \quad (1.3)$$

Proof. Let us start with claim (a). By taking $y = 0$ in (1.2), we see that $\alpha + \epsilon \leq 0$ and hence $\alpha + \frac{2}{3}\epsilon < 0$. Defining $\varphi' := \frac{\varphi}{|\alpha + \epsilon/3|}$ we get

$$\langle \varphi', x \rangle \leq \frac{\alpha}{|\alpha + \frac{\epsilon}{3}|} \leq -1 < \frac{\alpha + \frac{2}{3}\epsilon}{|\alpha + \frac{\epsilon}{3}|} =: -1 + \epsilon' < 0 \leq \langle \varphi', y \rangle \quad \forall x \in M, y \in N.$$

As for claim (b), if $N = tN$ for all $t > 0$ then $0 \in N$, because N is closed. Thus, we already know that we can take $\alpha = -1$. From (1.2) we see that $\langle \varphi, y \rangle \geq -1 + \epsilon$ for all $y \in N$. Since N is now closed under multiplication by positive scalars, for a fixed $y \in N$ we have also $t \langle \varphi, y \rangle \geq -1 + \epsilon$ for all $t > 0$. We conclude that $-1 + \epsilon \leq 0$ and that $\langle \varphi, y \rangle \geq 0$. This amounts to taking $\epsilon = 1$ from the start. \square

Another useful consequence of the Hahn-Banach separation theorem is a dual formula for the norm of a vector. We present it as a separated statement since we will make repeated use of it later.

Corollary 1.9. [77, §4.3]. *Let E be a Banach space and $x \in E$ a vector. Then*

$$\|x\| = \sup_{\varphi \in E^*, \|\varphi\|_* \leq 1} |\langle \varphi, x \rangle|. \quad (1.4)$$

Proof. On the one hand, by (1.1) we obtain $|\langle \varphi, x \rangle| \leq \|\varphi\|_* \|x\| \leq \|x\|$ for all $\varphi \in E^*$ such that $\|\varphi\|_* \leq 1$. On the other hand, pick $\epsilon > 0$ and apply Theorem 1.7 with $M = \{x\}$ and $N = \{y \in E : \|y\| \leq (1 - \epsilon)\|x\|\}$, which is evidently closed and convex. We obtain a functional $\varphi_1 \in E^*$ such that $\langle \varphi_1, x \rangle < \langle \varphi_1, y \rangle$ for all $y \in N$. As a matter of fact, we can consider instead $\varphi_2 := -\varphi_1$, which is such that $\langle \varphi_2, x \rangle > \langle \varphi_2, y \rangle$. By definition of dual norm (1.1), we can find $y \in N$ such that $\langle \varphi_2, y \rangle \geq (1 - \epsilon)\|\varphi_2\|_*(1 - \epsilon)\|x\| \geq (1 - 2\epsilon)\|\varphi_2\|_*\|x\|$. Putting all together, we see that $\varphi_3 := \frac{\varphi_2}{\|\varphi_2\|_*}$ satisfies $\|\varphi_3\|_* = 1$ and $\langle \varphi_3, x \rangle > (1 - 2\epsilon)\|x\|$. Since the construction can be repeated for all $\epsilon > 0$, we conclude that $\sup_{\varphi \in E^*, \|\varphi\|_* \leq 1} |\langle \varphi, x \rangle| \geq \|x\|$, yielding the claim. \square

1.4.3 Some classic results in Banach space theory

As we saw before, the assumption of completeness is crucial in order to establish some more profound facts that lie at the heart of the theory of Banach spaces. One such fact is the *uniform boundedness principle* (also called Banach-Steinhaus theorem), which we are now set to state in its simplest form. The most common way to prove it is via Baire's category theorem, but also direct proofs [83] or arguments based on non-standard analysis [84] are available.

Theorem 1.10 (Uniform boundedness principle). [77, §2.5] or [79, Theorem 1.6.9]. *Let E be a Banach space. If a family $\{\varphi_i\}_{i \in I}$ of continuous functionals $\varphi_i \in E^*$ satisfies*

$$\sup_{i \in I} |\langle \varphi_i, x \rangle| < \infty$$

for all $x \in E$, then

$$\sup_{i \in I} \|\varphi_i\|_* < \infty.$$

At this point, the reader might wonder, whether there is a real need to worry about the general concept of topological vector space if what we really care about here is Banach spaces, which carry a natural topology induced by their norms. As a matter of fact, although the norm topology really is the most intuitive topology one can endow a Banach space with, it is by no means the only one. In fact, the following alternative definitions will play an important role in what follows. For details, we refer the reader to [77, §§3.11 and 3.14] or [79, §§2.5 and 2.6].

Definition 1.11. *Let E be a Banach space with dual E^* . The initial topology on E generated by all continuous functionals $\varphi \in E^*$ is called **weak topology** and denoted by ω or by $\sigma(E, E^*)$. The initial topology on E^* generated by all vectors $x \in E$ (seen as functionals on E^*) is called **weak* topology** and denoted by ω^* or by $\sigma(E^*, E)$.*

As is easy to see, E and E^* equipped with these topologies are topological vector spaces in the sense of Definition 1.4. Moreover, it turns out that the above definitions are nontrivial as soon as we step out of the finite-dimensional realm. In fact, norm topology and weak topology coincide on a Banach space

E iff E is finite-dimensional [79, Proposition 2.5.13]. The same is true for the norm and weak* topologies on E^* [79, Corollary 2.6.3]. Topological notions pertaining to either of the above two topologies are referred to by juxtaposing weak- (or weakly) and weak*- in front of the relevant term. For instance, the concept of weak*-denseness will play a relevant role in what follows. Weak- and weak*-limits will be denoted by w -lim and w^* -lim, respectively.

Despite being different, weak topology and norm topology stem from the same underlying structure, i.e. the norm. Thanks to this common origin, there are indeed surprising connections between the two. An example of such a connection is provided by the following corollary to the Hahn-Banach separation theorem (Theorem 1.7).

Corollary 1.12. [77, §3.12] or [79, Theorem 2.5.16]. *Norm closure and weak closure coincide for convex subsets of a Banach space E . In particular, a convex subset of E is norm-closed iff it is weakly closed.*

Proof. Since the weak topology is coarser than the norm topology, the norm closure $\text{cl}(K)$ of any set $K \subseteq E$ is clearly contained inside its weak closure $\text{cl}_w(K)$. However, by Theorem 1.7 any $x \in E$ with $x \notin \text{cl}(K)$ can be separated strictly from the latter set (and in particular from K) by means of a functional $\varphi \in E^*$. This shows that there is a weak neighbourhood of x that does not contain points of K . Consequently, $x \notin \text{cl}_w(K)$, concluding the proof. \square

Let us now explore some basic facts about the weak* topology. Naturally, all functionals on the dual E^* that are induced by a vector $x \in E$ are considered to be continuous (this is in fact the definition of weak* topology). One could then wonder, whether these exhaust all the weak*-continuous functionals $E^* \rightarrow \mathbb{R}$. This is in fact the case, as the following lemma establishes.

Lemma 1.13. [77, p.66] or [79, Proposition 2.6.4]. *Let E be a Banach space. Then for a linear functional $f : E^* \rightarrow \mathbb{R}$ the following are equivalent:*

- (a) f is weak*-continuous;
- (b) its kernel $\ker(f)$ is weak*-closed;
- (c) $f(\cdot) = \langle \cdot, x \rangle$ for some $x \in E$.

Proof. The equivalence between (a) and (b) is standard, and stems from the linearity of f . If f is weak*-continuous then $f^{-1}(K)$ is weak*-closed for all closed sets $K \subseteq \mathbb{R}$, and in particular $f^{-1}(0) = \ker(f)$ is weak*-closed. Conversely, pick a net $(\varphi_\alpha)_\alpha \subseteq E^*$ such that $w^*\text{-lim}_\alpha \varphi_\alpha = 0$, and let us show that $\lim_\alpha f(\varphi_\alpha) = 0$. Up to considering subnets, we can assume that $f(\varphi_\alpha)$ converges to some real number r . Pick $\lambda \notin \ker(f)$, and construct the weak*-converging net formed by the functionals $\varphi'_\alpha := \varphi_\alpha - f(\varphi_\alpha) \frac{\lambda}{f(\lambda)}$. Since $\varphi'_\alpha \in \ker(f)$ for all α and $\ker(f)$ is weak*-closed,

$$\ker(f) \ni w^*\text{-lim}_\alpha \varphi'_\alpha = w^*\text{-lim}_\alpha \varphi_\alpha - \frac{\lambda}{f(\lambda)} \lim_\alpha f(\varphi_\alpha) = -\frac{\lambda}{f(\lambda)} r,$$

which is possible iff $r = 0$.

Now, since (c) clearly implies (a), we have just to show that if f is weak*-continuous then it acts as $f(\varphi) = \langle \varphi, x \rangle$ for some $x \in E$ and all $\varphi \in E^*$. By definition of continuity in topology, for all $\epsilon > 0$ the set $f^{-1}((-\epsilon, \epsilon)) \subseteq E^*$ is weak*-open. But weak*-open sets are unions of finite intersections of the form $\bigcap_{i=1}^n x_i^{-1}(U_i)$, where $x_i \in E$ is seen as a functional $x_i : E^* \rightarrow \mathbb{R}$, and $U_i \subseteq \mathbb{R}$ are open sets of real numbers (see the discussion in Section 1.3). Thus, since $0 \in f^{-1}((-\epsilon, \epsilon))$, we can pick a finite number of vectors $x_i \in E$ such that $0 \in \bigcap_{i=1}^n x_i^{-1}(U_i) \subseteq f^{-1}((-\epsilon, \epsilon))$. From $0 \in \bigcap_{i=1}^n x_i^{-1}(U_i)$ we deduce immediately $0 \in \bigcap_{i=1}^n U_i$. Take $\delta > 0$ such that $(-\delta, \delta) \subseteq \bigcap_{i=1}^n U_i$, which is possible since all sets U_i are open. Then for all $\varphi \in E^*$ one has

$$\begin{aligned} |\langle \varphi, x_i \rangle| &< \delta \quad \forall i = 1, \dots, n \\ \implies \varphi &\in \bigcap_{i=1}^n x_i^{-1}((-\delta, \delta)) \subseteq \bigcap_{i=1}^n x_i^{-1}(U_i) \subseteq f^{-1}((-\epsilon, \epsilon)) \\ \implies |f(\varphi)| &< \epsilon. \end{aligned}$$

As a consequence, when $\varphi \in \bigcap_{i=1}^n \ker x_i$ we have $|t \langle \varphi, x_i \rangle| < \delta$ for all i and $t \in \mathbb{R}$, and thus $|t f(\varphi)| < \epsilon$ for all $t \in \mathbb{R}$. This can happen only if $f(\varphi) = 0$. We have just shown that

$$\bigcap_{i=1}^n \ker x_i \subseteq \ker \varphi,$$

a condition that is well-known to be equivalent to $\varphi \in \text{span}\{x_1, \dots, x_n\}$ from elementary linear algebra. \square

Many strange features of infinite-dimensional Banach spaces stem from the fact that the unit ball $B = \{x \in E : \|x\| \leq 1\}$ is never compact in the norm topology.⁴ However, since the weak topology is coarser than the norm topology, it can happen that B is weakly compact. In fact, an important theorem says that this is the case iff the space is reflexive [79, Theorem 2.8.2]. One could expect similar exotic characterisations of the unit ball compactness to hold for the weak* topology. What comes as surprise now is that *the dual unit ball is always weak*-compact*. This statement is the content of the classic Banach-Alaoglu theorem.

Theorem 1.14 (Banach-Alaoglu). [77, §3.15] or [79, Theorem 2.6.18]. *Let E be a Banach space. The dual unit ball $B_* := \{\varphi \in E^* : \|\varphi\|_* \leq 1\}$ is weak*-compact.*

We can put in practise what we learnt until now and establish the following well-known result in Banach space theory. Since to the best of our knowledge the available demonstrations require some more advanced techniques, we

⁴In fact, with the help of the so-called *Riesz' lemma* it is not difficult to construct a sequence $(x_n)_n \subset B$ such that $\|x_n - x_m\| \geq 1/2$ for all $n \neq m$. Such a sequence does not admit any Cauchy (hence, convergent) subsequence, therefore B can not be compact.

include an elementary proof that makes instructive use of both the Banach-Alaoglu theorem and the uniform boundedness principle.

Proposition 1.15. [79, Corollary 2.7.12]. *Let $W \subseteq E^*$ be a subspace of the dual E^* of a Banach space E . Denote by $B_* = \{\varphi \in E^* : \|\varphi\|_* \leq 1\}$ the dual unit ball. Then W is weak*-closed iff so is $W \cap B_*$.*

Proof. Since B_* is weak*-compact by Theorem 1.14 and hence also weak*-closed, if W is weak*-closed then so is $W \cap B_*$. Conversely, assume that $W \cap B_*$ is weak*-closed, and let us show that the same holds for the whole W . First of all, since closed subsets of compact sets are again compact, we see that $W \cap B_*$ must in fact be weak*-compact (again, a consequence of Theorem 1.14). Then, take a net $(\varphi_\alpha)_\alpha \subseteq W$ that converges to $\varphi \in E^*$ in the weak* topology, i.e. $w^*\text{-lim}_\alpha \varphi_\alpha = \varphi$. If we prove that $\varphi \in W$, we are done. Rescale the functionals φ_α by defining $\lambda_\alpha := \frac{\varphi_\alpha}{\|\varphi_\alpha\|_*}$. Now, since $(\lambda_\alpha)_\alpha \subseteq W \cap B_*$ and the latter set has been shown to be weak*-compact, the characterisation of compactness in terms of nets (see Section 1.3) guarantees that there is a subnet $(\lambda_\beta)_\beta$ of $(\lambda_\alpha)_\alpha$ such that $w^*\text{-lim}_\beta \lambda_\beta = \lambda$ for some $\lambda \in W \cap B_*$. Taking subnets of converging nets does not change the limit, so we still have $w^*\text{-lim}_\beta \varphi_\beta = \varphi$.

Now, since $\lim_\beta \langle \varphi_\beta, x \rangle = \langle \varphi, x \rangle$ and in particular $\sup_\beta |\langle \varphi_\beta, x \rangle| < \infty$ for all $x \in E$, the uniform boundedness principle (Theorem 1.10) implies that there is $k \in \mathbb{R}$ such that $\|\varphi_\beta\|_* \leq k$ for all β , or in other words that the net of real numbers $(\|\varphi_\beta\|_*)_\beta \subseteq \mathbb{R}$ is bounded. We can then extract from $(\varphi_\beta)_\beta$ a subnet $(\varphi_\gamma)_\gamma$ such that $\lim_\gamma \|\varphi_\gamma\|_* = r \in \mathbb{R}$. Of course, $(\lambda_\gamma)_\gamma$ converges again to λ in the weak* topology, or $w^*\text{-lim}_\gamma \lambda_\gamma = \lambda$. Going back to the original net $(\varphi_\alpha)_\alpha$, which still satisfies $w^*\text{-lim}_\gamma \varphi_\gamma = \varphi$, we see that we must have

$$\begin{aligned} \varphi &= w^*\text{-lim}_\gamma \varphi_\gamma \\ &= w^*\text{-lim}_\gamma \|\varphi_\gamma\|_* \frac{\varphi_\gamma}{\|\varphi_\gamma\|_*} \\ &= w^*\text{-lim}_\gamma \|\varphi_\gamma\|_* \lambda_\gamma \\ &= (\lim_\gamma \|\varphi_\gamma\|_*) (w^*\text{-lim}_\gamma \lambda_\gamma) \\ &= r\lambda, \end{aligned}$$

and since $\lambda \in W$ this implies that also $\varphi \in W$. □

1.5 Ordered Banach spaces

Throughout Section 1.2, we discussed some natural ways to equip a vector space with an ordering. Instead, Section 1.3 was devoted to the study of vector spaces endowed with topological structures, and in Section 1.4 we further specialised to Banach spaces. Now, we will see how these notions interplay, giving rise to the beautiful theory of ordered topological vector spaces. For a complete exposition of the subject, we refer the reader to [42, V] or to the monographs [43, 44]. A more concise introduction can be found also in [81, Appendix A]. Part of our

presentation is taken from the lecture notes [76, ‘Duality of cones in locally convex spaces’], but we try to be as self-contained as possible and to report all the proofs of the results we need.

In what follows, we are interested in ordered vector spaces that are also Banach spaces. Going back to the definitions, we see that we enforced no connection whatsoever between the order and the Banach structure. Hence, the reader will not be surprised by the fact that we need to do so before starting to prove nontrivial results. For an ordered Banach space E , an example of an outstanding problem in this context concerns the relation between its order dual E^\otimes and its Banach dual E^* , as defined in Sections 1.2 and 1.4. Up to now, E^\otimes and E^* are two a priori different subspaces of the algebraic dual E' , and we already mentioned that in fact they can happen to have trivial intersection [43, Example 2.15(b)]. In light of this, the following question arises: *under what hypotheses can we show a relation between order and Banach dual? For instance, when is it the case that every positive functional is automatically continuous, i.e. $E^\otimes \subseteq E^*$?*

We can of course take a pragmatic approach and restrict ourselves to work inside the Banach dual from the start. Observe that the cone E_+^* formed by all positive *and continuous* functionals turns E^* into an ordered Banach space. We observe immediately that E_+^* is always weak*-closed, for any net $(\varphi_\alpha)_\alpha \subseteq E_+^*$ such that $w^*\text{-lim}_\alpha \varphi_\alpha = \varphi \in E^*$ satisfies also $\langle \varphi, x \rangle = \lim_\alpha \langle \varphi_\alpha, x \rangle \geq 0$ for any given $x \in E_+$, implying that $\varphi \in E_+^*$.

Note. We collected a few different ways to take duals of spaces and cones, so perhaps a quick recap is helpful now.

- E vector space: E' is the algebraic dual, i.e. the space formed by all possible linear functionals $E \rightarrow \mathbb{R}$.
- E, E_+ ordered vector space: E_+^\otimes denotes the set of all positive functionals inside the algebraic dual E' .
- E, E_+ ordered vector space: $E^\otimes = E_+^\otimes - E_+^\otimes$ is the order dual.
- E Banach space: E^* is the Banach dual, i.e. the space formed by all continuous (bounded) linear functionals.
- E, E_+ ordered Banach space: E_+^* denotes the cone of all positive and continuous linear functionals on E . This in turn defines an ordering on E^* .

1.5.1 Duality of cones

We start by asking ourselves: what are the possible requirements to impose on an ordered Banach space E with the purpose of connecting Banach and order structure? A very natural choice is the closedness of the positive cone E_+ with respect to the norm topology (for more general topologies, this leads to the definition of *ordered topological vector spaces*). Other central notions in the

theory of ordered Banach spaces are that of *normal cone* and *strict \mathcal{B} -cone*. In order to discuss them, we need some preliminary definitions.

Definition 1.16. Let $M \subseteq E$ be a subset of an ordered vector space E . Its *saturated hull* $[M]$ and *convex kernel* $]M[$ are defined as

$$[M] := (M + E_+) \cap (M - E_+), \quad (1.5)$$

$$]M[:= \text{co}((M \cap E_+) \cup -(M \cap E_+)). \quad (1.6)$$

Alternatively, one can think of the saturated hull as the union of all the intervals whose extreme points lie in M , in formula

$$[M] = \bigcup_{x, y \in M, x \leq y} [x, y]. \quad (1.7)$$

Clearly, saturated hulls and convex kernels are always convex sets, and moreover contain the origin as long as M does so. However, neither of the two sets $[M]$ or $]M[$ is a priori guaranteed to be closed, not even when both M and E_+ are such.

Definition 1.17. Let $\alpha, \beta \geq 1$ be real numbers. The positive cone E_+ of an ordered Banach space E with unit ball $B := \{x \in E : \|x\| \leq 1\}$ is said to be *α -normal* if $[B] \subseteq \alpha B$, and *β -generating* if $B \subseteq \beta]B[$.

The above definition can be cast into the following alternative form: E_+ is α -normal if whenever $\|x\|, \|y\| \leq 1$ and $x \leq z \leq y$, one has $\|z\| \leq \alpha$, while it is β -generating if whenever $\|x\| \leq 1$ there are $x_{\pm} \geq 0$ such that $x = x_+ - x_-$ and $\|x_{\pm}\| \leq \beta$. In particular, any β -generating cone is spanning. In the existing literature, a cone is usually called **normal** if it is α -normal for some $\alpha \in \mathbb{R}$, and a **strict \mathcal{B} -cone** if it is β -generating for some $\beta \in \mathbb{R}$. Our immediate goal is to gain a better understanding the interplay between the two concepts of α -normal and β -generating cones, which turn out to be dual to each other in a precise sense. Let us first give another definition.

Definition 1.18. Let E be a Banach space with dual E^* . The *polars* of subsets $M \subseteq E$ and $N \subseteq E^*$ are defined as

$$M^\circ := \{\varphi \in E^* : \langle \varphi, x \rangle \leq 1 \ \forall x \in M\}, \quad (1.8)$$

$$N^\circ := \{x \in E : \langle \varphi, x \rangle \leq 1 \ \forall \varphi \in N\}. \quad (1.9)$$

Naturally, starting from $M \subseteq E$ one can apply two times the polar operation and consider $M^{\circ\circ} := (M^\circ)^\circ$. We show in Lemma A.1 of Appendix A that if $M \subseteq E$ and $N \subseteq E^*$ are convex and contain the origin then

$$M^{\circ\circ} = \text{cl}(M), \quad (1.10)$$

$$N^{\circ\circ} = \text{cl}_{w^*}(N). \quad (1.11)$$

Polar sets behave well when it comes to taking saturated hulls or convex kernels. Namely, let E be an ordered Banach space with closed positive cone E_+ ,

and let its Banach dual E^* be ordered by the cone E_+^* of positive continuous functionals. We denote by B and B° the unit balls in E and E^* , respectively (with the notation of Theorem 1.14, $B^\circ = B_*$). Then we have the following computation rules:

$$[B]^\circ =]B^\circ[, \quad (1.12)$$

$$[B^\circ]^\circ = \text{cl}(]B[), \quad (1.13)$$

where cl denotes norm closure. In Appendix A we report the proof of these formulae, adapted from the existing literature, see for instance [42, IV §1.5] and [85]. Here, we limit ourselves to the observation that if $]B^\circ[$ is always weak*-compact (in particular, closed), which can be seen as follows: (1) $B^\circ \cap E_+^*$ is the intersection of a weak*-closed set (i.e. E_+^*) with a weak*-compact set (i.e. B° , see Theorem 1.14), hence it is weak*-compact; (2) convex hulls of the form $\text{co}(M \cup N)$, where both M and N are compact, are themselves compact, because for all nets $(x_\alpha)_\alpha = (p_\alpha y_\alpha + (1 - p_\alpha)z_\alpha)_\alpha \subseteq \text{co}(M \cup N)$ we can successively extract a convergent subnet of $(y_\alpha)_\alpha \subseteq M$, of $(z_\alpha)_\alpha \subseteq N$, and finally of $(p_\alpha)_\alpha \subseteq [0, 1]$.

With these tools in our hand, we can prove a duality result that will turn out to be very useful to develop the theory further. We need a preliminary lemma, which was stated for the first time in [86] in a slightly different form. In [87] it is reported without proof, while Ellis [45, Lemma 7] gives a complete argument, presented in a slightly simplified fashion in [76, p.18].

Lemma 1.19 (Tukey). *Let $M, N \subseteq E$ be closed, convex, and bounded subsets of E . Then*

$$\text{int}(\text{cl}(\text{co}(M \cup N))) \subseteq \text{co}(M \cup N). \quad (1.14)$$

Here interior and closure are taken with respect to the norm topology.

Proof. Let $x \in \text{int}(\text{cl}(\text{co}(M \cup N)))$, i.e. assume that there is $\delta > 0$ such that $\|x - y\| \leq 2\delta \Rightarrow y \in \text{int}(\text{cl}(\text{co}(M \cup N)))$ for all $y \in E$. Pick $\tilde{y}_1 \in E$ such that $\|x - \tilde{y}_1\| \leq \delta$. Clearly, $\tilde{y}_1 \in \text{int}(\text{cl}(\text{co}(M \cup N)))$. Since in particular $\tilde{y}_1 \in \text{cl}(\text{co}(M \cup N))$, we can choose $y_1 \in \text{co}(M \cup N)$ such that $\|y_1 - \tilde{y}_1\| \leq \delta$. Now, define $z_1 := 2x - y_1$. One has $\|x - z_1\| = \|x - y_1\| \leq \|x - \tilde{y}_1\| + \|\tilde{y}_1 - y_1\| \leq \delta + \delta = 2\delta$, from which we deduce that $z_1 \in \text{int}(\text{cl}(\text{co}(M \cup N)))$.

Until now, we have shown how to rewrite $x \in \text{int}(\text{cl}(\text{co}(M \cup N)))$ as an average $x = \frac{1}{2}(y_1 + z_1)$, where $y_1 \in \text{co}(M \cup N)$ and $z_1 \in \text{int}(\text{cl}(\text{co}(M \cup N)))$. We can repeat this procedure with z_1 in place of x , and write $z_1 = \frac{1}{2}(y_2 + z_2)$ with $y_2 \in \text{co}(M \cup N)$ and $z_2 \in \text{int}(\text{cl}(\text{co}(M \cup N)))$. Going back to x , this gives us the representation $x = \frac{1}{2}y_1 + \frac{1}{4}y_2 + \frac{1}{4}z_2$. Iterating the process, we end up with an identity of the form $x = \sum_{n=1}^{\infty} 2^{-n}y_n$, where the series converges in norm. This latter fact can be seen as follows: since M is bounded, the same is true for $\text{int}(\text{cl}(\text{co}(M \cup N)))$, thus the remainder term must vanish in norm asymptotically.

Now, since $y_n \in \text{co}(M \cup N)$ for all n and both M and N are convex, we can write $y_n = p_n a_n + (1 - p_n)b_n$ with $p_n \in [0, 1]$, $a_n \in M$ and $b_n \in N$. Define

$p := \sum_{n=1}^{\infty} 2^{-n} p_n$, which is legitimate because the series on the right-hand side converges. We find

$$\begin{aligned} x &= \lim_n \sum_{k=1}^n 2^{-k} y_k \\ &= \lim_n \left(\sum_{k=1}^n 2^{-k} p_k a_k + \sum_{k=1}^n 2^{-k} (1 - p_k) b_k \right) \\ &= \lim_n \sum_{k=1}^n 2^{-k} p_k a_k + \lim_n \sum_{k=1}^n 2^{-k} (1 - p_k) b_k \\ &= pa + (1 - p)b, \end{aligned}$$

where $a := \lim_n \sum_{k=1}^n \frac{p_k}{2^k p} a_k \in M$, since the sequence of partial sums is clearly Cauchy (because M is bounded), M is closed (hence complete) and convex. This proves that $x \in \text{co}(M \cup N)$, as claimed. \square

Let us state as a separate lemma the following elementary fact, which will turn out to be useful multiple times in what follows.

Lemma 1.20. *Let $M \subseteq E$ be a norm-bounded subset of a Banach space E . Then*

$$\bigcap_{\epsilon > 0} (1 + \epsilon)M \subseteq \text{cl}(M). \quad (1.15)$$

Proof. Let $x \in \bigcap_{\epsilon > 0} (1 + \epsilon)M$. Then for all $n \in \mathbb{N}$ with $n \geq 1$ there is some $x_n \in M$ such that $x = (1 + \frac{1}{n})x_n$. The norm-boundedness of M implies that the sequence $(x_n)_n \subseteq M$ converges to x in norm, since $\lim_n \|x - x_n\| = \lim_n \frac{1}{n} \|x_n\| = 0$. \square

Now we are ready to present the following result, arguably a cornerstone of the theory of ordered Banach spaces. The two claims it is made of are due to Grosberg and Krein [88] and to Ellis [45]. Here, we report the simplified proof found in [85] and thoroughly discussed in [76, p.18].

Theorem 1.21 (Grosberg-Krein-Ellis). *Let E be an ordered Banach space with closed positive cone. Then:*

- (a) E_+ is α -normal iff E_+^* is α -generating; and
- (b) E_+^* is α -normal iff E_+ is $(\alpha + \epsilon)$ -generating for all $\epsilon > 0$.

Proof. Let us start with statement (a). By definition, E_+ is α -normal iff $[B] \subseteq \alpha B$. We claim that this condition is in turn equivalent to $[B]^\circ \supseteq \frac{1}{\alpha} B^\circ$. That the former implies the latter follows from the general rules $M \subseteq N \Rightarrow N^\circ \subseteq M^\circ$ and $(\alpha M)^\circ = \frac{1}{\alpha} M^\circ$ (where $\alpha > 0$). Conversely, if $\frac{1}{\alpha} B^\circ \subseteq [B]^\circ$ by applying the same two rules we obtain $\alpha B^{\circ\circ} \supseteq [B]^{\circ\circ}$ and hence

$$[B] \subseteq \text{cl}[B] = [B]^{\circ\circ} \subseteq \alpha B^{\circ\circ} = \alpha B,$$

where we applied twice (1.10), the first time with $M = [B]$, and the second with $M = B$. Now, we can use the computation rule (1.12) to deduce that E_+ is α -normal iff $\frac{1}{\alpha}B^\circ \subseteq [B^\circ] =]B^\circ[$, i.e. iff E_+^* is α -generating.

Proving claim (b) requires a bit more care. By definition, E_+^* is α -normal iff $[B^\circ] \subseteq \alpha B^\circ$. Using the fact that $[B^\circ]$ is weak*-closed (see the discussion after (1.13)) and hence $[B^\circ]^{\circ\circ} = [B^\circ]$ by (1.11), plus the usual identity $B^{\circ\circ} = B$, one sees that this is in turn equivalent to $[B^\circ]^\circ \supseteq \frac{1}{\alpha}B$. Applying (1.13), this yields the equivalent condition $B \subseteq \alpha \text{cl}(]B[)$. Applying (1.15) to the norm-bounded set $]B[$, we see immediately that if E_+ is $(\alpha + \epsilon)$ -generating for all $\epsilon > 0$ then E_+^* is α -normal.

To show the converse, assume that $B \subseteq \alpha \text{cl}(]B[)$, and let us prove that $B \subseteq (\alpha + \epsilon)]B[$ for all $\epsilon > 0$. By hypothesis,

$$\text{int}(\alpha^{-1}B) \subseteq \text{int}(\text{cl}(]B[)) \subseteq]B[,$$

where the last step follows from Lemma 1.19, which is applicable since $]B[= \text{co}((B \cap E_+) \cup -(B \cap E_+))$ is the convex hull of the union of two closed, convex and bounded sets. Then, for all $\epsilon > 0$ observe that $(\alpha + \epsilon)^{-1}B \subseteq \text{int}(\alpha^{-1}B) \subseteq]B[$, which amounts to saying that E_+ is $(\alpha + \epsilon)$ -generating. \square

1.5.2 Order dual vs Banach dual

We started this Section by posing the question of the interplay between order and Banach duals of an ordered Banach space. The question is partially settled by the following results.

Proposition 1.22. [42, V §3.5] or [43, Corollary 1.28]. *Let E be an ordered Banach space with closed and spanning positive cone E_+ . Then E_+ is a strict \mathcal{B} -cone, i.e. it is β -generating for some $\beta \in \mathbb{R}$.*

Proof. Define $K := \text{cl}(]B[)$. We see immediately that K is closed and convex. Furthermore, for any given $x \in E$ we can find a decomposition $x = x_+ - x_-$ with $x_\pm \geq 0$ by hypothesis. From this we deduce $x \in (\|x_+\| + \|x_-\|)]B[$ and thus $E = \bigcup_{n \in \mathbb{N}} n]B[= \bigcup_{n \in \mathbb{N}} nK$. Then, Lemma 1.6 implies that K contains a neighbourhood of the origin, i.e. there is $\beta > 1$ such that $\frac{1}{\beta-1}B \subseteq K$. As in the proof of Theorem 1.21, thanks to (1.13) we obtain

$$(\beta - 1)B^\circ = \left(\frac{1}{\beta - 1}B \right)^\circ \supseteq (\text{cl}(]B[))^\circ = ([B^\circ]^\circ)^\circ = [B^\circ],$$

implying that E_+^* is $(\beta - 1)$ -normal. By Theorem 1.21(b), we conclude that E_+ is β -generating, as claimed. \square

Proposition 1.23. [42, V §5.5] or [43, Corollary 2.17(b)]. *Let E be an ordered Banach space with closed and spanning positive cone. Then every positive linear functional on E is also continuous, in formula $E^\circledast \subseteq E^*$ (i.e. the order dual is contained inside the Banach dual).*

Proof. Assume by contradiction that there is a positive functional $\varphi \in E_+^*$ that is not continuous. Continuity and boundedness are equivalent concepts for linear functionals, hence φ will be unbounded on $E = E_+ - E_+$. This is possible only if φ is unbounded on the positive cone E_+ itself. In fact, thanks to Proposition 1.22, we know that E_+ is β -generating for some $\beta \geq 1$, i.e. that for all $x \in E$ there exists a decomposition $x = x_+ - x_-$ such that $x_{\pm} \geq 0$ and $\|x_{\pm}\| \leq \beta\|x\|$. If φ were bounded on E_+ , i.e. $|\langle \varphi, x_{\pm} \rangle| \leq k\|x_{\pm}\|$ for all $x_{\pm} \geq 0$, we would find $|\langle \varphi, x \rangle| \leq |\langle \varphi, x_+ \rangle| + |\langle \varphi, x_- \rangle| \leq k(\|x_+\| + \|x_-\|) \leq 2k\beta\|x\|$, contrary to the assumption that φ is unbounded.

Therefore, we can construct a sequence $(x_n)_n \subset E_+$ with $\|x_n\| \leq 2^{-n}$ and $\langle \varphi, x_n \rangle \geq 1$ for all $n \in \mathbb{N}$. The partial sums $y_n := \sum_{k=0}^n x_k$ then form an increasing Cauchy sequence, which will converge to some $y = \lim_n y_n \in E_+$ (since E_+ is closed). From $y \geq y_n$ we get $\langle \varphi, y \rangle \geq \langle \varphi, y_n \rangle = \sum_{k=0}^n \langle \varphi, x_k \rangle \geq n + 1$, which yields a contradiction once we take the limit on n . \square

Proposition 1.24. [42, V §3.2]. *Let E be a Banach space with normal positive cone. Then the cone formed by all bounded positive functionals generates the whole Banach dual E^* . In particular, if E_+ is normal, closed, and spanning, then we have $E^{\otimes} = E^*$.*

Proof. We have only to show that if E_+ is normal then every bounded functional $\varphi \in E^*$ admits a representation of the form $\varphi = \lambda - \mu$, where λ, μ are positive and bounded. To this purpose, let us augment the Banach space E by taking the direct sum with a single copy of the real line. On the resulting space $\mathbb{R} \oplus E$ one can impose several different norms, but the choice of the norm is not going to affect the proof we are presenting. For the sake of the argument, one can consider $\mathbb{R} \oplus E$ to be normed by $\|(t, x)\| := |t| + \|x\|$.

Now, for all $x \geq 0$ define the function

$$p(x) := \sup\{\langle \varphi, y \rangle : 0 \leq y \leq x\}.$$

We see immediately that: (a) $p(\lambda x) = \lambda p(x)$ for all $\lambda \geq 0$; (b) $p(x + y) \geq p(x) + p(y)$. This in turn shows that the set

$$C := \{(t, x) : t \in \mathbb{R}, x \in E_+, 0 \leq t \leq p(x)\} \subseteq \mathbb{R} \oplus E$$

is in fact a cone. Now, since E_+ is normal we have that $\bigcup_{z_1, z_2 \in B, z_1 \leq z_2} [z_1, z_2] = [B] \subseteq \alpha B$ for some $\alpha \in \mathbb{R}$, where B is the unit ball of E . Therefore, any $y \in [0, x]$ satisfies $\frac{y}{\|x\|} \in [0, \frac{x}{\|x\|}] \subseteq [B] \subseteq \alpha B$, i.e. $\|y\| \leq \alpha\|x\|$. Consequently, we see that $p(x) \leq \alpha\|x\|$ holds for all $x \in E_+$. In turn, this implies that $(1, 0) \notin \text{cl}(C)$ (the closure of C in the norm topology), as otherwise we could construct a sequence $(x_n)_n \subset E_+$ such that $\lim_n \|x_n\| = 0$ but $\liminf_n p(x_n) \geq 1$, which is in contradiction with $p(x) \leq \alpha\|x\|$ for all $x \in E_+$.

Since the closed convex set $\text{cl}(C)$ is easily seen to satisfy $\text{cl}(C) = t \text{cl}(C)$ for all $t > 0$, and moreover $(1, 0) \notin \text{cl}(C)$, Corollary 1.8 (b) allows us to construct a functional $h \in (\mathbb{R} \oplus E)^*$ that is positive on $\text{cl}(C)$ and such that $h(1, 0) = -1$. Any bounded functional h on $\mathbb{R} \oplus E$ is of the form $h(s, y) := as + \langle \lambda, y \rangle$ for

some $a \in \mathbb{R}$ and $\lambda \in E^*$. In our case, we see immediately that $a = -1$ and that $-p(x) + \langle \lambda, x \rangle \geq 0$ holds for all $x \in E_+$. In particular, λ is a positive functional and $\langle \lambda, x \rangle \geq \langle \varphi, x \rangle$ for all $x \in E_+$, from which the representation $\varphi = \lambda - (\lambda - \varphi)$ of φ as the difference of two positive bounded functionals follows. \square

1.6 Order unit and base norm spaces

Throughout this section, we make another step toward increasing specialisation, and study two particular kinds of ordered Banach spaces, namely so-called *order unit* and *base norm* spaces. For details, we refer the reader to [44, III] and to [76]. For the sake of completeness, we will also provide more specific references to the original papers when appropriate. Unlike most of the content of Sections 1.3-1.5, which becomes almost trivial when we deal with (complete) finite-dimensional vector spaces, the assumptions we are going to make here concern the shape of the unit ball and therefore restrict significantly the analysis even in finite dimension.

Roughly speaking, an order unit space is a Banach space whose unit ball is made of two copies of the same cone, facing in opposite directions and glued together by their bases (Figure 1.2). The original definition of order unit spaces in the context of ordered Banach spaces seems to go back to Ellis [45]. The unit ball of a base norm space, instead, is generated via convex hull by two planar faces that are opposite to each other (Figure 1.3). These kind of spaces seems to have been considered for the first time by Edwards [47]. A remarkable feature of these kinds of spaces is that they are dual to each other. In fact, this is the case for the spaces depicted in Figures 1.2 and 1.3. The purpose of this section is to define rigorously these concepts in the context of ordered Banach spaces, and present a duality theorem whose proof has been the subject of works by Krein [89], Edwards [47] and Ellis [45].

1.6.1 Order unit spaces

Throughout this subsection we examine and discuss the notion of *order unit space* [45]. We start by introducing an important yet very natural requirement on vector orderings.

Definition 1.25. *An ordered vector space E is called **Archimedean** if $x \leq 0$ whenever there is $y \in E$ such that $nx \leq y$ for all $n \in \mathbb{N}$.*

Not all ordered vector spaces are Archimedean, not even in finite dimension. An example will help to understand why. Consider $E = \mathbb{R}^2$ ordered by the cone $E_+ = \{(x, y) : x > 0 \text{ and } y > 0\} \cup \{(0, 0)\}$. The vectors $x = (-1, 0)$, $y = (0, 1)$ satisfy $nx \leq y$ for all $n \in \mathbb{N}$, nevertheless $x \not\leq 0$. Clearly, the construction of this counterexample was possible because the E_+ we chose was not closed. In fact, a finite-dimensional ordered vector space is Archimedean iff its positive cone is closed [42, V §4].

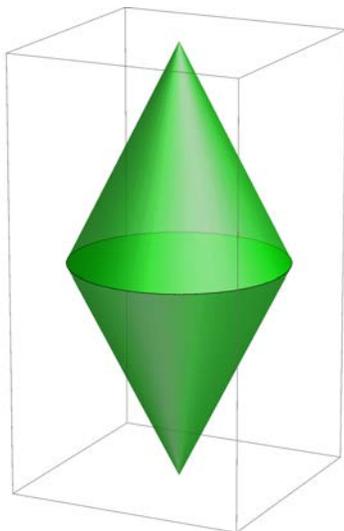


Figure 1.2: The unit ball of an order unit space, namely \mathbb{R}^3 with the norm $\|(x, y, z)\| = 2\sqrt{x^2 + y^2} + |z|$.

For the applications we have in mind, many ordered vector spaces come with a special element that is singled out by some physical axiom. Then, the following definition becomes relevant.

Definition 1.26. An *order unit* of an ordered vector space E is a positive element $u \geq 0$ such that $E = \bigcup_{n \in \mathbb{N}} n[-u, u]$, where $[-u, u] := \{x \in E : -u \leq x \leq u\}$.

Again, the finite-dimensional case helps us to understand better the idea behind the definition. In fact, when $\dim E < \infty$, order units are all those elements that belong to the interior of E_+ . An order unit is useful because it allows us to construct a norm on E . We are ready to give the following definition.

Definition 1.27. An Archimedean ordered vector space E with an order unit u is called an *order unit space* if it is complete when equipped with the norm

$$p_u(x) := \inf \{t > 0 : x \in t[-u, u]\}. \quad (1.16)$$

Note that the Archimedean axiom ensures that (1.16) is a norm and not only a semi-norm. In fact, $p_u(x) = 0$ implies that $x \in \frac{1}{n}[-u, u]$ for all $n \in \mathbb{N}$. If E is Archimedean, this is possible only if $x = 0$.

Example 1.28. A paradigmatic example of an order unit space is the sequence space l^∞ [43, §1.6], defined by

$$l^\infty := \left\{ (a_n)_{n \in \mathbb{N}} \subset \mathbb{R} : \sup_{n \in \mathbb{N}} |a_n| < \infty \right\}. \quad (1.17)$$

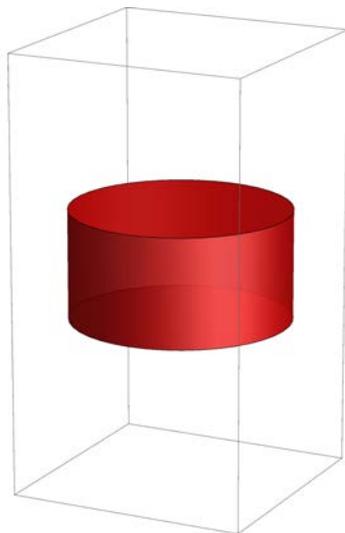


Figure 1.3: The unit ball of a base norm space, namely \mathbb{R}^3 with the norm $\|(x, y, z)\| = \max\{\sqrt{x^2 + y^2}, 2|z|\}$.

It is easy to verify that l^∞ becomes complete when equipped with the norm

$$\|a\|_\infty := \sup_{n \in \mathbb{N}} |a_n|. \quad (1.18)$$

Furthermore, when ordered by the (closed) cone $C := \{(a_n)_n : a_n \geq 0 \forall n\}$, the norm (1.18) turns out to be induced by the order unit $u := (1)_n$ (the constant sequence). Therefore, l^∞ is an order unit space.

Since order unit spaces are very special examples of ordered Banach spaces, they enjoy many of the properties we discussed in Section 1.5 in relation to the interplay between norm and order structure. An example is given by the following result, whose proof and corollaries can be deduced straight from the observations in [76, ‘Order unit and base norm spaces’].

Proposition 1.29. *If E is an order unit space, then the positive cone E_+ is spanning, closed in the norm topology, 1-normal and 2-generating. In particular, order dual and Banach dual coincide, i.e. $E^\circledast = E^*$.*

Proof. Clearly, the existence of an order unit u already implies that the positive cone is spanning. Let us prove that E_+ is closed. If $(x_n)_n \subset E_+$ is a sequence of positive vectors that converges in norm to $x \in E$, then for all $n \in \mathbb{N}$ we will have $\|x - x_m\| \leq \frac{1}{n}$ eventually in $m \in \mathbb{N}$. Thus, $x - x_m \in \frac{1}{n}[-u, u]$ and in particular $-x \leq \frac{1}{n}u - x_m \leq \frac{1}{n}u$, i.e. $-nx \leq u$. Since this happens for all $n \in \mathbb{N}$ and E is Archimedean, we conclude that $-x \leq 0$ and hence that $x \geq 0$, as claimed. By Proposition 1.23, this implies also $E^\circledast \subseteq E^*$.

To show that E_+ is 1-normal, we employ the formula (1.7) for the saturated hull $[B]$ of the unit ball $B = [-u, u]$:

$$[B] = \bigcup_{x, y \in [-u, u], x \leq y} [x, y] = [-u, u] = B.$$

Checking that E_+ is 2-generating is also easy, for any $x \in [-u, u]$ can be decomposed as $x = y_+ + y_- = 2\left(\frac{1}{2}y_+ + \frac{1}{2}y_-\right)$ with $y_{\pm} := \frac{x \pm u}{2}$. Since $y_+ \in [0, u] = [-u, u] \cap E_+ = B \cap E_+$ and $y_- \in [-u, 0] = -B \cap E_+$, we get $x \in 2\operatorname{co}((B \cap E_+) \cup -(B \cap E_+)) =]B[$, and since x was generic $B \subseteq]B[$. Finally, the equality $E^{\otimes} = E^*$ between order dual and Banach dual follows from Proposition 1.24. \square

Corollary 1.30. *In an order unit space with order unit space u , the unit ball coincides with $[-u, u]$.*

Proof. From the definition (1.16) it is clear that $[-u, u] \subseteq B$, where B is the unit ball. On the other hand, if $x \in B$ then for all $\epsilon > 0$ one has $x \in (1 + \epsilon)[-u, u]$, and consequently $x \in \bigcap_{\epsilon > 0} (1 + \epsilon)[-u, u] \subseteq \operatorname{cl}([-u, u]) = [-u, u]$, where we applied (1.15), and the last inclusion holds because the positive cone is closed hence so are all intervals. \square

For the sake of further developments of the theory, it is useful to have a criterion that allows us to establish that a space is an order unit space while checking as few conditions as possible. The following simple observation that partially reverses one of the claims of Proposition 1.29 will come in handy.

Lemma 1.31. *Let E be an ordered Banach space whose norm is given by (1.16) for some $u \geq 0$. If the positive cone E_+ is closed, then E is an order unit space.*

Proof. We have just to check that E is Archimedean. This is readily verified, since $-nx \leq y$ for all $n \in \mathbb{N}$ implies that $\frac{1}{n}y - x \in E_+$, which in turn yields $-x \in \operatorname{cl}(E_+) = E_+$ upon taking the limit $n \rightarrow \infty$. From $-x \geq 0$ we deduce $x \leq 0$, as claimed. \square

1.6.2 Base norm spaces

There is another way to induce a norm on an ordered vector space. In order to discuss this second alternative, we need a definition.

Definition 1.32. *Let E be an ordered vector space. A non-empty convex subset K is called a **base** of the positive cone E_+ if for all $x \geq 0$ there is a unique $t \geq 0$ such that $x \in tK$.*

Before we delve into the definition of base norm space, it is useful to have in hand a lemma characterising the concept of base in a way that lends itself better to computations. Let us start by introducing some terminology. As before, let E be an ordered vector space. A positive linear functional $\varphi \in E_+^* \subseteq E^{\otimes}$ in

the order dual is called **strictly positive** if $\langle \varphi, x \rangle > 0$ whenever $x \in E_+$ with $x \neq 0$. In finite dimension, φ is strictly positive iff it belongs to the relative interior of E_+^* . The following elementary result was originally found in [45, 47].

Lemma 1.33. *Let E be an ordered vector space with positive cone E_+ . A base K for E_+ (if it exists) is always of the form*

$$K = \{x \geq 0 : \langle u, x \rangle = 1\}, \quad (1.19)$$

where $u \in E_+^* \subset E^{\otimes}$ is strictly positive.

Proof. Sets defined by (1.19) are clearly bases, because for all $x \in E_+$, $x \in tK$ happens iff $t = \langle u, x \rangle$. Conversely, let K be a base for E_+ . Then let us show that the affine hull $\text{aff}(K)$ of K does not contain the origin. In fact, if $\sum_{i=1}^n \lambda_i a_i = 0$ for some $a_i \in K$ and coefficients λ_i summing to 1, then constructing $a := \frac{1}{n} \sum_{i=1}^n a_i \in K$ we obtain

$$(1-t)a = (1-t)\frac{1}{n} \sum_{i=1}^n a_i = (1-t)\frac{1}{n} \sum_{i=1}^n a_i + t \sum_{i=1}^n \lambda_i a_i = \sum_{i=1}^n \left(\frac{1-t}{n} + t\lambda_i \right) a_i.$$

Since this is a convex combination of elements of K for sufficiently small t , say $0 < t < \epsilon$, we conclude that $(1-t)a \in K$ for all $0 < t < \epsilon$, absurd since K is a base. Now that we have shown that $0 \notin \text{aff}(K)$, we can construct a functional $u \in E'$ with the property that $\langle u, x \rangle = 1$ for all $x \in \text{aff}(K)$. From the fact that $\langle u, a \rangle = 1$ for all $a \in K$ we see that u must be strictly positive on E_+ . \square

The following easy corollary, which is concerned with cones and bases in ordered *Banach* spaces, will be needed later.

Corollary 1.34. *Let E be an ordered Banach space whose positive cone E_+ admits a base K . If the functional u determined via Lemma 1.33 is continuous and satisfies $\|x\| \leq k \langle u, x \rangle$ for some $k \in \mathbb{R}$ and all $x \in E_+$, then $\text{cl}(E_+)$ is again a cone, it has a base $\text{cl}(K)$, and moreover $\text{cl}(K) = \text{cl}(E_+) \cap u^{-1}(1)$.*

Proof. We start by showing that under the aforementioned hypotheses u is strictly positive on $\text{cl}(E_+)$. Indeed, if $(x_n)_n \subset E_+$ is a sequence converging in norm to $x \in \text{cl}(E_+)$, i.e. $\lim_n \|x - x_n\| = 0$, we have $k \langle u, x \rangle = k \lim_n \langle u, x_n \rangle \geq \lim_n \|x_n\| = \|x\|$, where we used the continuity of u and the fact that $\lim_n \|x_n\| = \|x\|$, easily seen to be a consequence of the triangular inequality $|\|x\| - \|x_n\|| \leq \|x - x_n\|$. Thus, $\langle u, x \rangle = 0$ is possible only if $x = 0$.

An immediate consequence of u being strictly positive on $\text{cl}(E_+)$ is that $\text{cl}(E_+)$ is a legitimate cone, i.e. that $\text{cl}(E_+) \cap -\text{cl}(E_+) = \{0\}$. The set $\bar{K} := \text{cl}(E_+) \cap u^{-1}(1)$ is then a base by Lemma 1.33, and it is clear that $\text{cl}(K) \subseteq \bar{K}$ since the latter set is closed and contains K . On the other hand, if $(x_n)_n \subset E_+$ converges in norm to $x \in \text{cl}(E_+) \cap u^{-1}(1)$, then $\lim_n \langle u, x_n \rangle = \langle u, x \rangle = 1$ and thus the sequence $a_n := \frac{x_n}{\langle u, x_n \rangle} \in K$ satisfies $\lim_n \|x - a_n\| = 0$, in turn implying that $x \in \text{cl}(K)$ and finally that $\text{cl}(K) = \bar{K} = \text{cl}(E_+) \cap u^{-1}(1)$. \square

Given an ordered vector space whose positive cone E_+ is spanning and equipped with a base K , there is a natural way to turn it into a (semi-)normed space. In order to do this, it is useful (although not necessary) to employ the strictly positive functional u that is uniquely associated with K via (1.19).

Definition 1.35. *An ordered vector space E whose positive cone E_+ is spanning and admits a base K is called a **base norm space** if*

$$q_K(x) := \inf \{ \langle u, x_+ + x_- \rangle : x = x_+ - x_-, x_{\pm} \geq 0 \} \quad (1.20)$$

defines a norm on E with respect to which E is complete. Here, u is associated with K via (1.19).

Some observations on the above definition are in order.

- (a) The fact that K is a base of a spanning cone ensures that the infimum appearing on the right-hand side of (1.20) is always finite. However, in general (1.20) will only define a semi-norm, unless the convex set $\text{co}(K \cup (-K))$ is **linearly bounded**, meaning that its intersections with one-dimensional subspaces are always bounded.
- (b) In practice, the reader will not be surprised by the fact that an example of the failure of q_K to be a norm can already be constructed in finite dimension as long as we use non-closed cones. For instance, taking $E = \mathbb{R}^2$ and $E_+ = \{(x, y) : y > 0\} \cup \{(0, 0)\}$ yields $q_K(x, y) = |y|$, which is only a semi-norm. The non-closed nature of E_+ was a necessary ingredient in this example. In fact, in finite dimension q_K is always a norm if E_+ is closed, or equivalently if K is compact.⁵
- (c) As anticipated, it is useful but not mandatory to have the functional u in hand for the sake of the above definition. We could have defined the base norm directly as

$$\|x\| = \inf \{ |\alpha| + |\beta| : x = \alpha a + \beta b, a, b \in K, \alpha, \beta \in \mathbb{R} \}. \quad (1.21)$$

To see why (1.20) translates to (1.21), it suffices to realise that the optimal representation $x = \alpha a + \beta b$ in (1.21) can always be chosen to have $\alpha \geq 0$ and $\beta \leq 0$. For instance, if $\alpha, \beta > 0$ then $x = (\alpha + \beta) \frac{\alpha a + \beta b}{\alpha + \beta}$ is an alternative representation with $\alpha' = \alpha + \beta$ and $\beta' = 0$, which then satisfies $|\alpha'| + |\beta'| \leq |\alpha| + |\beta|$.

- (d) The functional u of (1.19) is always continuous with respect to the base norm. In fact, whenever $x = x_+ - x_-$ with $x_{\pm} \geq 0$ we find $|\langle u, x \rangle| = |\langle u, x_+ \rangle - \langle u, x_- \rangle| \leq \langle u, x_+ \rangle + \langle u, x_- \rangle = \langle u, x_+ + x_- \rangle$, and taking the infimum as in (1.20) we see that $\|u\|_* \leq 1$.

⁵In finite dimension, a base of a closed cone is necessarily compact, and in fact every closed cone admits a compact base (see [90] or Lemma 2.4).

Before we proceed further, let us discuss an explicit example of a base norm space.

Example 1.36. The simplest example of a base norm space is perhaps the sequence space l^1 [43, §1.6], defined by

$$l^1 := \left\{ (a_n)_{n \in \mathbb{N}} \subset \mathbb{R} : \sum_n |a_n| < \infty \right\}. \quad (1.22)$$

Such a space is complete when equipped with the norm

$$\|a\|_1 := \sum_n |a_n|, \quad (1.23)$$

and once ordered by the (closed) cone $E_+ := \{(a_n)_n : a_n \geq 0 \forall n\}$ becomes a base norm space with base given by

$$K := \left\{ (a_n)_n \in E_+ \subset l^1 : \sum_n a_n = 1 \right\}.$$

Now, let us study a bit more in detail certain aspects of the theory of base norm spaces. Such spaces enjoy a lot of useful properties, one of the most important of which is undoubtedly the additivity of the norm on positive vectors. This simple observation is the content of the following lemma, for which we refer the reader again to [76, ‘Order unit and base norm spaces’].

Lemma 1.37. *Let E be a base norm space with positive cone E_+ , and let u be the functional identified by the base K via Lemma 1.33. Then $\|x\| = \langle u, x \rangle$ for all $x \geq 0$, and thus $\|a\| = 1$ for all $a \in K$.*

Proof. If $x \geq 0$ then (1.20) implies immediately that $\|x\| \leq \langle u, x \rangle$, but the converse is also true since any representation of the form $x = x_+ - x_-$ with $x_{\pm} \geq 0$ (and so $\langle u, x_{\pm} \rangle \geq 0$) satisfies $\langle u, x_+ \rangle + \langle u, x_- \rangle \geq \langle u, x_+ \rangle - \langle u, x_- \rangle = \langle u, x \rangle$. \square

We can also wonder, how the unit ball of a base norm space can be characterised in terms of its base K . The following result will answer the question and help us to develop the theory further. For its proof and that of its corollary, see [76, ‘Order unit and base norm spaces’].

Proposition 1.38. *Let E be a base norm space with associated base K . Then its unit ball is*

$$B = \text{cl}(\text{co}(K \cup -K)).$$

Proof. From (1.20) together with the fact that $\langle u, a \rangle = 1$ for all $a \in K$ we see immediately that $\text{co}(K \cup -K) \subseteq B$. Since the latter set is closed, we deduce

$\text{cl}(\text{co}(K \cup -K)) \subseteq B$. To show the converse, we resort once more to (1.20), which tells us that $B \subseteq (1 + \epsilon)\text{co}(K \cup -K)$ holds for all $\epsilon > 0$. Then we find

$$B \subseteq \bigcap_{\epsilon > 0} (1 + \epsilon)\text{co}(K \cup -K) \subseteq \text{cl}(\text{co}(K \cup -K)),$$

where for the last step we used (1.15). \square

As an easy corollary of the above lemma, we find a slightly simplified formula for the norm associated with the dual of a base norm space.

Corollary 1.39. *Let E be a base norm space with base K . Then the dual norm $\|\cdot\|_*$ on E^* admits the expression*

$$\|\varphi\|_* = \sup_{a \in K} |\langle \varphi, a \rangle|. \quad (1.24)$$

Proof. By (1.1) and using Proposition 1.38, we have

$$\begin{aligned} \|\varphi\|_* &= \sup \{ |\langle \varphi, x \rangle| : x \in B \} \\ &= \sup \{ |\langle \varphi, x \rangle| : x \in \text{cl}(\text{co}(K \cup -K)) \} \\ &= \sup \{ |\langle \varphi, x \rangle| : x \in \text{co}(K \cup -K) \} \\ &= \sup \{ |\langle \varphi, a \rangle| : a \in K \}. \end{aligned}$$

Observe that for the third equality we used the continuity of φ . \square

An obvious consequence of the expression (1.24) is that $\|\cdot\|_*$ is monotone with respect to the dual order, i.e. that $\lambda \geq \varphi$ implies $\|\lambda\|_* \geq \|\varphi\|_*$.

We saw that the positive cone in an order unit space is necessarily closed (Proposition 1.29). We can also wonder, whether some analogue of this result holds for base norm spaces. The answer is partially affirmative, for the positive cone is not guaranteed to be closed, but one can take its closure without affecting the base norm, a result that goes back to Ellis [46].

Proposition 1.40. *Let E be a base norm space with positive cone E_+ and base K . Then:*

- (a) E_+ is spanning, 2-normal and $(1 + \epsilon)$ -generating for all $\epsilon > 0$;
- (b) $\text{cl}(E_+)$ is again a cone with base $\text{cl}(K)$, and the base norms associated with K and $\text{cl}(K)$ coincide;

Proof. Let us start with claim (a). The positive cone of a base norm space is spanning by definition. To show that it is 2-normal, we pick $x \in [B] = (B + E_+) \cap (B - E_+)$ (where B is the unit ball) and proceed to show that $\|x\| \leq 2$, i.e. that $x \in 2B$. By hypothesis we can write $x = y_+ + z_+ = y_- - z_-$, where $\|y_{\pm}\| \leq 1$ and $z_{\pm} \geq 0$. Using Lemma 1.37, we see that $\|z_+\| + \|z_-\| = \|z_+ + z_-\| = \|y_- - y_+\| \leq \|y_-\| + \|y_+\| \leq 2$. Then by the triangular inequality

$$\|2x\| = \|y_+ + z_+ + y_- - z_-\| \leq \|z_+\| + \|z_-\| + \|y_+\| + \|y_-\| \leq 4,$$

i.e. $\|x\| \leq 2$. Checking that E_+ is $(1 + \epsilon)$ -generating for all $\epsilon > 0$ is even easier. For a fixed $\epsilon > 0$ and $x \in B$, from (1.20) we see that we can find $x_{\pm} \geq 0$ such that $x = x_+ - x_-$ and $\langle u, x_+ + x_- \rangle \leq 1 + \epsilon$. The representation $\frac{x}{1+\epsilon} = \frac{1}{1+\epsilon}(x_+ - x_-)$ then shows that $\frac{x}{1+\epsilon} \in]B[$, yielding the claim.

Now we move on to proving claim (b). We already saw that the functional u identified by (1.19) is necessarily continuous with respect to the base norm (observation (d) above). Since $\langle u, x \rangle = \|x\|$ holds for all $x \in E_+$ by Lemma 1.37, we can apply Corollary 1.34 and conclude that $\text{cl}(E_+)$ is a legitimate cone with base $\text{cl}(K) = \text{cl}(E_+) \cap u^{-1}(1)$. It remains to show that $\text{cl}(K)$ induces the same base norm. This follows almost immediately from Proposition 1.38, because it is easy to verify that

$$\text{cl}(\text{co}(K \cup -K)) = \text{cl}(\text{co}(\text{cl } K \cup -\text{cl } K))$$

and hence the two unit balls coincide. □

Remark. The reader might have noticed a certain degree of asymmetry between the concept of order unit space (Subsection 1.6.1) and that of base norm space. That this is the case is manifest if one compares, for instance, Propositions 1.29 and 1.40: the positive cone is automatically closed for order unit spaces, but does not need to be such for base norm spaces (as examples in finite dimension already show). Ultimately, this asymmetry is due to the fact that we imposed the Archimedean axiom on order unit spaces from the start. If in Definition 1.27 we had only required p_u to be a norm for some $u \geq 0$, then symmetry would have been somewhat restored. However, in most of the literature on the subject it is customary to refer to order unit spaces as complying with the requirements in Definition 1.27 as it is.

1.6.3 Duality of order unit and base norm spaces

As the reader might have guessed from our Examples 1.28 and 1.36, deliberately chosen to be closely related, the concepts of order unit and base norm spaces are in a certain sense dual to each other. This subsection is concerned with giving a precise sense to this intuition. Let us start by establishing an easy corollary to the characterisation of bases given in Lemma 1.33.

Corollary 1.41. *Let E be an order unit space with order unit u . Then*

$$K_* := \{\varphi \in E_+^* : \langle \varphi, u \rangle = 1\} \tag{1.25}$$

is a base for the dual cone $E_+^ \subseteq E^*$.*

Proof. Once viewed as a functional on E^* , u is clearly strictly positive on the dual cone E_+^* , for any $\varphi \in E_+^*$ with $\langle \varphi, u \rangle = 0$ satisfies also $\langle \varphi, x \rangle = 0$ for all $x \in [-u, u]$ and hence $\varphi = 0$, because $[-u, u]$ is the unit ball of E by Corollary 1.30. By Lemma 1.33, we conclude that K_* defined via (1.25) is a base of E_+^* . □

Now, we are ready to present the duality theorem for base norm and order unit spaces, as was established by Edwards [47] and Ellis [45] building upon previous progresses made by Krein [89]. Our exposition of the proof is based on [76].

Theorem 1.42 (Edwards-Ellis). [76, Theorem 3.1]. *Let E be an ordered Banach space whose positive cone E_+ is closed. Let the Banach dual E^* be ordered by the cone E_+^* of positive and continuous functionals. Then:*

- (a) *E is a base norm space iff E^* is an order unit space; in this case, base and order unit are linked by (1.19).*
- (b) *dually, E is an order unit space iff E^* is a base norm space with weak*-compact base; in this case, order unit and base are linked by (1.25).*

Proof. Let us break down the proof into several steps.

- We start by showing that the dual of a base norm space is an order unit space with order unit given by (1.19). Thanks to Lemma 1.33, given a base K of the positive cone E_+ , we can identify a strictly positive functional $u \in E'$ such that $K = \{x \in E_+ : \langle u, x \rangle = 1\}$. We know that u is indeed bounded with respect to the base norm (see observation (d) in Subsection 1.6.2), so that $u \in E^*$ and we have a candidate for the order unit in E^* . Since E_+ is closed by hypothesis, Lemma 1.31 tells us that what is left to show is only that $\|\varphi\| = p_u(\varphi)$ for all $\varphi \in E^*$, where p_u is given by (1.16). Applying Corollary 1.39, we obtain

$$\begin{aligned}
\|\varphi\|_* &= \sup \{ |\langle \varphi, a \rangle| : a \in K \} \\
&= \inf \{ t > 0 : -t \leq \langle \varphi, a \rangle \leq t \quad \forall a \in K \} \\
&= \inf \{ t > 0 : -t \langle u, a \rangle \leq \langle \varphi, a \rangle \leq t \langle u, a \rangle \quad \forall a \in K \} \\
&= \inf \{ t > 0 : -t \langle u, x \rangle \leq \langle \varphi, x \rangle \leq t \langle u, x \rangle \quad \forall x \in E_+ \} \\
&= \inf \{ t > 0 : \varphi \in t[-u, u] \} \\
&= p_u(\varphi),
\end{aligned}$$

as claimed.

- Now we demonstrate that if E^* is an order unit space with unit u then E is a base norm space with base $K := \{x \geq 0 : \langle u, x \rangle = 1\}$, completing the proof of claim (a). We have to show that K is really a base, and that the norm $\|\cdot\|$ on E coincides with the function $q_K(x)$ in (1.20). First of all, as in the proof of Lemma 1.37 we know that $q_K(x) = \langle u, x \rangle$ for all $x \geq 0$. We now employ formula (1.4) to show that $\|x\| \leq q_K(x)$ whenever $x \in E_+ - E_+$. Indeed, take a representation $x = x_+ - x_-$ such that $x_{\pm} \geq 0$ and some $\varphi \in E^*$ such that $\|\varphi\|_* \leq 1$, i.e. $\varphi \in [-u, u]$. Then we have

$$|\langle \varphi, x \rangle| = |\langle \varphi, x_+ \rangle - \langle \varphi, x_- \rangle| \leq |\langle \varphi, x_+ \rangle| + |\langle \varphi, x_- \rangle| \leq \langle u, x_+ \rangle + \langle u, x_- \rangle.$$

Taking the supremum over $\varphi \in [-u, u]$ and the infimum over x_{\pm} this yields $\|x\| \leq q_K(x)$, as claimed. An immediate consequence is that u is strictly positive on E_+ , because any $x \geq 0$ such that $\langle u, x \rangle = 0$ satisfies also $\|x\| \leq q_K(x) = \langle u, x \rangle = 0$ and hence $x = 0$. Then, Lemma 1.33 ensures that K is really a base of E_+ .

It remains to show that $q_K(x) \leq \|x\|$ for all $x \in E$ (implying in particular that E_+ is spanning). If B is the unit ball of $\|\cdot\|$ we have clearly

$$\|x\| = \inf\{t > 0 : x \in tB\}.$$

Since E_+^* is 1-normal by Proposition 1.29, using Theorem 1.21 we see that E_+ must be $(1 + \epsilon)$ -generating for all $\epsilon > 0$, and in particular spanning. Then, picking $t > 0$ such that $x \in tB$, since $B \subseteq (1 + \epsilon)B[$ we obtain also

$$x \in (1 + \epsilon)t]B[= (1 + \epsilon)t \operatorname{co}((B \cap E_+) \cup -(B \cap E_+)),$$

implying that there is a representation of the form $x = (1 + \epsilon)t(py_+ - (1 - p)y_-)$, where $y_{\pm} \in B \cap E_+$ and $p \in [0, 1]$. By definition of $q_K(x)$ we get

$$q_K(x) \leq (1 + \epsilon)t(p \langle u, y_+ \rangle + (1 - p) \langle u, y_- \rangle) \leq (1 + \epsilon)t,$$

where the last step follows because $\|u\|_* \leq 1$ and $y_{\pm} \in B$ together imply that $\langle u, y_{\pm} \rangle \leq \|u\|_* \|y_{\pm}\| \leq 1$. The above inequality holds for all $\epsilon > 0$, so we conclude that $q_K(x) \leq t$, and after taking the infimum over t this yields $q_K(x) \leq \|x\|$.

- Let us move on to claim (b), and prove that if E is an order unit space then E^* is a base norm space with a weak*-compact base given by (1.25). According to Definition 1.35, we have to check that: (i) the cone E_+^* is spanning; (ii) K_* defined via (1.25) is a base of E_+^* ; (iii) K_* is weak*-compact; and (iv) the norm on E^* is given by (1.20).

Now, (i) follows from Proposition 1.29, while (ii) is established by Corollary 1.41. As for (iii), note first that $K_* = E_+^* \cap u^{-1}(1)$ is the intersection of two weak*-closed sets and hence itself weak*-closed. For the argument that justifies why E_+^* is weak*-closed, we refer the reader to the discussion at the beginning of Section 1.5. Next, we see that $\varphi \in K_*$ implies $\varphi \in E_+^*$ and hence $|\langle \varphi, x \rangle| \leq \langle \varphi, u \rangle = 1$ for all $x \in B = [-u, u]$, with equality when $x = u$. By (1.1), we conclude that $\|\varphi\|_* = 1$, and in particular that $\varphi \in B^\circ$ with B° being the dual unit ball. Since B° is weak*-compact by the Banach-Alaoglu theorem (Theorem 1.14), and K_* is weak*-closed, K_* is a closed subset of a compact set and hence itself compact (in our case, weak*-compact).

We now show (iv). Call $q_{K_*}(\varphi)$ the function defined as in (1.20), i.e.

$$\begin{aligned} q_{K_*}(\varphi) &= \inf \{ \langle \varphi_+ + \varphi_-, u \rangle : \varphi = \varphi_+ - \varphi_-, \varphi_{\pm} \in E_+^* \} \\ &= \inf \{ t > 0 : \varphi \in t \operatorname{co}(K_* \cup -K_*) \}. \end{aligned}$$

Now, while proving claim (iii) we saw that $\|\varphi\|_* = \langle \varphi, u \rangle$ whenever $\varphi \in E_+^*$ (rescale and use the fact that $\varphi \in K_* \Rightarrow \|\varphi\|_* = 1$). As a consequence, one gets $K_* = B^\circ \cap E_+^*$ and hence $\text{co}(K_* \cup -K_*) =]B^\circ[$ with the notation of (1.6). Since E_+ is 1-normal by Proposition 1.29 and hence E_+^* is 1-generating by Theorem 1.21, we see that

$$]B^\circ[= B^\circ.$$

Therefore,

$$q_{K_*}(\varphi) = \inf \{t > 0 : \varphi \in t]B^\circ[\} = \inf \{t > 0 : \varphi \in tB^\circ\} = \|\varphi\|_*,$$

concluding the proof of (iv).

- It remains to prove that if E^* is a base norm space with a weak*-compact base K_* then E must be an order unit space with order unit u satisfying (1.25). By Lemma 1.33, we see that K_* must be of the form $K_* = \{\varphi \in E_+^* : f(\varphi) = 1\}$, where $f : E^* \rightarrow \mathbb{R}$ is strictly positive on E_+^* . Then, we have to show that indeed f acts as $f(\varphi) = \langle \varphi, u \rangle$, for some $u \in E$. Thanks to Lemma 1.13, it is enough to show that f is weak*-continuous, i.e. that $\ker(f)$ is weak*-closed. Proposition 1.15 tells us that we have just to check that $\ker(f) \cap B^\circ$ is weak*-closed.

To this purpose, let us show that $\ker(f) \cap B^\circ = \frac{1}{2}(K_* - K_*)$. We start by proving that $\frac{1}{2}(K_* - K_*) \subseteq \ker(f) \cap B^\circ$. Pick $\varphi \in \frac{1}{2}(K_* - K_*)$, i.e. assume that there is a decomposition $\varphi = \frac{1}{2}(\lambda - \mu)$ with $\lambda, \mu \in K_*$. Then, we see that $f(\varphi) = \frac{1}{2}(f(\lambda) - f(\mu)) = 0$ and $\|\varphi\|_* \leq \frac{1}{2}(f(\lambda) + f(\mu)) = 1$, i.e. $\varphi \in \ker(f) \cap B^\circ$. Note that we used the fact that E^* is a base norm space to estimate $\|\varphi\|_*$ via (1.20). Conversely, take $\varphi \in \ker(f) \cap B^\circ$ and let us show that $\varphi \in \frac{1}{2}(K_* - K_*)$. Since $\|\varphi\|_* \leq 1$, by hypothesis for all $n > 0$ one can find a decomposition $\varphi = \varphi_n^+ - \varphi_n^-$ such that $f(\varphi_n^+) + f(\varphi_n^-) \leq 1 + \frac{1}{n}$. From $\varphi \in \ker(f)$ we deduce $f(\varphi_n^+) = f(\varphi_n^-)$, so that $f(\varphi_n^\pm) \leq \frac{1+1/n}{2}$. Consequently, $\frac{2}{1+1/n}\varphi_n^\pm \in K_*$ for all $n \in \mathbb{N}$. Since K_* is weak*-compact, up to extracting subsequences we can assume that $2\varphi^\pm := w^*\text{-lim}_n \frac{2}{1+1/n}\varphi_n^\pm \in K_*$ exist. We obtain $\varphi^+ - \varphi^- = \varphi$ and hence $\varphi \in \frac{1}{2}(K_* - K_*)$, as claimed.

With the above argument we have shown that $\ker(f) \cap B^\circ = \frac{1}{2}(K_* - K_*)$ (and indeed that $B^\circ = \text{co}(K_* \cup -K_*)$, for that matter). Since K_* is weak*-compact, the same is true for $\frac{1}{2}(K_* - K_*)$. Thus, also $\ker(f) \cap B^\circ$ is weak*-compact and in particular weak*-closed. As we said, via Proposition 1.15 and Lemma 1.13 this implies that $f(\cdot) = \langle \cdot, u \rangle$ for some $u \in E$. Because of Lemma 1.31, we have only to verify that the norm $\|\cdot\|$ on E is given by (1.16). Thanks to the general formula (1.4) and to Proposition 1.38, we can basically repeat the reasoning used to show the

first part of claim (a):

$$\begin{aligned}
 \|x\| &= \sup \{ |\langle \varphi, x \rangle| : \|\varphi\|_* \leq 1 \} \\
 &= \sup \{ |\langle \varphi, x \rangle| : \varphi \in \text{cl}(\text{co}(K_* \cup -K_*)) \} \\
 &= \sup \{ |\langle \varphi, x \rangle| : \varphi \in \text{co}(K_* \cup -K_*) \} \\
 &= \sup \{ |\langle \varphi, x \rangle| : \varphi \in K_* \} \\
 &= \inf \{ t > 0 : -t \leq \langle \varphi, x \rangle \leq t \ \forall \varphi \in K_* \} \\
 &= \inf \{ t > 0 : -t \langle \varphi, u \rangle \leq \langle \varphi, x \rangle \leq t \langle \varphi, u \rangle \ \forall \varphi \in K_* \} \\
 &= \inf \{ t > 0 : -t \langle \varphi, u \rangle \leq \langle \varphi, x \rangle \leq t \langle \varphi, u \rangle \ \forall \varphi \in E_+^* \} \\
 &= \inf \{ t > 0 : x \in t[-u, u] \} \\
 &= p_u(x),
 \end{aligned}$$

where for the second last step we observed that $\langle \varphi, x \rangle \leq \langle \varphi, y \rangle$ for all $\varphi \in E_+^* = -E_+^\circ$ implies that $x - y \in (E_+^*)^\circ = -E_+^{\circ\circ} = -E_+$, this latter equality being a consequence of (1.10) and of the identity $E_+ = \text{cl}(E_+)$.

□

Remark. In the statement of the above Theorem 1.42, one can freely change the expression ‘weak*-compact’ to ‘weak*-closed’. In fact, a base of a base norm space is always a subset of the dual unit ball, which is in any case weak*-compact by Theorem 1.14. Thus, in this situation weak*-closedness and weak*-compactness are equivalent concepts.

1.7 Ludwig’s embedding theorem

We are finally ready to state and prove Ludwig’s embedding theorem, a cornerstone of general probabilistic theories. Referring to the discussion at the end of Subsection 1.1.2, this result accomplishes two goals at once. First, it provides a geometrically intuitive representation of the set of states and effects as described in Subsection 1.1.2. Secondly, it gives us a computationally convenient way to evaluate the probability functional μ subjected to Axioms 1, 2, 3. In doing so, we arrive at the so-called ‘state space formalism’ of general probabilistic theories, to be described in the next chapter, building it on firm and rigorous mathematical foundations. The theorem can be found in [7, IV] and [27] (with proof) and also quoted in [65, III, Theorem 3.1] and [76, ‘The structure of ordered Banach spaces in axiomatic quantum mechanics’] (without proof).

Theorem 1.43 (Ludwig’s embedding theorem). [7, IV, Theorem 3.7] or [65, III, Theorem 3.1]. *Let Ω, Λ be sets equipped with a function $\mu : \Lambda \times \Omega \rightarrow [0, 1]$ that satisfies Axioms 1, 2, 3. Then, there is a base norm space E with closed positive cone and an embedding of Ω into E and of Λ into the dual order unit space E^* such that:*

- (a) *(the images of) Ω and Λ are convex, $0, u \in \Lambda$ (where $0 \in E^*$ is the zero functional, and u is the unit effect), and $\Lambda = u - \Lambda$;*

- (b) μ coincides with the canonical bilinear form on $E^* \times E$, i.e. $\mu(\lambda, \omega) = \langle \lambda, \omega \rangle$ for all $\omega \in \Omega$ and $\lambda \in \Lambda$;
- (c) E is normed by a base $K = \text{cl}(\Omega)$ which is the norm closure of Ω , and u is the order unit that norms E^* ;
- (d) the linear span of Λ is weak*-dense in E^* .

Moreover, the embedding satisfying all the above requirements is substantially unique, in the sense that any two embeddings are connected by an isometric order isomorphism (see the discussion at the beginning of the forthcoming Subsection 1.7.9).

The proof we present through Subsections 1.7.1 – 1.7.9 is a re-elaborated version of the one in [7, pp.104-119]. The argument is quite long and especially in certain passages it relies heavily on many technical tools we have been developing through Sections 1.3 – 1.5, so for the sake of the presentation we break it down into simpler steps, and devote a subsection of the present section to discuss each of them.

1.7.1 Constructing the basic vector spaces

Let us start by defining the real vector space V consisting of all the functions $x : \Lambda \rightarrow \mathbb{R}$ that admit a representation of the form

$$x(\cdot) = \sum_{i=1}^p \alpha_i \mu(\cdot, \omega_i), \quad (1.26)$$

for $p \in \mathbb{N}$, $\alpha_i \in \mathbb{R}$, and $\omega_i \in \Omega$. Clearly, V is a vector space once equipped with the pointwise sum between functions. Moreover, there is an obvious embedding $\Omega \subset V$ given by $\Omega \ni \omega \mapsto \omega(\cdot) := \mu(\omega, \cdot)$. Clearly, one can repeat the same construction exchanging Ω and Λ , and define a vector space W as composed of all functions $y : \Omega \rightarrow \mathbb{R}$ of the form

$$y(\cdot) = \sum_{j=1}^q \beta_j \mu(\lambda_j, \cdot), \quad (1.27)$$

with $q \in \mathbb{N}$, $\beta_j \in \mathbb{R}$, and $\lambda_j \in \Lambda$. Again, there is a natural embedding $\Lambda \subset W$ given by $\Lambda \ni \lambda \mapsto \lambda(\cdot) := \mu(\cdot, \lambda)$. Let us stress here that the fact that these embeddings are indeed injective is a consequence of Axiom 1. We can also use Axiom 2 to deduce that: (1) the zero of W as a vector space belongs to Λ ; (2) Λ contains also a special element u , i.e. (the image of) the unit effect, such that $\langle u, \omega \rangle = 1$ for all $\omega \in \Omega$.

One can endow the two vector spaces we just constructed with a natural bilinear form $\langle \cdot, \cdot \rangle : W \times V \rightarrow \mathbb{R}$. For $x \in V$ and $y \in W$ given by (1.26) and (1.27), respectively, we define

$$\langle y, x \rangle := \sum_{i=1}^p \sum_{j=1}^q \alpha_i \beta_j \mu(\lambda_j, \omega_i). \quad (1.28)$$

It is very easy to verify that once one thinks of Ω, Λ as subsets of V, W , one has $\langle \cdot, \cdot \rangle|_{\Lambda \times \Omega} = \mu(\cdot, \cdot)$. This bilinear form allows us to think of the elements of V as linear functionals on W . This can be done by defining $W \ni y : V \rightarrow \mathbb{R}$ as given by $y(x) := \langle y, x \rangle$. In symbols, we write $W \subseteq V'$, where V' is the algebraic dual of V .

An immediate advantage of having embedded Ω and Λ into vector spaces is that Axiom 3 now implies that $\Omega \subseteq V$ and $\Lambda \subseteq W$ are convex sets. Let us verify this claim by taking $\omega_1, \omega_2 \in \Omega$, $p \in [0, 1]$, and by showing that $p\omega_1 + (1-p)\omega_2 \in \Omega$ (the proof for Λ is totally analogous). Axiom 3 ensures the existence of some $\tau \in \Omega$ such that $\mu(\lambda, \tau) = p\mu(\lambda, \omega_1) + (1-p)\mu(\lambda, \omega_2)$ for all $\lambda \in \Lambda$. This can be rephrased by saying that the function $\mu(\cdot, \tau)$ defined on Λ coincides with the convex combination of the two functions $\mu(\cdot, \omega_1)$ and $\mu(\cdot, \omega_2)$. By definition of the embedding $\Omega \subseteq V$, we see that inside V we have really $\tau = p\omega_1 + (1-p)\omega_2$, i.e. Ω is convex.

It is also clear from Axiom 2 that the equation $\Lambda = u - \Lambda$ holds. In fact, for $\lambda \in \Lambda$ there must exist $\lambda' \in \Lambda$ such that $\mu(\lambda', \cdot) = \mu(u, \cdot) - \mu(\lambda, \cdot)$ as functions on Ω . Consequently, $\lambda' = u - \lambda$ and therefore $u - \lambda \in \Lambda$. As we shall see, this completes the verification of claim (a).

1.7.2 Introducing a Banach space structure

Now, let us define

$$R := \left\{ y \in V' : \sup_{\omega \in \Omega} |\langle y, \omega \rangle| < \infty \right\}, \quad (1.29)$$

which is readily verified to be a vector subspace of V' . We can of course turn R into a normed vector space by equipping it with the norm

$$\|y\|^R := \sup_{\omega \in \Omega} |\langle y, \omega \rangle|. \quad (1.30)$$

To check that this gives a true norm and not a semi-norm, observe that $0 = \langle y, \omega \rangle = y(\omega)$ for all $\omega \in \Omega$ implies that $y : \Omega \rightarrow \mathbb{R}$ is the zero function.

Note. We chose the notation $\|\cdot\|^R$ instead of the simpler $\|\cdot\|$ because we want to reserve the latter for the norm on E , the space we are aiming for.

Now, the following holds.

Proposition 1.44. *The space R defined in (1.29) becomes a Banach space once it is endowed with the norm (1.30).*

Proof. We have only to verify that R is complete with the norm (1.30). Since showing this requires a bit of care, we break down the proof into elementary steps.

- (1) Take a Cauchy sequence $(y_n)_{n \in \mathbb{N}} \subset R$. A pointwise limit y can be constructed via the formula $y(x) := \lim_{n \rightarrow \infty} \langle y_n, x \rangle$, where the right-hand side is well defined because for a fixed x as in (1.26) we see that $(\langle y_n, x \rangle)_{n \in \mathbb{N}} \subset \mathbb{R}$ is a Cauchy sequence, as follows from the inequalities

$$\begin{aligned} |\langle y_n, x \rangle - \langle y_m, x \rangle| &= \left| \sum_{i=1}^p \alpha_i (\mu(y_n, \omega_i) - \mu(y_m, \omega_i)) \right| \\ &\leq \sum_{i=1}^p |\alpha_i| |\mu(y_n, \omega_i) - \mu(y_m, \omega_i)| \\ &\leq \sum_{i=1}^p |\alpha_i| (\|y_n\|^R - \|y_m\|^R) \\ &\leq \left(\sum_{i=1}^p |\alpha_i| \right) \|y_n - y_m\|^R \end{aligned}$$

together with the definition of Cauchy sequence in a Banach space.

- (2) Now that we have constructed a plausible limit y , let us examine its properties. The bilinearity of $\langle \cdot, \cdot \rangle$ and the linearity of limits together ensure that $y(\cdot)$ is a linear functional, so that $y \in V'$.
- (3) We have to verify that $y \in R$, i.e. that $\sup_{\omega \in \Omega} |\langle y, \omega \rangle| < \infty$. This can be done by writing

$$\begin{aligned} \sup_{\omega \in \Omega} |\langle y, \omega \rangle| &= \sup_{\omega \in \Omega} \lim_{n \rightarrow \infty} |\langle y_n, \omega \rangle| \\ &\leq \sup_{\omega \in \Omega} \sup_{n \in \mathbb{N}} |\langle y_n, \omega \rangle| \\ &= \sup_{n \in \mathbb{N}} \sup_{\omega \in \Omega} |\langle y_n, \omega \rangle| \\ &= \sup_{n \in \mathbb{N}} \|y_n\|^R \\ &< \infty, \end{aligned}$$

where the last inequality follows from the fact that $(\|y_n\|^R)_{n \in \mathbb{N}} \subset \mathbb{R}$ is itself a Cauchy sequence, which is in turn a consequence of the triangular inequality $|\|y_n\|^R - \|y_m\|^R| \leq \|y_n - y_m\|^R$. Cauchy sequences are convergent and in particular bounded, which yields the last step.

- (4) Finally, we have to check that the convergence $y_n \rightarrow y$ happens in the Banach space norm and not only pointwise. In other words, we have to prove that $\lim_{n \rightarrow \infty} \|y - y_n\|^R = 0$. Assume by contradiction that there exists $\epsilon_0 > 0$ such that $\|y - y_n\|^R > 2\epsilon_0$ holds for infinitely many values of n . Then we will show that $(y_n)_{n \in \mathbb{N}} \subset R$ can not be a Cauchy sequence, i.e. for all $N \in \mathbb{N}$ there are $n, m \geq N$ such that $\|y_n - y_m\|^R \geq \epsilon_0 > 0$. To this purpose, let $N \in \mathbb{N}$ be given. By hypothesis, we can pick $n \geq N$

such that $\sup_{\omega \in \Omega} |\langle y - y_n, \omega \rangle| = \|y - y_n\|^R > 2\epsilon_0$, and consequently $\omega \in \Omega$ such that $|\langle y - y_n, \omega \rangle| \geq 2\epsilon_0$. Since $\langle y - y_n, \omega \rangle = \lim_{m \rightarrow \infty} \langle y_m - y_n, \omega \rangle$, this shows the existence of $m \geq N$ such that $|\langle y_m - y_n, \omega \rangle| \geq \epsilon_0$, finally implying that $\|y_n - y_m\|^R \geq \epsilon_0$.

□

Now, let us examine few further properties of R that we will need in the following. First, W and R are both vector subspaces of V' , and in fact it turns out that $W \subseteq R$. In fact, for all $y \in W$ represented as in (1.27), we obtain

$$\begin{aligned} \sup_{\omega \in \Omega} |\langle y, \omega \rangle| &= \sup_{\omega \in \Omega} \left| \sum_{j=1}^q \beta_j \mu(\lambda_j, \omega) \right| \\ &\leq \sum_{j=1}^q |\beta_j| |\mu(\lambda_j, \omega)| \\ &\leq \sum_{j=1}^q |\beta_j| \\ &< \infty, \end{aligned}$$

which implies $y \in R$, by definition. This fact has an analogous at the level of primal spaces. Namely, one can show in pretty much the same way that elements of V are continuous (equivalently, bounded) when seen as linear functionals on R , or in other words that $V \subseteq R^*$, where R^* is the Banach dual of R . This is because an element $x \in V$ represented as in (1.26), once it is seen in a natural way as a functional on R via the formula $y \mapsto x(y) := \langle y, x \rangle$, is clearly bounded, since

$$|x(y)| = |\langle y, x \rangle| = \left| \sum_{i=1}^p \alpha_i \mu(y, \omega_i) \right| \leq \left(\sum_{i=1}^p |\alpha_i| \right) \sup_{\omega \in \Omega} |\langle y, \omega \rangle| = \left(\sum_{i=1}^p |\alpha_i| \right) \|y\|^R.$$

This shows that the norm of $x \in V$ as a functional is bounded from above, or more explicitly that $\|x\|_*^R \leq \sum_{i=1}^p |\alpha_i| < \infty$. In turn, this proves that $x \in R^*$.

1.7.3 Inducing new norms on V

Now that we have a norm on R , we can use it to construct new norms on V . For instance, for any given subspace Z such that $W \subseteq Z \subseteq R$, we can construct a norm $\|\cdot\|_Z$ on V given by

$$\|x\|_Z := \sup_{z \in Z, \|z\|^R \leq 1} |\langle z, x \rangle|. \quad (1.31)$$

Because of the constraint $\|z\|^R \leq 1$, one sees that $\|x\|_Z \leq \|x\|_*^R < \infty$, which implies in particular that (1.31) is well defined for all $x \in V$. Clearly, we

obtain also the identity $\|\cdot\|_R = \|\cdot\|_*^R$. Observe that (1.31) defines a norm, since $\langle \lambda, x \rangle = 0$ for all $\lambda \in \Lambda \subseteq W \subseteq Z$ shows that x is the zero function on Λ .

Observe that by its very definition (1.30) the norm $\|\cdot\|^R$ satisfies

$$\|\lambda\|^R \leq 1 \quad \forall \lambda \in \Lambda. \quad (1.32)$$

This implies that all effects can be plugged into (1.31), yielding $\|x\|_Z \geq \sup_{\lambda \in \Lambda} |\langle \lambda, x \rangle|$. In particular, when one considers $x = \omega \in \Omega$, the unit effect $u \in \Lambda$ gives us $\|\omega\|_Z \geq |\langle u, \omega \rangle| = 1$. On the other hand, since $|\langle y, \omega \rangle| \leq 1$ whenever $\|y\|^R \leq 1$ (by the very definition of $\|\cdot\|^R$), we know that $\|\omega\|_Z \leq 1$. Putting these two observations together we see that

$$\|\omega\|_Z = \langle u, \omega \rangle = 1 \quad \forall \omega \in \Omega. \quad (1.33)$$

1.7.4 Constructing new Banach spaces

Now that we have endowed V with the family of norms $\|\cdot\|_Z$, we can consider the associated completions V_Z of V . Observe that even if we know that $V \subseteq R^*$, we can not a priori claim that $V_Z \subseteq R^*$, because R^* is complete with respect to the norm $\|\cdot\|_*^R$, while the completion V_Z is taken with respect to another (a priori incomparable) norm $\|\cdot\|_Z$. In other words, it will not be the case in general that all elements $x \in V_Z$ are bounded when viewed as functionals on R with respect to $\|\cdot\|^R$. They are however bounded on Z with the same norm $\|\cdot\|^R$, since for $y \in Z$ one has $|\langle y, x \rangle| \leq \|y\|^R \|x\|_Z$. This shows that $V_Z \subseteq Z^*$ holds true, a weaker inclusion than the one discussed above because the set of linear functionals on R that are bounded only on Z is in general a proper superset of R^* (inside the algebraic dual R').

Dually, with the same reasoning we see that every $y \in Z$ is also bounded as a functional on V with respect to the norm $\|\cdot\|_Z$, which tells us that it can be extended to a bounded functional on V_Z . In other words, we can regard Z as a subset of V_Z^* . In turn, this latter subspace of V' happens to be contained inside R . In fact, consider a functional $y \in V_Z^* \subseteq V'$, and let us show that it satisfies (1.29). Write

$$\begin{aligned} \sup_{\omega \in \Omega} |\langle y, \omega \rangle| &\stackrel{(1)}{=} \sup_{\omega \in \Omega \subseteq V} \frac{|\langle y, \omega \rangle|}{\|\omega\|_Z} \\ &\stackrel{(2)}{\leq} \sup_{0 \neq x \in V} \frac{|\langle y, x \rangle|}{\|x\|_Z} \\ &\stackrel{(3)}{<} \infty \end{aligned}$$

The justification of these steps is as follows: (1) we employed (1.33) to divide by $\|\omega\|_Z$; (2) we extended the supremum from Ω to the whole $V \supset \Omega$; (3) we used the fact that y is bounded as a functional on the normed space $(V, \|\cdot\|_Z)$ by hypothesis. The above derivation shows also that

$$\|y\|_{Z^*} \leq \|y\|^R \quad \forall y \in V_Z^*. \quad (1.34)$$

Let us summarise the above discussion by stating that

$$W \subseteq Z \subseteq V_Z^* \subseteq R. \quad (1.35)$$

At this point, let us ask ourselves whether the expression (1.31) is still valid for all $x \in V_Z$, while it was initially conceived only for $x \in V$. The answer is affirmative, as we proceed to show.

Lemma 1.45. *Equation (1.31) is valid on the whole completion V_Z of V .*

Proof. First of all, pick $x \in V_Z$ and $z \in Z \subseteq V_Z^*$ such that $\|z\| \leq 1$. Using (1.34), we immediately deduce that $|\langle z, x \rangle| \leq \|x\|_Z \|z\|_{Z^*} \leq \|x\|_Z \|z\| \leq \|x\|_Z$, so that

$$\|x\|_Z \geq \sup_{z \in Z, \|z\|^R \leq 1} |\langle z, x \rangle|.$$

To show the opposite inequality, consider a sequence $(v_n)_n \subset V$ such that $\lim_n \|x - v_n\|_Z = 0$. This sequence satisfies $\langle z, x \rangle := \lim_n \langle z, v_n \rangle$ (remember that we are allowed to consider z as a functional in V_Z^* , thus acting also on x). In fact, such a sequence can be used to *define* the scalar product $\langle z, x \rangle$. But there is more: the above limit is also uniform on the set of $z \in Z$ such that $\|z\|^R \leq 1$. This is because once we fix an $\epsilon > 0$ and construct $N \in \mathbb{N}$ such that $\|x - v_n\|_Z \leq \epsilon$ for all $n \geq N$, using (1.34) we will also have $|\langle z, x \rangle - \langle z, v_n \rangle| = |\langle z, x - v_n \rangle| \leq \|z\|_{Z^*} \|x - v_n\|_Z \leq \|z\|^R \|x - v_n\|_Z \leq \epsilon$ independently of $z \in Z$ as long as $\|z\|^R \leq 1$. In turn, this allows us to conclude that

$$\sup_{z \in Z, \|z\|^R \leq 1} |\langle z, x \rangle| \geq \sup_{z \in Z, \|z\|^R \leq 1} |\langle z, v_n \rangle| - \epsilon = \|v_n\|_Z - \epsilon.$$

Taking the limit on n yields

$$\sup_{z \in Z, \|z\|^R \leq 1} |\langle z, x \rangle| \geq \|x\|_Z - \epsilon$$

for all $\epsilon > 0$, i.e.

$$\sup_{z \in Z, \|z\|^R \leq 1} |\langle z, x \rangle| \geq \|x\|_Z,$$

as claimed. In the above derivation, we used the fact that also the norm of x can be written as a limit, namely $\|x\|_Z = \lim_n \|v_n\|_Z$. This can be seen easily by applying the triangular inequality to show that $\lim_n |\|x\|_Z - \|v_n\|_Z| \leq \limsup_n \|x - v_n\|_Z = 0$. \square

1.7.5 Identifying symmetric subspaces

The Banach spaces V_Z are our candidates for the E we want to construct. Observe that the expression (1.30) of the norm on R coincides with the formula for the dual of a base norm given by Corollary 1.39. Since we want to retain this feature, we would like to consider subspaces Z such that the norm on V_Z^* as a Banach space coincides with the restriction of the norm on R . We proceed now to identify a class of Z s such that this happens. We need the following elementary result.

Lemma 1.46. [7, IV, Theorem 3.4]. *For all Z such that $W \subseteq Z \subseteq R$, the norm associated with V_Z^* and the restriction of that of R coincide at least on Z itself.*

Proof. Since (1.34) holds, we have only to show that $\|z\|^R \leq \|z\|_{Z^*}$ for all $z \in Z$. We proved that the formula (1.31) is satisfied not only for $x \in V$ but more generally for $x \in V_Z$, hence we see immediately that $\frac{|\langle z, x \rangle|}{\|z\|^R} \leq \|x\|_Z$ for all $z \in Z, z \neq 0$. Therefore,

$$\|z\|_{Z^*} = \sup_{x \in V_Z, \|x\|_Z \leq 1} |\langle z, x \rangle| \leq \sup_{x \in V_Z, \|x\|_Z \leq 1} \|z\|^R \|x\|_Z \leq \|z\|,$$

concluding the proof of the claim. \square

The above calculation shows that there is a simple way to ensure that the norm of R coincides with that of V_Z^* on the whole V_Z^* , i.e. requiring that $Z = V_Z^*$. Subspaces Z such that $W \subseteq Z \subseteq R$ and $Z = V_Z^*$ will play a special role in what follows, and will be called *symmetric*. The simplest example of a symmetric subspace is R itself, as (1.35) with $Z = R$ implies that $R = V_R^*$.

1.7.6 Constructing the smallest symmetric subspace

Symmetric subspaces have the remarkable properties of being closed under intersections, as we proceed to show.

Proposition 1.47. *If $\{Z_i\}_{i \in I}$ is a family (finite or infinite) of symmetric subspaces, their intersection $Z_* := \bigcap_{i \in I} Z_i$ is again a symmetric subspace.*

Proof. We need only to show that $V_{Z_*}^* \subseteq Z_*$, since the converse is generally true (1.35). We start by noticing that because of the very definition of $\|\cdot\|_Z$ (1.31), from the inclusions $Z_* \subseteq Z_i$ the inequality $\|\cdot\|_{Z_*} \leq \|\cdot\|_{Z_i}$ follows immediately. In turn, such inequalities show that functionals that are continuous with respect to $\|\cdot\|_{Z_*}$ are also continuous with respect to $\|\cdot\|_{Z_i}$, in formula $V_{Z_*}^* \subseteq V_{Z_i}^*$. Taking the intersection over $i \in I$ and using the identities $V_{Z_i}^* = Z_i$, one obtains

$$V_{Z_*}^* \subseteq \bigcap_{i \in I} V_{Z_i}^* = \bigcap_{i \in I} Z_i = Z_*,$$

yielding the claim. \square

Now, we can meaningfully consider the smallest symmetric subspace, call it E_* . Evidently, E_* can also be written as the intersection of all symmetric subspaces, in formula

$$E_* := \bigcap_{\substack{W \subseteq Z \subseteq R, \\ Z \text{ symmetric}}} Z. \quad (1.36)$$

Now, define

$$E := V_{E_*}. \quad (1.37)$$

Since E_* is symmetric, we have $E_* = V_{E_*}^* = E^*$, i.e. E^* is the dual Banach space of E . Thus, from now on we will forget about E_* and rather call it E^* . The norm on E will be denoted by $\|\cdot\|$, while $\|\cdot\|_*$ will be the corresponding dual norm on E^* .

We have $V \subseteq E$ and $W \subseteq E^*$, and clearly the canonical bilinear form $\langle \cdot, \cdot \rangle$ on $E^* \times E$ coincides with $\mu(\cdot, \cdot)$ on $\Lambda \times \Omega$, meeting requirement (b). Applying Lemma 1.46 to the symmetric subspace E^* , and using the expression (1.31) generalised to all $x \in V_{E_*} = E$ by Lemma 1.45, we establish that

$$\|\varphi\|_* = \|\varphi\|^R = \sup_{\omega \in \Omega} |\langle \varphi, \omega \rangle| \quad \forall \varphi \in E^*. \quad (1.38)$$

The norm on E is correspondingly given by (1.4):

$$\|x\| = \sup_{\varphi \in E^*, \|\varphi\|_* \leq 1} |\langle \varphi, x \rangle| = \|x\|_{E^*} \quad \forall x \in E. \quad (1.39)$$

1.7.7 Order structure: meeting requirement (c)

We now show that E defined by (1.36) and (1.37) can be given a structure of base norm space. In order to do this, we first define a cone on E given by $C := \mathbb{R}_+ \cdot \Omega = \{t\omega : t \geq 0, \omega \in \Omega\}$. Obviously, Ω is a base for C , the strictly positive functional determined by Lemma 1.33 coinciding with the unit effect $u \in \Lambda$. By (1.32) and (1.38), we see that u satisfies $\|u\|_* \leq 1$, in particular it is continuous. Moreover, (1.33) guarantees that $\|x\| = \langle u, x \rangle$ holds for all $x \in C$. Then, we can apply Corollary 1.34 and conclude that $E_+ := \text{cl}(C)$ is a legitimate cone and thus defines an ordering on E . Furthermore, we also know that $\text{cl}(\Omega) = \text{cl}(C) \cap u^{-1}(1)$ is a base of E_+ .

It remains to show that: (1) E equipped with the closed cone E_+ is a base norm space with corresponding base $\text{cl}(\Omega)$; and (2) E is an order unit space with order unit u . Thanks to Theorem 1.42, we see that (2) implies (1), so we restrict ourselves to proving (2). We can follow the same style of reasoning as the one adopted in the proof of Theorem 1.42, and write

$$\begin{aligned} \|\varphi\|_* &\stackrel{(1)}{=} \sup \{|\langle \varphi, \omega \rangle| : \omega \in \Omega\} \\ &\stackrel{(2)}{=} \sup \{|\langle \varphi, \omega \rangle| : \omega \in \text{cl}(\Omega)\} \\ &= \inf \{t > 0 : -t \leq \langle \varphi, \omega \rangle \leq t \quad \forall \omega \in \text{cl}(\Omega)\} \\ &\stackrel{(3)}{=} \inf \{t > 0 : -t \langle u, \omega \rangle \leq \langle \varphi, \omega \rangle \leq t \langle u, \omega \rangle \quad \forall \omega \in \text{cl}(\Omega)\} \\ &\stackrel{(4)}{=} \inf \{t > 0 : -t \langle u, x \rangle \leq \langle \varphi, x \rangle \leq t \langle u, x \rangle \quad \forall x \in E_+\} \\ &= \inf \{t > 0 : \varphi \pm tu \geq E_+\} \\ &= \inf \{t > 0 : \varphi \in t[-u, u]\} \\ &= p_u(\varphi), \end{aligned}$$

where p_u is given by (1.16). The justification of the above steps is as follows: (1) we applied (1.38); (2) we recalled that $\langle \varphi, \cdot \rangle$ is continuous as a functional

acting on E ; (3) we employed the identity $\text{cl}(\Omega) = \text{cl}(C) \cap u^{-1}(1)$; (4) we used the fact that $\text{cl}(\Omega)$ is a base of the cone E_+ . Now, E_+^* is weak*-closed and hence norm-closed because of the discussion at the beginning of Section 1.5. Then, Lemma 1.31 ensures that $\|\varphi\|_* = p_u(\varphi)$ for all $\varphi \in E^*$ is sufficient to guarantee that E^* is an order unit space. As we mentioned above, Theorem 1.42 tells us that E is a base norm space with corresponding base $\text{cl}(\Omega)$.

1.7.8 Weak*-denseness: meeting requirement (d)

This is perhaps the most delicate part of the proof of Ludwig's theorem, as it relies crucially on many results in Banach space theory as well as in general topology. Since many different topologies will play a significant role in what follows, throughout this subsection we suspend the usage of expressions like weak and weak* topology on (say) E and E^* , and rather more pedantically we spell them out as $\sigma(E, E^*)$ and $\sigma(E^*, E)$, respectively.

Let us start with some notation. Given any Banach space F with dual F^* , and subsets $M \subseteq F$ and $N \subseteq F^*$, the **annihilators** M^\perp and ${}^\perp N$ are given by

$$M^\perp := \{\varphi \in F^* : \langle \varphi, x \rangle = 0 \ \forall x \in M\}, \quad (1.40)$$

$${}^\perp N := \{x \in F : \langle \varphi, x \rangle = 0 \ \forall \varphi \in N\}. \quad (1.41)$$

By confronting these definitions with those of polars, (1.8) and (1.41), we see that $M^\perp = M^\circ$ and ${}^\perp N = N^\circ$ whenever M or N are subspaces. Using Lemma A.1, we arrive at the formulae

$${}^\perp(M^\perp) = \text{cl}(M), \quad (1.42)$$

$$({}^\perp N)^\perp = \text{cl}_{w^*}(N), \quad (1.43)$$

valid when M, N are subspaces. The interested reader can look up this standard observation in many textbooks in functional analysis (see for instance [77, Theorem 4.7(b)] or [79, Proposition 2.6.6(c)]).

We now turn to the first results of this subsection. Let us remind the reader that because of (1.35) one can identify R with the dual subspace V_R^* .

Lemma 1.48. *For $W \subseteq Z \subseteq R = V_R^*$ and $V \subseteq V_R$, pick $x \in {}^\perp Z \subseteq V_R$ and construct a sequence $(v_n)_n \subset V$ such that $\lim_n \|x - v_n\|_R = 0$. Then*

$$\lim_n \|v_n\|_Z = 0.$$

Proof. In fact, if this were not the case we could find $\epsilon_0 > 0$ such that $\|v_n\|_Z \geq 2\epsilon_0$ happens frequently in n . We could then construct a sequence $(z_n)_n \subset Z$ such that $\|z_n\|^R \leq 1$ and $|\langle z_n, v_n \rangle| \geq \epsilon_0$ frequently in n . Since $x \in {}^\perp Z$, we would have

$$\|x - v_n\|_R = \sup_{y \in R, \|y\|^R \leq 1} |\langle y, x - v_n \rangle| \geq |\langle z_n, x - v_n \rangle| = |\langle z_n, v_n \rangle| \geq \epsilon_0,$$

in contradiction with the assumption that $\lim_n \|x - v_n\|_R = 0$. \square

Proposition 1.49. [7, IV, Theorem 3.6]. *Let Z be a subspace such that $W \subseteq Z \subseteq R = V_R^*$. If Z is $\sigma(V_R^*, V_R)$ -closed, then it is symmetric.*

Proof. We have to show that a $\sigma(V_R^*, V_R)$ -closed subspace Z satisfies also $V_Z^* \subseteq Z$. Naturally, the topology $\sigma(V_R^*, V_R)$ is nothing by the weak* topology induced by the Banach space V_R on its dual V_R^* . By (1.43) we can write $Z = (\perp Z)^\perp$, where the annihilators have to be taken with respect to V_R and V_R^* . Then, pick $y \in V_Z^*$ and $x \in \perp Z \subseteq V_R$, and let us show that $\langle y, x \rangle = 0$, so that we will have $y \in (\perp Z)^\perp$ and the claim will follow.

Take a sequence $(v_n)_n \subset V$ converging to x in the norm $\|\cdot\|_R$. Since $y \in V_Z^* \subseteq V_R^*$, on the one hand we have $\langle y, x \rangle = \lim_n \langle y, v_n \rangle$. On the other hand, since y is bounded on V with respect to $\|\cdot\|_Z$, one has $|\langle y, v_n \rangle| \leq \|y\|_{Z^*} \|v_n\|_Z$ and consequently $\lim_n |\langle y, v_n \rangle| \leq \|y\|_{Z^*} \lim_n \|v_n\|_Z = 0$, where the last step is an application of Lemma 1.48. \square

Until now, we only looked at topologies induced by the global Banach spaces V_R, V_R^* . However, claim (d) of Theorem 1.43 concerns an ‘intrinsic’ topology, induced by the smallest symmetric subspace E^* on itself. Therefore, to complete the proof we need some tools to relate the two concepts. This is the purpose of the forthcoming Proposition 1.51.

Lemma 1.50. *Let Z be a subspace such that $W \subseteq Z \subseteq V_Z^* \subseteq R = V_R^*$. Then the topologies $\sigma(V_R^*, V_R)|_{V_Z^*}$ and $\sigma(V_Z^*, V_Z)$ satisfy $\sigma(V_R^*, V_R)|_{V_Z^*} \subseteq \sigma(V_Z^*, V_Z)$. In other words, the restriction of the global weak* topology to V_Z^* is coarser than its own weak* topology.*

Proof. First of all, observe that if you have a family \mathcal{F} of functions $f : X \rightarrow \mathbb{R}$ on a set X , and a subset $Y \subseteq X$, the restriction to Y of the initial topology generated by \mathcal{F} on X coincides with the initial topology on Y , i.e. $\sigma(X, \mathcal{F})|_Y = \sigma(Y, \mathcal{F})$. This is a trivial consequence of the identities $(f|_Y)^{-1}(U) = f^{-1}(U) \cap Y$, which are valid for all (open) subsets $U \subseteq \mathbb{R}$.

Therefore, the restriction of $\sigma(V_R^*, V_R)$ to V_Z^* is identical to $\sigma(V_Z^*, V_R)$, the initial topology generated directly on V_Z^* by the family of functionals of the form $x : V_Z^* \rightarrow \mathbb{R}$ (acting as $x(\cdot) = \langle \cdot, x \rangle$), where $x \in V_R$. By definition of initial topology (Section 1.3), in order to show that $\sigma(V_Z^*, V_R) \subseteq \sigma(V_Z^*, V_Z)$, we have to prove that every function $x : V_Z^* \rightarrow \mathbb{R}$ with $x \in V_R$ is continuous with respect to $\sigma(V_Z^*, V_Z)$.

Using the characterisation of continuity with nets, our claim is that every net $(\varphi_\alpha)_\alpha \subseteq V_Z^*$ that converges to $\varphi \in V_Z^*$ in the $\sigma(V_Z^*, V_Z)$ topology satisfies also $\lim_\alpha \langle \varphi_\alpha, x \rangle = \langle \varphi, x \rangle$ for all $x \in V_R$ (i.e. $x : V_Z^* \rightarrow \mathbb{R}$ is continuous). Now, pick $x \in V_R$, so that there exists a sequence $(v_n)_n \subset V$ with $\lim_n \|x - v_n\|_R = 0$. As a consequence, we have also $\langle \psi, x \rangle = \lim_n \langle \psi, v_n \rangle$ for all $\psi \in V_R^*$, and in particular for all $\psi \in V_Z^*$. Then, our claim becomes

$$\lim_\alpha \lim_n \langle \varphi_\alpha, v_n \rangle = \lim_n \langle \varphi, v_n \rangle . \quad (1.44)$$

Now, $(v_n)_n$ is Cauchy with respect to the norm $\|\cdot\|_R$, hence also with respect to $\|\cdot\|_Z \leq \|\cdot\|_R$. Thus, there is $x' \in V_Z$ such that $\lim_n \|x' - v_n\|_Z = 0$. Since $\varphi_\alpha, \varphi \in V_Z^*$, this implies that $\lim_n \langle \varphi_\alpha, v_n \rangle = \langle \varphi_\alpha, x' \rangle$ and $\lim_n \langle \varphi, v_n \rangle = \langle \varphi, x' \rangle$. Thus, (1.44) becomes

$$\lim_\alpha \langle \varphi_\alpha, x' \rangle = \langle \varphi, x' \rangle,$$

which is satisfied by hypothesis since $(\varphi_\alpha)_\alpha \subseteq V_Z^*$ converges to $\varphi \in V_Z^*$ in the $\sigma(V_Z^*, V_Z)$ topology, and $x' \in V_Z$. \square

Proposition 1.51. [7, IV, Theorem 3.5]. *Let Z be a symmetric subspace, and let $T \subseteq Z$ be a subspace of $Z = V_Z^*$ that is $\sigma(V_Z^*, V_Z)$ -closed. Then T is also $\sigma(V_R^*, V_R)$ -closed, in particular symmetric.*

Proof. In what follows, we will denote by B_F the unit ball of a generic Banach space F . Let us start by noting that a subspace T is $\sigma(V_R^*, V_R)$ -closed (i.e. weak*-closed with respect to the Banach spaces V_R and V_R^*) iff $T \cap B_R$ is $\sigma(V_R^*, V_R)$ -closed, by Proposition 1.15. From now on, we will thus focus on this latter set. Since by Lemma 1.46 the restriction of the norm of $R = V_R^*$ to $Z = V_Z^*$ coincides with the norm naturally carried by the space V_Z^* , we have $B_Z = B_R \cap Z = B_{V_Z^*}$. Then, from $T \subseteq Z$ we find

$$T \cap B_R = (T \cap Z) \cap B_R = T \cap (Z \cap B_R) = T \cap B_Z.$$

Now, using again Proposition 1.15 (the 'easy' direction) we see that since T is $\sigma(V_Z^*, V_Z)$ -closed, the same happens to $T \cap B_{V_Z^*}$, i.e. to $T \cap B_Z$ (remember that $B_Z = B_{V_Z^*}$). In fact, $T \cap B_{V_Z^*}$ is a $\sigma(V_Z^*, V_Z)$ -closed subset of the unit ball $B_{V_Z^*}$, which is in turn $\sigma(V_Z^*, V_Z)$ -compact by the Banach-Alaoglu theorem (Theorem 1.14). Then, we conclude that $T \cap B_Z = T \cap B_{V_Z^*}$ is in fact $\sigma(V_Z^*, V_Z)$ -compact. On $T \cap B_Z \subseteq V_Z^*$ the $\sigma(V_Z^*, V_Z)$ topology is weaker than (the restriction of) the $\sigma(V_R^*, V_R)$ topology because of Lemma 1.50, and clearly Hausdorff because any weak* topology is such. But by Lemma 1.3 a coarser Hausdorff topology on a compact set must coincide with the original topology, hence $\sigma(V_Z^*, V_Z)$ and $\sigma(V_R^*, V_R)|_{V_Z^*}$ are indeed the same. In particular, since $T \cap B_Z = T \cap B_R$ was $\sigma(V_Z^*, V_Z)$ -compact, it is also $\sigma(V_R^*, V_R)$ -compact, and in particular $\sigma(V_R^*, V_R)$ -closed. As discussed above, this implies that T is $\sigma(V_R^*, V_R)$ -closed, and by Proposition 1.49 that T is symmetric. \square

Now, we are ready to state the result that completes the proof of claim (d) of Theorem 1.43.

Proposition 1.52. *Let E, E^* be defined by (1.37), (1.36). Then W is $\sigma(E^*, E)$ -dense in E^* .*

Proof. We remind the reader that E^* is nothing but the smallest symmetric subspace. Assume by contradiction that W is not $\sigma(E^*, E)$ -dense in E^* , and consider its $\sigma(E^*, E)$ -closure in E^* , call it $\bar{W} \subsetneq E^*$. Since E^* is symmetric and \bar{W} is $\sigma(E^*, E)$ -closed, we can apply Proposition 1.51 and conclude that also \bar{W}

is symmetric. This is a contradiction because E^* was the smallest symmetric subspace. \square

Remark. We now have another way to characterise the smallest symmetric subspace E^* : it is the $\sigma(V_R^*, V_R)$ -closure of W inside V_R^* .

1.7.9 Verifying uniqueness up to isomorphisms

Before we start proving this last claim, let us clarify the concept of order and isometric isomorphisms in the context of ordered Banach spaces. We remind the reader that an invertible linear map $\Phi : E_1 \rightarrow E_2$ between two ordered vector spaces E_1, E_2 is called an **order isomorphism** if for all $x, y \in E_1$ one has $x \leq y \Leftrightarrow \Phi(x) \leq \Phi(y)$, or equivalently if $x \geq 0 \Leftrightarrow \Phi(x) \geq 0$. If instead E_1, E_2 are Banach spaces, the relevant concept is that of **isometric isomorphism**, which is a surjective linear map $\Phi : E_1 \rightarrow E_2$ such that $\|\Phi(x)\|_2 = \|x\|_1$ for all $x \in E_1$. Isometric isomorphisms are automatically injective and hence invertible. We are mainly interested in the case of E_1, E_2 being ordered Banach spaces. To preserve these two structures simultaneously, a surjective linear map $\Phi : E_1 \rightarrow E_2$ must be at the same time an order isomorphism and an isometric isomorphism, i.e. an **isometric order isomorphism**.

Let us come to the proof of the uniqueness claim. As it turns out, we can rely on Ludwig's construction of E (see (1.36) and (1.37)) to show that such a subspace is isometrically and order isomorphic to any other Banach space F satisfying requirements (b), (c) and (d) of Theorem 1.43. Here we are going to follow this route, which is also Ludwig's route, to conclude the proof of the theorem. However, it is also instructive to prove the uniqueness of E up to isometric order isomorphisms from scratch, using only the constraints expressed by the statement of Theorem 1.43. For the interested readers, we report such an argument in Appendix B. Here, let us confine ourselves to arguing as follows.

- (1) Since V, W are obtained via finite linear combinations from Ω, Λ , we can think of them as linear subspaces of F, F^* , i.e. $V \subseteq F$ and $W \subseteq F^*$.
- (2) F is a base norm space, hence by Corollary 1.39 the norm on F^* is given by

$$\|\varphi\|_* = \sup_{\omega' \in \text{cl}(\Omega)} |\langle \varphi, \omega' \rangle| = \sup_{\omega \in \Omega} |\langle \varphi, \omega \rangle|.$$

This means that one can embed F^* in R , i.e. $F^* \subseteq R$, and that on F^* the norm of R coincides with that of F^* .

- (3) Moreover, because of the above reasoning F^* turns out to be a symmetric subspace, which implies that $E^* \subseteq F^*$.
- (4) On the other hand, we saw that E^* is nothing but the closure of W inside $V_R^* = R$ with respect to the $\sigma(V_R^*, V_R)$ topology. By Lemma 1.50, one has $\sigma(V_R^*, V_R)|_{F^*} \subseteq \sigma(F^*, F)$. Using the fact that W is $\sigma(F^*, F)$ -dense in F^* , we have

$$F^* = \text{cl}_{\sigma(F^*, F)}(W) \subseteq \text{cl}_{\sigma(V_R^*, V_R)}(W) = E^*.$$

This proves that $E^* = F^*$.

- (5) Now, $E = V_{E^*}$ by definition (1.36), while $F = V_{F^*}$ because by requirement (c) V is norm-dense in F (as is easy to verify). Therefore, one gets also $E = F$ (as Banach spaces).
- (6) Finally, the construction of the order structure in E discussed in Subsection 1.7.7 ensures that E and F are also order isomorphic with the same isomorphism that makes $E = F$ as Banach spaces.

Chapter 2

Composite systems in GPTs

2.1 Introduction

Throughout the previous chapter we laid the foundation of the so-called abstract state space formalism of general probabilistic theories (GPTs). Now, we will rather take this as a starting point, and turn our attention to the study of more advanced constructions. A wise approach is to let ourselves guide by the history of quantum mechanics, and a fundamental lesson we learnt in the past century is that many interesting phenomena appear when we look at composite systems, i.e. systems made of different parties, each one with its share of a physical system. In quantum mechanics this approach ultimately led to the discovery of nonlocality [3, 6], which has revealed us a fundamental feature of Nature that had escaped our previous investigations. Within the context of GPTs, we will therefore study how the state space of a bipartite system can be related to those of its components.

The present chapter is organised as follows. The rest of this section is devoted to presenting the fundamental questions we want to address in a non-technical language (Subsection 2.1.1) and to discussing our original contributions (Subsection 2.1.2). For the sake of completeness, throughout Section 2.2 we provide an overview of the mathematical machinery of GPTs for single systems and introduce the formalism to deal with bipartite systems. In trying to make this chapter as self-contained as possible, we will quickly review some of the concepts already introduced in Chapter 1. In Section 2.3, we list some examples of GPTs or classes of GPTs. Besides that of acquainting the reader with the formalism, this has the purpose of providing us with an essential sample of models on which we can test conjectures and hypotheses, something that will turn out useful also in Chapter 3. In Section 2.4, which constitutes in some respects a mathematical detour, we are concerned with examining the separability problem, well-known from quantum mechanics, in the broader GPT setting. Despite spurring primarily from mathematical curiosity, many tools we will develop there will be needed in the rest of the thesis. The last two sections constitute the heart of the chapter and contain the most significant original

material. In Section 2.5 we introduce a problem that is in our view central to our understanding of composite systems in general probabilistic theories, and review what is already known about it. Section 2.6 is then devoted to presenting our progress toward a solution of this fundamental question. In doing so, we will also provide novel insights regarding some mathematical problems in functional analysis.

2.1.1 Some fundamental questions

Among all the scientific revolutions the twentieth century has brought to us, we are mainly concerned here with the dramatic shift of paradigm that led us to change our picture of the physical world forever. I am naturally speaking of the discovery that our naive understanding of Nature as a classical reality is inconsistent at a fundamental level. As part of a standard undergraduate course, a student is usually taught that the reason for this is the *superposition principle*, which dictates that microscopic objects have access to an infinite number of incomparable states, incomparable meaning that none of them can be prepared as a probabilistic mixture of the others (in mathematical language, they are extreme points of the set of states). This phenomenon corresponds to what is usually called the *non-classicality* of quantum theory. As we will see, this type of non-classicality is indeed common to almost all GPTs, since classical theories have a very peculiar geometric structure (Subsection 2.3.1).

There is however another situation where the incompatibility with a classical theory emerges clearly, namely when one considers bipartite quantum states and the correlations they can exhibit. Can we analyse this kind of questions also in the more general GPT setting? As a start, given two GPTs A and B that describe two physical systems, what can we say about the system obtained by joining them? As we saw in Chapter 1, general probabilistic theories are in a sense nothing more than a way to embed reasonable physical axioms into a coherent mathematical picture, and for this mathematical picture to have some reason to be, we should be able to describe the joint system by another GPT, call it AB .

We will explore the constraints that have to be imposed on AB in Subsection 2.2.3. Leaving aside the technical details, we will see that there is no unique way to construct a meaningful AB that is compatible with the local models A and B . And in fact we should not expect to find such a uniqueness, for there can be many ways one can (or can not) have access to global states that were not available when the systems were physically disjoint. As is well known, this is exactly what happens for quantum mechanical systems: states that are accessible with local operations assisted by classical communication between the two parties are commonly called *separable* (originally ‘classically correlated’ [91]), while global states that can not be prepared in such a way are dubbed *entangled*. As the reader may know, entanglement is necessary (although not sufficient [91]) to generate non-local correlations out of a physical state via local measurements. States that are sufficiently entangled to be able to produce such correlations are called *nonlocal*.

It would be hard to overestimate the importance of entanglement from the purely conceptual and – we are tempted to say – ontological level. Already from the early days of quantum mechanics, the centrality of the concept was very clearly perceived, although not all of the consequences were immediately appreciated. The memorable debate on the status of this phenomenon had as protagonists scientists of the calibre of Einstein [2], Schrödinger [5] and Bohr [4], and ultimately led to the discovery of nonlocality by Bell [3]. Besides, and on a different line, over the last 30 years quantum information science has proven, among other things, how crucial entanglement is to overcome classical constraints at the operational level [92].

As it turns out (Subsection 2.5.1), the same definitions of separable, entangled and non-local state make sense in the GPT setting as well. In view of the above discussion, we want to gain a deeper understanding of what happens when one combines two different non-classical theories. This goal can be seen as the common thread that binds together the whole first part of the present thesis. Before detailing our programme further, let us stress an important point. In order to make the questions independent of the particular way of joining the systems, we should focus not on a specific composition rule but rather on all the possible composition rules. In fact, in Subsection 2.2.3 it will emerge that we are always free to prescribe *de iure* that the only allowed states of the bipartite system are separable. We articulate our programme as follows.

- (i) First, and most importantly, we want to answer the following question: *is the existence of entangled states always contemplated when one combines two non-classical theories?* This question, which will be adequately formalised and treated in Section 2.6, plays a major role in our investigation and in our conception of the problem. One could complement it with an analogous yet stronger query, with the word ‘Bell nonlocality’ in place of ‘entanglement’. This latter question is very well motivated from the foundational and operational point of view, as it is formulated using concretely measurable probabilities instead of rather abstract algebraic properties.
- (ii) An even stronger (and much more difficult to tackle) question is then: *is it possible to lower bound the maximal entanglement that can be displayed by a composite theory once it is known how non-classical the local theories are?* The problem here is intentionally ill-defined, and we refer the reader to Section 2.7 for more details. Again, the same question with ‘Bell nonlocality’ in place of ‘entanglement’ is of interest.
- (iii) In general, *out of the many non-classical phenomena quantum information helped us to understand, which ones are generic features of non-classical theories?* This very commendable programme [60] has been pursued by many authors and already led to important results [93, 59, 41]. Here, we want to lay emphasis on the quantitative version the the above (general) query: *how strong is the non-classicality exhibited by quantum mechanics, in all its various forms?* Is it the strongest that is possibly

conceivable, or are there more extreme examples? We will devote part of Chapter 3 to answering this question for the particular non-classical phenomenon known as *data hiding* [9, 10].

- (iv) Even though it is of perhaps secondary importance, we want to *develop tools to tackle the separability problem in general probabilistic theories*, in analogy with what has been done in quantum mechanics. This means finding necessary and sufficient conditions for a state of a bipartite theory to be separable, and naturally the problem can be tackled both in general and for specific classes of models. Section 2.4 is devoted to presenting some results in this direction.

To the careful reader it will not escape that we are putting in practise the approach we outlined in the discussion at the beginning of Section 1.1, which in turn reflects our personal view on the field: general probabilistic theories should not be seen merely only as a framework within which one tries to *derive* quantum theory¹, but rather – or complementarily – as a chance to put some order into the plethora of surprising phenomena we have discovered after the advent of the quantum revolution.² Arguably, not all of these phenomena are at the same level, since some are common to all non-classical theories, some are not, and the latter must be regarded as features that are specifically quantum and thus tell us something even deeper about the structure of reality.

2.1.2 Original contributions

The purpose of this subsection is to give the reader a brief overview of the original contributions contained in this chapter, mostly in Sections 2.4 and 2.6. Our interest in the problems we discuss here was spurred mainly by the study of state discrimination subjected to locality constraints, as conducted in Chapter 3. Contrary to the order of appearance in the present thesis, the forthcoming Chapter 3 contains what came out of our early research project, while we arrived at formulating many of the problems presented in the present chapter later during my PhD. In fact, how it is often the case in science, more foundational and abstract questions come at a later stage, when one gains a deeper understanding of some of the subtleties of the theory.

In view of all this, the reader will not be surprised by our lack of a full understanding of the answers to the many questions we formulate here. This is reflected in the abundance of conjectures and questions we put forward here (Conjectures 2.1 and 2.2, Questions 2.3 and 2.4), which contribute to make this

¹Obviously, I do not think that the axiomatic approach to quantum theory boils down to this, which would be both ungenerous and simplistic. Instead, the many excellent works on the subject are praiseworthy in their effort to condense the peculiarity of quantum mechanics in a short list of requirements.

²Let me add a purely personal note. I would find it very disappointing if someone were to come up with a derivation of quantum mechanics from first and indisputable principles, that is, with a *proof* that Nature *must be* quantum mechanical. To me, it would mean that there is – at a fundamental level – basically nothing else we have to understand about why things are the way they are, which is, by definition, disheartening.

chapter more an exposition of a research project than a complete work. And in fact, almost all the results we obtain here are still unpublished, although we are confident that they will be given some publishable form soon. However, as for every research project worthy of the name, what we *do* have clear in mind are the questions themselves, and motivating and presenting those questions is our main goal here.

That being said, let us try to point the reader to some of the original contributions of this chapter. Section 2.4 is devoted to the study of the separability problem in the context of GPTs, something that will come in handy later, and that to the extend of our knowledge was not the subject of previous investigation. Besides an elementary extension of the Woronowicz criterion [94] to the GPT realm (Proposition 2.15 in Subsection 2.4.1), in the following Subsection 2.4.3 we show that a low enough tensor rank is sufficient to ensure separability of a state in an arbitrary GPT (Theorem 2.23). This extends an analogous quantum mechanical result by Cariello [95, 96]. There are at least two good reasons why we include this generalisation among our original results: first, taking this abstract approach ends up telling us something we did not know even in the quantum mechanical case (Corollary 2.24); and second, after a few common initial steps, the proof departs significantly from the one in [95], and becomes significantly more involved. The final Subsection 2.4.4 discusses the separability problem for a special class of GPTs, encompassing the so-called *centrally symmetric models*. The main result there, Proposition 2.25, establishes an unexpected connection between the separability problem for these models and the beautiful theory of tensor norms as originally put forward by Grothendieck [97, 98].

Section 2.5 is devoted to turning some of the problems discussed in the above Subsection 2.1.1 into mathematically precise conjectures. Particularly central to our investigation will be Conjecture 2.1, the very formulation of which constitutes, to the extent of our knowledge, an original contribution. Despite the fact that we deem it of prime importance for our understanding of the intrinsic features of the phenomenon of nonlocality in physical theories, the problem does not seem to have drawn enough attention in the existing literature. The rest of the section is dedicated a detailed exposition of what is currently known about the question, mostly contained in Theorem 2.37, which is due to Namioka and Phelps [8].

Next, Section 2.6 contains the presentation of our progress toward answering Conjecture 2.1 and some related questions. Subsection 2.6.1 is concerned with the investigation of Question 2.3, which asks for a strengthening of Conjecture 2.1. In Theorem 2.39, we extend Namioka-Phelps theorem (Theorem 2.37) to show that such a strengthening holds true at least for the special case considered in Theorem 2.37. Subsection 2.6.2 is instead devoted to examining Conjecture 2.1 (or more precisely its slightly weaker form, Conjecture 2.2) within the context of centrally symmetric models. We will explain in detail why we deem this restriction of the problem worth investigation. While looking at the specific features of this restricted problem, we will stumble upon some universal Banach space constants, defined in (2.92) and (2.93), that might be of inde-

pendent interest in functional analysis. Here we limit ourselves to showing a nontrivial lower bound on the second such constant (Theorem 2.40), which can be regarded as the main original achievement of the chapter, since it leads us to answering Conjecture 2.2 in the affirmative for the special case of centrally symmetric models (Corollary 2.41). Although we will try to give the reader some intuition why Theorem 2.40 holds, the fact that its claim concerns all Banach spaces makes its proof somewhat long and laborious. We discuss some limitations on the constant appearing in Theorem 2.40 through a somewhat illuminating example in Appendix D.

Finally, let us mention some minor contributions that are believed to be partly original in spirit, either for the approach we adopt or for the proof technique we use. Lemma 2.7, which tells us how to join two physical systems that are locally described by GPTs, is very well known. However, we hope that the some readers will find the simple proof we present in Appendix C useful and worth the space it takes. The definition of locally constrained sets of measurements (Definition 2.8) comes straightforwardly from the quantum mechanical case. In spite of this, some amount of work is required to formalise what we mean by LOCC protocols in the GPT setting.³ Next, centrally symmetric models as given in Definition 2.11 seem to have been considered by the author for the first time in [99]. Finally, although it does not represent a substantial progress in any respect, we are particularly proud of the approach to Bell inequalities and Bell nonlocality taken in Subsection 2.3.4, for we found it conceptually clarifying and notationally convenient (but this is also a matter of personal taste).

2.2 Basic theory

2.2.1 Some preliminaries on Banach spaces and tensor norms

We start this section with some mathematical notions on finite-dimensional Banach spaces and tensor norms. The assumption of finite dimension simplifies heavily the technical tools needed in analysing many of the problems we will investigate, but this is not the only reason why we make it. In fact, some of those problems themselves make only sense in the finite-dimensional setting. This is the case, for instance, for the questions raised in Chapter 3.

Note. A quick disclaimer for the reader: from now on, all vector spaces we will deal with are understood to be finite-dimensional.

One of the immediate advantages of the finite-dimensional setting is that we have no longer to worry about specifying the topology carried by our vector space V . In fact, there is a natural topology that comes from the isomorphism with \mathbb{R}^d ($d = \dim V$). When we make use of topological concepts like

³I am grateful to Howard Barnum and Matthias Christandl for enlightening discussions on this topic.

closedness, this is the topology we are referring to. A well-known property of this kind of topologies is that a subset $K \subseteq V$ is compact iff it is closed and bounded. Another consequence of the isomorphism with \mathbb{R}^d is that the assumption of completeness in Definition 1.5 can be dropped with no repercussions. We will still speak of Banach spaces, but in finite dimension this notion coincides with that of normed space. Furthermore, there are some results that hold specifically for finite-dimensional Banach spaces. An example that is worth mentioning since we will make use of it is the following.

Lemma 2.1 (Auerbach). [97, Appendix A.4]. *Let $(V, \|\cdot\|)$ be a Banach space. If $d = \dim V < \infty$, then there is a basis $\{v_1, \dots, v_d\}$ of V (called the **Auerbach basis**) such that $\|v_i\| = 1 = \|v_i^*\|_*$ for all $i = 1, \dots, d$, where $\{v_i^*\}_i$ denotes the corresponding dual basis, and $\|\cdot\|_*$ the dual norm as in (1.1).*

Remark. The above result is nontrivial because we are requiring a basis and its dual basis to be *simultaneously* normalised.

We now move on to discussing possible ways to turn a tensor product of two Banach spaces into a Banach space itself. In the finite-dimensional setting the only problem we have to worry about is how to construct a suitable tensor norm, i.e. a norm on the tensor product with some additional properties. The reason why we want to devote a full subsection to discussing these mathematical objects may not be apparent yet, but tensor norms will turn out to play a major role throughout the present chapter (particularly in Subsection 2.4.4 and 2.6.2) and to a larger extent in the forthcoming Chapter 3. There, we will even sharpen some of our technical tools and use them to tackle problems related to the ultimate effectiveness of state discrimination subjected to locality constraints.

We start by reminding the reader [97, 98] that given two finite-dimensional Banach spaces V_A, V_B (whose norms will be equally denoted by $\|\cdot\|$ for simplicity), there are two notable ways in which one can construct a norm on the tensor product $V_A \otimes V_B$. The first construction yields the so-called **injective norm**, which can be expressed as

$$\|Z\|_\varepsilon := \max \{ \langle \varphi \otimes \lambda, Z \rangle : \varphi \in V_A^*, \|\varphi\|_* \leq 1, \lambda \in V_B^*, \|\lambda\|_* \leq 1 \}. \quad (2.1)$$

The second norm we are interested in goes under the name of **projective norm**, and is defined as

$$\|Z\|_\pi := \min \left\{ \sum_{i=1}^n \|x_i\| \|y_i\| : n \in \mathbb{N}, Z = \sum_{i=1}^n x_i \otimes y_i \right\}. \quad (2.2)$$

These two norms are dual to each other in the following sense. Thinking of V_A^*, V_B^* as Banach spaces equipped with the dual norms $\|\cdot\|_*$, we can construct the associated injective and projective tensor norms on $V_A^* \otimes V_B^*$, denoted by $\|\cdot\|_{*\varepsilon}$ and $\|\cdot\|_{*\pi}$, respectively. One could wonder how these norms compare to the dual to (2.1) and (2.2), denoted by $\|\cdot\|_{\varepsilon*}$, $\|\cdot\|_{\pi*}$, respectively. As it turns out, one has

$$\|\cdot\|_{*\varepsilon} = \|\cdot\|_{\pi*}, \quad \|\cdot\|_{*\pi} = \|\cdot\|_{\varepsilon*}. \quad (2.3)$$

Note. The notation is intended to help the reader via simple graphic rules. For instance, the symbol $\|\cdot\|_{\varepsilon*}$ stands for ‘first construct the injective norm, then take the dual’, and conversely $\|\cdot\|_{*\pi}$ means ‘first take the dual norms, then construct the projective norm out of them’. Then, the above identities can be recovered easily by remembering that taking a $*$ from the external to the internal position (or vice versa) causes an exchange $\varepsilon \leftrightarrow \pi$.

Concerning the comparison between injective and projective norms, the inequality $\|X\|_{\varepsilon} \leq \|X\|_{\pi}$ is easily seen to hold for all $X \in V_A \otimes V_B$. We will write symbolically

$$\|\cdot\|_{\varepsilon} \leq \|\cdot\|_{\pi}. \quad (2.4)$$

As it turns out, for simple tensors $x \otimes y$ both norms coincide with the product of the local norms, i.e. $\|x \otimes y\|_{\varepsilon} = \|x \otimes y\|_{\pi} = \|x\| \|y\|$ (‘reasonable cross norms’). Establishing an inequality that complements (2.4) will be one of the main steps for the solution of the problems presented in Chapter 3 (Theorem 3.18).

Let us now present an example of the construction of injective and projective norms we sketched above. The reader will soon recognise in our case of study the more familiar notions of matrix norms.

Example 2.2. Consider two Euclidean spaces V_A, V_B , whose norms we denote by $|\cdot|_2$. We can identify $V_A \otimes V_B$ with the set of $d_A \times d_B$ real matrices, denoted by $\mathbb{R}^{d_A \times d_B}$. With this convention, for a given vector $Z \in V_A \otimes V_B$ and two functionals $v \in \mathbb{R}^{d_A} \simeq V_A^*$, $w \in \mathbb{R}^{d_B} \simeq V_B^*$, we can write $\langle v \otimes w, Z \rangle = v^T Z w$. Then, remembering that Euclidean norms are self-dual, from (2.1) we obtain in this special case

$$\begin{aligned} \|Z\|_{\varepsilon} &= \max\{v^T Z w : |v|_{2*}, |w|_{2*} \leq 1\} \\ &= \max\{v^T Z w : |v|_2, |w|_2 \leq 1\} \\ &= \|Z\|_{\infty}, \end{aligned}$$

where $\|Z\|_{\infty}$ denotes the operator norm of the matrix Z , i.e. its largest singular value. The fact that $|\cdot|_{2*} = |\cdot|_2$ also implies that $\|\cdot\|_{*\varepsilon} = \|\cdot\|_{\varepsilon}$. This together with (2.3) shows that

$$\|\cdot\|_{\pi} = \|\cdot\|_{*\varepsilon*} = \|\cdot\|_{\varepsilon*} = \|\cdot\|_{\infty*} = \|\cdot\|_1,$$

where $\|\cdot\|_1$ is the trace norm, i.e. the sum of all singular values.

2.2.2 Single systems

As we already mentioned, the theory we develop in the present chapter builds upon Ludwig’s embedding theorem (Theorem 1.43). In fact, the definition of GPT we are going to adopt makes use of the concepts appearing in the claim of Theorem 1.43.

The assumption of finite dimension simplifies heavily the technical tools needed in analysing many of the problems we will investigate, but this is not

the only reason why we make it. In fact, some of those problems themselves make only sense in the finite-dimensional setting. This is the case, for instance, for the questions raised in Chapter 3.

One of the immediate advantages of the finite-dimensional setting is that we have no longer to worry about specifying the topology carried by our vector space V . In fact, there is a natural topology that comes from the isomorphism with \mathbb{R}^d ($d = \dim V$). When we make use of topological concepts like closedness, this is the topology we are referring to. A well-known property of this kind of topologies is that a subset $K \subseteq V$ is compact iff it is closed and bounded.

Among the others, the theory of *convexity* is one of those example where the picture becomes particularly simple and intuitive in finite dimension. We remind the reader that a subset K of a real vector space V is called **convex** if $px + (1-p)y \in K$ whenever $x, y \in K$ and $p \in [0, 1]$. For an introduction to convex geometry that will suffice for our purposes, we refer the reader to [100]. Let us present here a brief review of the few properties we will need in the following.

A point x of a convex set K is called **extreme** if it can not be written as a nontrivial convex combination of other points in the set, i.e. if $x = py + (1-p)z$ for $y, z \in K$ and $p \in (0, 1)$ implies that $y = z = x$. The extreme points of a compact convex set K , collectively denoted as $E(K)$, have the remarkable property that their convex combinations cover the whole K , as expressed by the celebrated Minkowski-Carathéorodý theorem [101, 102].

Theorem 2.3 (Minkowski-Carathéorodý). [100, Theorem 8.11]. *Let $K \subseteq V$ be a compact convex subset of a finite-dimensional real vector space V of dimension d . Then every point of K can be written as a convex combination of at most $d+1$ extreme points of K . Moreover, if $x \in \text{int}(K)$ then for all extreme points $v_0 \in E(K)$ we can find $v_1, \dots, v_d \in E(K)$ such that $x = \sum_{i=1}^d \alpha_i v_i$ with $\alpha_0 > 0$, $\alpha_i \geq 0$ for all $i = 1, \dots, d$ and $\sum_{i=0}^d \alpha_i = 1$.*

We now move on to reviewing the basics of general probabilistic theories, or GPTs for short. Before we define rigorously the notion of GPT, a quick recap of the concepts discussed in Chapter 1 and of their specialisations to the finite dimensional case will help. The reader who went through Sections 1.2 and 1.6 of Chapter 1 can jump directly to Definition 2.6.

An **ordered vector space** (Section 1.2) is a real vector space V equipped with a **cone** C (also denoted by V_+) that defines the positive elements. A cone (Definition 1.2) is any subset $C \subset V$ that is: (i) closed under sums; (ii) closed under multiplications by positive scalars; and (iii) does not contain nontrivial vector subspace, or equivalently is such that $C \cap -C = \{0\}$. If $\text{span}(C) = C - C = V$, then C is called **spanning**. Most cones we will encounter in the rest of this thesis are spanning. The ordering on V is defined via $x \leq y \Leftrightarrow y - x \in C$. A positive vector $x \in C$ that is also internal to C , in symbols $x \in \text{int}(C)$, is called **strictly positive**, and we will write $x > 0$.

When V is an ordered vector space with spanning positive cone C , its **dual** V^* , which consists of all linear functionals $V \rightarrow \mathbb{R}$, can also be given the structure of an ordered vector space. The positive cone $C^* = V_+^*$, also called **dual cone**, is defined to be formed by all functionals $\varphi \in V^*$ that are positive on C , i.e. such that $\langle \varphi, x \rangle \geq 0$ for all $x \geq 0$ (which means $x \in C$). As usual, we use the notation $\langle \varphi, x \rangle := \varphi(x)$, which defines the canonical bilinear form $\langle \cdot, \cdot \rangle : V^* \times V \rightarrow \mathbb{R}$. In analogy with the nomenclature adopted for vectors, a functional $\varphi \in C^*$ is said to be **strictly positive** if it belongs to the interior of C^* , or equivalently if $\langle \varphi, x \rangle = 0$ for some $x \in C$ is possible only if $x = 0$.

As is easy to realise, dual cones are always closed and spanning. Furthermore, if a spanning cone is closed then it coincides with its double dual, i.e. $C = C^{**}$.

Let us take the chance to fix some more nomenclature. Since we are dealing with finite-dimensional spaces, V and V^* are always isomorphic. If V is equipped with a cone $C \subset V$, it makes sense to ask whether there is a linear isomorphism $T : V \rightarrow V^*$ such that $T(C) = C^*$. If this is the case, the cone C is said to be **weakly self-dual**. If C is weakly self-dual with the corresponding isomorphism T being induced by a nondegenerate scalar product on V , then C is called **strongly self-dual**.

Let us now come to the definition of a special class of positive functionals. Given an ordered vector space V with positive cone C , a positive functional $\varphi \in C^*$ that satisfies $\langle \varphi, x \rangle > 0$ for all nonzero $x \geq 0$ is called **strictly positive**. Strictly positive functionals are exactly those functionals that belong to the relative interior of C^* (i.e. the interior of C^* inside its linear span).

A **base** of a cone $C \subset V$ is a convex subset $\Omega \subset C$ such that for all $x \in C$ there is a unique real number $t \geq 0$ satisfying $x \in t\Omega$ (Definition 1.32). It turns out that a base of the positive cone C of an ordered vector space is always of the form

$$\Omega = \{x \geq 0 : \langle u, x \rangle = 1\}, \quad (2.5)$$

where $u \in V^*$ is a strictly positive functional (Lemma 1.33). Using the above characterisation, it is not difficult to show the following.

Lemma 2.4. [90]. *Let Ω be a base of a cone $C \subset V$, with V finite-dimensional. Then Ω is bounded. If C is in addition closed, then Ω is even compact. Furthermore, every finite-dimensional cone admits a bounded base, and, if closed, a compact one.*

Proof. Throughout this proof, we assume without loss of generality that the cone is spanning, i.e. that $V = C - C$. Let us start with the first claim. Let $\Omega \subset C \subset V$ be a base of a cone in dimension $d = \dim V$, and let $u \in \text{int}(C^*)$ be the strictly positive functional associated with Ω via (2.5). Then, it is not difficult to verify that u can be written as a strictly positive linear combination of positive functionals forming a basis, i.e. $u = \sum_{i=1}^d \alpha_i \varphi_i$ with $\varphi_i \in \text{int}(C^*)$ and $\alpha_i > 0$. This can be seen as follows. Take a ball centered on u and entirely contained inside $\text{int}(C^*)$. Pick functionals $\varphi_1, \dots, \varphi_d$ that: (i) lie on the boundary of said ball; and (ii) are such that $u \in \text{int}(\text{co}\{\varphi_1, \dots, \varphi_d\})$.

In particular, they must form a basis of the dual space. Now, by construction u can be written as a strictly positive linear combination of $\varphi_1, \dots, \varphi_d$, say $u = \sum_{i=1}^d \alpha_i \varphi_i$ with $\alpha_i > 0$ for all i . For $x \in \Omega$ we obtain $1 = \langle \varphi_i, x \rangle = \sum_{i=1}^d \alpha_i \langle \varphi_i, x \rangle$, that together with the positivity of each addend leads to $0 \leq \langle \varphi_i, x \rangle \leq 1/\alpha_i$ for all $i = 1, \dots, d$. Since all functionals φ_i (for $i = 1, \dots, d$) are bounded on Ω and together they form a basis of the dual space, Ω is bounded as a set.

Observe that since by (2.5) the base Ω is the intersection of C with a hyperplane, if C is closed the same is true for Ω . Since in finite dimension closed and bounded sets are compact, we obtain that bases of closed cones are necessarily compact.

As for the second part of the claim, it suffices to show that all finite-dimensional cones admit a base, i.e. by (2.5) a strictly positive functional. Start by observing that in finite dimension all spanning cones have internal points. Here, we are naturally concerned with the dual C^* of a cone $C \subset V$. Since C^* is spanning, it will contain a basis $\varphi_1, \dots, \varphi_d$ of V^* . Then it is easy to verify that $\sum_{i=1}^d \varphi_i$ belongs to the interior of C^* , thus is a strictly positive functional and as such induces a basis of C via (2.5). \square

Thanks to the above result, we can translate properties of compact sets into properties of cones (in finite dimension). See for instance the discussion in [100, XI]. A simple example of this strategy in action concerns the Minkowski-Carathéodory theorem (Theorem 2.3). Let us start with a simple definition. A nonzero element $x \in V_+$ of the positive cone of an ordered vector space V is called an **extremal vector** if for all $y \in V$ the inequalities $0 \leq y \leq x$ together imply that $y = \alpha x$ for some scalar α . In this case the set $\{\alpha x : \alpha \in \mathbb{R}\}$ is called an **extremal ray** of V_+ . Once a base Ω of V_+ has been fixed, and u is the corresponding strictly positive functional given by (2.5), it is not difficult to see that a nonzero $x \in V_+$ is an extremal vector of V_+ iff $\frac{x}{\langle u, x \rangle} \in \Omega$ is an extreme point of Ω as a convex set. For a more thorough discussion of these and related concepts, we refer the reader to [103, §1.6]. We can now deduce the following corollary of Theorem 2.3.

Corollary 2.5. *Let $C \subset V$ be a closed cone in finite dimension $d = \dim V < \infty$. Then every point of C can be written as a sum of at most d extremal vectors of C . Moreover, if $x \in \text{int}(C)$ then for all extremal vectors $v_0 \in C$ we can find extremal vectors v_1, \dots, v_k (where $k \leq d - 1$) such that $x = \sum_{i=1}^k v_i$.*

Proof. The claim follows immediately by putting together Theorem 2.3 and the above outlined correspondence between extremal vectors of V_+ and extreme points of Ω . \square

An ordered vector space whose positive cone is spanning and has a base Ω is called a **base norm space** (Definition 1.35) if its norm is given by

$$\|x\| = \inf \{ \langle u, x_+ + x_- \rangle : x = x_+ - x_-, x_{\pm} \geq 0 \} \quad (2.6)$$

for all $x \in V$, where the functional u is uniquely identified by the base Ω via (2.5). It is easy to verify that in a base norm space $\|x\| = \langle u, x \rangle$ whenever $x \geq 0$, and that in particular $\|\omega\| = 1$ for all $\omega \in \Omega$ (Lemma 1.37). The base norm is then seen to be additive on the positive cone.

The dual of a base norm space is a so-called **order unit space** (Definition 1.27), and its norm is given by

$$\|\varphi\|_* = \max \{t > 0 : \varphi \in t[-u, u]\}, \quad (2.7)$$

where $[-u, u] := \{\varphi \in V^* : -u \leq \varphi \leq u\}$, the ordering in the dual space being defined as usual with the help of the dual cone C^* . Observe that the above equation (2.7) defines a norm iff u is a strictly positive functional, i.e. if it is internal to the dual positive cone. It is clear from (2.7) that $[-u, u]$ is the dual unit ball. Observe that the maximum on the right-hand side of (2.7) exists because C^* is closed and thus so are all intervals. Let us now come to the rigorous mathematical definition of a general probabilistic theory.

Definition 2.6 (General probabilistic theories [28]). *A general probabilistic theory (GPT) is a finite-dimensional base norm space whose positive cone is closed.*

Note. Let us note in passing that in [28] GPTs are rather called *state spaces* (and no assumption of finite dimension is made). However, here we will reserve this latter term for the set Ω of normalised states.

In what follows, we will specify a GPT by giving the triple (V, C, u) , where: (i) V is the underlying vector space; (ii) C is its positive cone; and (iii) u is a strictly positive functional that identifies a base via (2.5). We will call u the **unit effect**. Usually, capital letters like A, B are reserved for GPTs. The base norm of a GPT A will be occasionally denoted by $\|\cdot\|_A$ when there is some form of ambiguity, for instance if there are multiple GPTs defined on the same vector space.

Now, Definition 2.6 does not tell us anything about the physics behind a GPT model. To be considered a valid physical theory, a mathematical object must be endowed with a set of rules that specify how to translate physical experiments into its language and how to extract predictions out of the mathematical machinery. These rules can be called collectively ‘physical interpretation’, since they connect the physical world with the mathematics we use to model it. We already discussed the most basic of those rules in Subsection 1.1.2, where we translated them to postulates on the μ probability function (Axioms 1, 2, 3). Let us briefly recap what we have established until now, and recast it into the language provided by Theorem 1.43. In the following, let (V, C, u) be a GPT.

- (I) The base Ω identified by (2.5) is the **state space** of the theory. States (or ensemble of states) of the physical system under examination are represented by elements of Ω .

- (II) An **effect**, i.e. a configuration of the measuring apparatus together with a pattern of outputs of the counters that constitute it is represented by a functional $e \in [0, u]$. Thus, a **measurement** is represented by a (finite) collection of effects $(e_i)_{i \in I} \subset [0, u]$ such that $\sum_{i \in I} e_i = u$. If $|I| = 2$ then the measurement is called **binary**.
- (III) Given a state $\omega \in \Omega$ and an effect $e \in [0, u]$, the real number $\langle e, \omega \rangle \in [0, 1]$ is the probability of registering the outcome specified by e when the system is measured (according to the procedure identified by e).
- (IV) If the preparation of the system in a state ω_i is conditioned on the outcome of a (discrete) random variable i , and after the preparation all the information on particular realisation of i is deleted, the resulting state of the system is represented by the convex combination $\bar{\omega} = \sum_i p_i \omega_i$. Analogously, if our apparatus measures the effect e_i conditioned on i the net measured effect is $\bar{e} = \sum_i p_i e_i$.

Some comments are in order. First, and most importantly, in the simple picture we just sketched only *destructive measurements* are taken into account. In other words, there is no rule in the above list that specifies how a system transforms *after* a measurement has been performed. We will see how this can be amended when we will discuss the definition of LOCC protocols, before the forthcoming Definition 2.8. However, including the post-measurement collapses in the picture can come at a cost, especially if one wants to impose further restrictions. An example of such a restriction is that the outcomes coming with probability one should not cause collapses. In [56, 57] it is shown that this rules out all GPTs whose state space is a polytope.

Second, in (ii) we considered only complete measurements, i.e. collections $(e_i)_{i \in I} \subset [0, u]$ such that the probabilities of detection add up to 1 for all input states, i.e. $\sum_{i \in I} \langle e_i, \omega \rangle = 1$ for all $\omega \in \Omega$ or equivalently $\sum_{i \in I} e_i = u$. This is no loss of generality, as any incomplete measurement can be completed by adding to the collection of effects the functional $u - \sum_{i \in I} e_i$, corresponding to the ‘error symbol’. If the reader is familiar with the basics of quantum information, the measurements as defined here correspond exactly to the so-called *positive operator-valued measurements* (POVM). On a different line, let us remark that for the sake of the presentation we restricted to finite alphabets when dealing with random variables or measurements. This poses no hurdle in principle, as all definitions make perfect sense when one uses instead integrals over measurable spaces.

There is at least another question that this very minimal scheme of a physical theory has left unanswered. We know that effects are associated with functionals $e \in [0, u]$, but *conversely, do all the functionals $e \in [0, u]$ correspond to physical effects?* And even more generally, *do all collections $(e_i)_{i \in I} \subset [0, u]$ correspond to physical measurements?* Here, a physical measurement is understood to be a configuration of the counters, and correspondingly an effect e_i that belongs to it represents one of the possible outputs of the counters. In the context of GPTs, this is usually called the *no-restriction hypothesis* [104].

Here, we are not going to discuss the status of this hypothesis, and we will limit ourselves to adding it as a separate assumption to the above list.

- (V) All collections $(e_i)_{i \in I} \subset [0, u]$ represent legitimate physical measurements. We shall denote the set of all such collections as \mathbf{M} .

In particular, the above assumption (v) implies that every effect $e \in [0, u]$ is actually physically measurable. In the notation of Theorem 1.43, this amounts to saying that $\Lambda = [0, u]$.

For the applications we have in mind, we will have a lot to do with restricted sets of measurements, i.e. subsets $\mathcal{M} \subseteq \mathbf{M}$. It is then useful to fix some nomenclature here. For a restricted set of measurements $\mathcal{M} \subseteq \mathbf{M}$, we will denote by $\langle \mathcal{M} \rangle$ the set generated by \mathcal{M} via **coarse graining**, i.e. by a posteriori declaring some of the outcomes of a measurement in \mathcal{M} as the same. In formula,

$$\langle \mathcal{M} \rangle := \left\{ (e_j)_{j \in J} : \exists I \text{ finite, } \{I_j\}_{j \in J} \text{ partition of } I : \right. \\ \left. (e_i)_{i \in I} \in \mathcal{M}, e_j = \sum_{i \in I_j} e_i \quad \forall j \in J \right\}. \quad (2.8)$$

We conclude this section by listing some alternative formulae to express the base norm (2.6) and its dual norm (2.7) associated with a GPT (V, C, u) . First, since the positive cone C is assumed to be closed, it is easy to see that the infimum in (2.6) can be replaced with a minimum. Moreover, we can apply the dual formula (1.4) for the norm of a Banach space to our special case and conclude that

$$\|x\| = \max_{\varphi \in [-u, u]} \langle \varphi, x \rangle \quad (2.9)$$

$$= \max_{e \in [0, u]} \{ |\langle e, x \rangle| + |\langle u - e, x \rangle| \}. \quad (2.10)$$

Apart from the expression (2.6), the dual norm of a base norm can be found directly by performing the maximisation

$$\|\varphi\|_* = \sup_{\omega \in \Omega} |\langle \varphi, \omega \rangle|, \quad (2.11)$$

as detailed in Corollary 1.39. In particular, we see that $\|\cdot\|_*$ is necessarily order monotone, i.e. $\varphi \leq \lambda$ implies $\|\varphi\|_* \leq \|\lambda\|_*$. In the finite-dimensional case, it is a simple and instructive exercise to translate minimisations like (2.6) to maximisations like (2.9) using the duality theory of convex programming [105].

2.2.3 Bipartite systems

Until now we only cared about modelling single systems. However, a lot of interesting physics emerges when one considers instead composite systems. Here we will focus only on bipartite systems, while for the treatment of larger composites we refer the reader to [41]. Clearly, if we want it to represent real physics,

the GPT machinery should encompass some kind of way to build a bipartite system $AB = (V_{AB}, C_{AB}, u_{AB})$ out of two single systems $A = (V_A, C_A, u_A)$ and $B = (V_B, C_B, u_B)$. Throughout this subsection, we will review the basic physical requirements to be imposed on such a construction.

First, preparing a composite state (ω_A, τ_B) should be possible for all local states $\omega_A \in \Omega_A$ and $\tau_B \in \Omega_B$. This ‘joining’ map must convex-linear in both arguments if the probabilistic interpretation (Assumption (IV) in Subsection 2.2.3) has to be respected. The same is naturally true for the effects, as all pairs of local measurements (e_A, f_B) must result in a legitimate global measurement, and the mapping must be convex-linear in both arguments. Moreover, the probability of measuring a combined effect (e_A, f_B) on a combined state (ω_A, τ_B) has to be given by $\langle e_A, \omega_A \rangle \langle f_B, \tau_B \rangle$. Finally, it is reasonable to postulate that ‘local statistics’ are sufficient to determine any state of the joint system, i.e. that from the set of numbers $\langle (e_A, f_B), \eta_{AB} \rangle$ one can perfectly reconstruct the state $\eta_{AB} \in \Omega_{AB}$. Conversely for the effects: measuring an effect on all ‘local states’ (ω_A, τ_B) should suffice to determine it completely.⁴ Summarising, we state the following additional axioms.

Axiom 4. There are convex-bilinear ‘composition maps’ $\mathcal{J} : \Omega_A \times \Omega_B \rightarrow \Omega_{AB}$ and $\mathcal{J}_* : [0, u_A] \times [0, u_B] \rightarrow [0, u_{AB}]$.

Axiom 5. For all states ω_A, τ_B and all effects e_A, f_B , it holds

$$\langle \mathcal{J}_*(e_A, f_B), \mathcal{J}(\omega_A, \tau_B) \rangle = \langle e_A, \omega_A \rangle \langle f_B, \tau_B \rangle .$$

Axiom 6 (Local tomography principle). Bipartite states are fully determined by the statistics resulting from local measurements, and conversely bipartite effects are fully determined by the statistics resulting from measuring local states.

Under these assumptions, the following can be shown [39, 40].

Lemma 2.7. *Let the finite-dimensional GPTs A, B together with their composite AB satisfy Axioms 4, 5, and 6. Then there are isomorphisms $\mathcal{F} : V_A \otimes V_B \rightarrow V_{AB}$ and $\mathcal{F}_* : V_A^* \otimes V_B^* \rightarrow V_{AB}^*$ such that:*

- (a) $\mathcal{F}(\omega_A \otimes \tau_B) = \mathcal{J}(\omega_A, \tau_B)$ for all states $\omega_A \in \Omega_A, \tau_B \in \Omega_B$;
- (b) $\mathcal{F}_*(e_A \otimes f_B) = \mathcal{J}_*(e_A, f_B)$ for all effects $e_A \in [0, u_A], f_B \in [0, u_B]$;
- (c) $\mathcal{F}_*(u_A \otimes u_B) = u_{AB}$; and
- (d) $\mathcal{F}_*^{-1} = \mathcal{F}^*$, where \mathcal{F}^* is the adjoint of \mathcal{F} .

⁴All this is very reasonable, but it does not mean that we can not come up with relevant examples of GPTs that do *not* satisfy this latter principle. This happens for instance when one considers real quantum mechanics, where the allowed density matrices are forced to be real symmetric instead of hermitian.

Since we are interested only in finite-dimensional GPTs, the proof of the above result is really elementary. We report it in Appendix C. Because of the above result, we will always identify the joint vector space V_{AB} with the tensor product of the two local spaces, i.e.

$$V_{AB} = V_A \otimes V_B. \quad (2.12)$$

Similarly, one has $u_{AB} = u_A \otimes u_B$, and facts (a), (b) and (d) in Lemma 2.7 tell us that we must add the following interpretation rule to the list in Subsection 2.2.2.

- (VI) If two physical systems with no previous interactions are prepared in locals states ω_A and τ_B , the resulting state of the system AB is represented by $\omega_A \otimes \tau_B$. Similarly, if measurements $(e_A^{(i)})_{i \in I} \subset [0, u_A]$ and $(f_B^{(j)})_{j \in J} \subset [0, u_B]$ are performed on any state of the global system, the resulting measurement is given by $(e_A^{(i)} \otimes f_B^{(j)})_{(i,j) \in I \times J}$.

As usual, particular care must be taken in ensuring that the resulting model is *non-signalling*, i.e. it does not allow instantaneous transmission of information between the two parties. As is easy to see, *within the GPT formalism the non-signalling principle is automatically implemented*: if AB is in a state ω_{AB} , whatever operation is carried out on the local party B , the reduced state of the system A will be given by the ‘partially evaluated’ vector $\langle u_B, \eta_{AB} \rangle \in \Omega_A$, implicitly defined by the equations

$$\langle \varphi_A, \langle u_B, \eta_{AB} \rangle \rangle := \langle \varphi_A \otimes u_B, \eta_{AB} \rangle \quad \forall \varphi_A \in V_A^*. \quad (2.13)$$

Until now we have only talked about the host space V_{AB} , saying nothing about the set of *states* we can legitimately prepare there. In other words, in order to specify the GPT $AB = (V_{AB}, C_{AB}, u_{AB})$ we have to tell what the positive C_{AB} is. Remarkably, C_{AB} is not fully determined by the above axioms, and moreover there seems to be no reasonable axiom we can posit to specify it completely. The reason why this is the case will be apparent from the zoo of examples we will examine in Section 2.3.

However, there are in fact two minimal requirements we have to impose on C_{AB} , and they descend directly from the above interpretation rule (VI). Namely, products of local states have to be legitimate global states, and since C_{AB} has to be convex we end up with requiring $C_A \otimes_{\min} C_B \subseteq C_{AB}$, where the **minimal tensor product** is defined by

$$C_A \otimes_{\min} C_B := \text{co}(C_A \otimes C_B), \quad (2.14)$$

where $C_A \otimes C_B = \{x \otimes y : x \in C_A, y \in C_B\}$. Since C_A, C_B are both closed cones, also $C_A \otimes_{\min} C_B$ can be shown to be closed. States in the minimal tensor product are usually called **separable**. This terminology comes from quantum

information, where the concept of separable states was originally defined in [91] (where those states are called ‘classically correlated’).

Naturally, a similar reasoning holds for at the level of effects, enforcing $C_A^* \otimes_{\min} C_B^* \subseteq C_{AB}^*$. We can deduce from this some information regarding the primal cone C_{AB} by taking the dual of both sides, and thus reversing the sense of the inclusion. We obtain $C_{AB} \subseteq C_A \otimes_{\max} C_B$ with

$$C_A \otimes_{\max} C_B := \{Z \in V_A \otimes V_B : \langle \varphi \otimes \lambda, Z \rangle \geq 0 \quad \forall \varphi \in C_A^*, \lambda \in C_B^*\} \quad (2.15)$$

$$= \left(C_A^* \otimes_{\min} C_B^* \right)^*, \quad (2.16)$$

where the equivalence of the above two definitions of **maximal tensor product** is elementarily established. Accordingly, one has also

$$\left(C_A \otimes_{\max} C_B \right)^* = C_A^* \otimes_{\min} C_B^*. \quad (2.17)$$

In what follows, borrowing the terminology from quantum information, functionals in the maximal tensor product are usually called **entanglement witnesses** or simply **witnesses**. The original definitions of minimal and maximal tensor products seem to go back to Peressini and Sherbert [106] and independently to Hulanicki and Phelps [107]. Both works were written toward the end of the Sixties, when the theory of ordered Banach spaces was reaching a certain degree of maturity. Substantial progress is due to Namioka and Phelps [8], as we shall see in Section 2.5. Summarising, we have established that the positive cone of the composite system must obey the two-sided bound

$$C_A \otimes_{\min} C_B \subseteq C_{AB} \subseteq C_A \otimes_{\max} C_B. \quad (2.18)$$

Minimal and maximal tensor products of cones will be the main subject of study in Section 2.5.

With a slight abuse of notation, given two GPTs $A = (V_A, C_A, u_A)$ and $B = (V_B, C_B, u_B)$, we will refer to the composites

$$A \otimes_{\min} B := \left(V_A \otimes V_B, C_A \otimes_{\min} C_B, u_A \otimes u_B \right), \quad (2.19)$$

$$A \otimes_{\max} B := \left(V_A \otimes V_B, C_A \otimes_{\max} C_B, u_A \otimes u_B \right), \quad (2.20)$$

as the minimal and maximal tensor product of the GPTs A and B , respectively.

2.2.4 Classes of measurements in bipartite GPTs

Now that we have the construction of bipartite GPTs in hand, we can define some interesting classes of measurements on a composite (bipartite) system. As usual, the intuition we have developed in studying quantum mechanics (assuming such an intuition can exist at all) can guide us in giving operationally

meaningful definitions. When dealing with bipartite quantum systems, some restricted sets of protocols come into play quite naturally as deriving from operational constraints. Examples of such sets include:

- local operations (LO), i.e. protocols that can be realised when both parties have full control over their local systems but are forbidden to communicate in any way when they can still access the systems (communication is allowed afterwards);
- local operations assisted by shared randomness (LOSR), when besides local operations also previous sharing of classically correlated random variables is allowed;
- local operations assisted by one-way classical communication ($\text{LOCC}_{\rightarrow}$), when one gives in addition the possibility of using a one-way classical communication device (say a television); and finally
- general LOCC protocols, when local operations are assisted by two-way classical communication (we give the two parties, say, a phone).

For the following comment, we are going to assume that the reader is familiar with basic concepts of quantum information such as the formalisation of general measurements as POVM and the notion of separability. We will anyway provide rigorous definition in the forthcoming Section 2.3. Once the above classes of protocols have been introduced, it is also convenient to introduce corresponding mathematical relaxations that can turn out to be much easier to characterise. For instance, consider the set of measurements $(e_i)_{i \in I}$ such that E_i is a separable positive operators for all $i \in I$. We call these measurements *separable*, and denote them collectively as SEP. It is easy to see that $\text{LOCC} \subseteq \text{SEP}$, and less trivially the inclusion can be shown to be strict. For a comprehensive review of these questions, we refer the reader to [108].

Until now we have discussed only quantum theory. Perhaps surprisingly, it turns out that all these restricted classes of measurements can be defined in a bipartite arbitrary GPT $AB = (V_A \otimes V_B, C_{AB}, u_A \otimes u_B)$ constructed out of two local theories A and B in such a way that the constraints (2.18) are met. Before we provide general definitions below, let us briefly discuss how to add *dynamical prescriptions* to our mathematical formalism. The purpose of these rules is to specify how states transform after a measurement, extending the picture we have been describing so far, mostly orientated towards the outcomes and their probabilities. In practice, we will not make use of these prescriptions, and in fact our results are totally independent of any assumption concerning them beyond the mere consistency with the operational interpretation of the theory. However, this apparatus is needed to define a generic LOCC protocol, which requires multiple, interactive rounds of operations on the same systems.⁵

⁵I thank Howard Barnum and Matthias Christandl for an enlightening discussion on this point, from which the forthcoming definition of LOCC protocols in an arbitrary GPT ultimately came out.

Following [28], we can define **instruments** on one of the two system, say $A = (V_A, C_A, u_A)$, as collections $(\phi_i)_{i \in I}$ of linear maps $\phi_i : V_A \rightarrow V_A$ that are **completely positive**, i.e. satisfy $((\phi_i)_A \otimes I_B)(C_{AB}) \subseteq C_{AB}$, and sum up to a normalisation-preserving map, i.e. $\sum_{i \in I} \phi_i^*(u_A) = u_A$, with $\phi_i^* : V_A^* \rightarrow V_A^*$ being the adjoint of ϕ_i (B.1). A totally analogous definition can be given for instruments on the B system. In the operational interpretation of the theory, an instrument describes a non-destructive measurement, with $\phi_i(\omega)$ representing the unnormalised post-measurement state when the outcome i has been recorded on the initial state ω , and the normalisation coefficient $\langle u, \phi_i(\omega) \rangle = \langle \phi_i^*(u), \omega \rangle$ being the probability that the process yields the outcome i (accordingly, observe that $(\phi_i^*(u))_{i \in I}$ is a valid measurement in the GPT sense thanks to the interpretation rule (V) in Subsection 2.2.2).

With the concept of instrument at hand, in order to define LOCC protocols we can follow the steps described in [108, §2.2]. We will not repeat the construction here since it is totally analogous to the quantum mechanical one once the concept of instrument in GPTs has been clarified, and it is not indispensable for the applications we have in mind.

Definition 2.8. *Let $A = (V_A, C_A, u_A)$ and $B = (V_B, C_B, u_B)$ be two GPTs, and let the composite system $AB = (V_A \otimes V_B, C_{AB}, u_A \otimes u_B)$ satisfy (2.18). Denote by $P(I)$ the set of probability distributions on a finite alphabet I . Then local operations (LO), local operations assisted by shared randomness (LOSR), one-way classical communication (LOCC $_{\rightarrow}$) or two-way classical communication (LOCC), and separable measurements (SEP) are subsets of the set \mathbf{M}_{AB} of all measurements on AB given by:*

$$\text{LO} := \{(e_i \otimes f_j)_{(i,j) \in I \times J} : (e_i)_{i \in I} \in \mathbf{M}_A, (f_j)_{j \in J} \in \mathbf{M}_B\}, \quad (2.21)$$

$$\begin{aligned} \text{LOSR} := \{ & (p_i e_j^{(i)} \otimes f_k^{(i)})_{(i,j,k) \in I \times J \times K} : I \text{ finite, } p \in P(I), \\ & (e_j^{(i)})_{j \in J} \in \mathbf{M}_A, (f_k^{(i)})_{k \in K} \in \mathbf{M}_B \quad \forall i \in I\}, \end{aligned} \quad (2.22)$$

$$\text{LOCC}_{\rightarrow} := \{(e_i \otimes f_j^{(i)})_{(i,j) \in I \times J} : (e_i)_{i \in I} \in \mathbf{M}_A, (f_j^{(i)})_{j \in J} \in \mathbf{M}_B \quad \forall i \in I\}, \quad (2.23)$$

$$\text{LOCC} := \{(\Phi_i^*(u_A \otimes u_B))_{i \in I} : (\Phi_i)_{i \in I} \text{ LOCC instrument on } AB\} \quad (2.24)$$

$$\text{SEP} := \{(e_i \otimes f_i)_{i \in I} \in \mathbf{M}_{AB}\}. \quad (2.25)$$

The above sets will be collectively called **locally constrained sets of measurements**.

It is easy to verify that

$$\text{LO} \subseteq \text{LOSR} \subseteq \text{LOCC}_{\rightarrow} \subseteq \text{LOCC} \subseteq \text{SEP}. \quad (2.26)$$

The last inclusion is slightly less trivial than the others, but its proof follows closely the quantum mechanical one. Namely, referring for details and

nomenclature to [108], one can observe that: (1) one-way local instruments are separable, in the sense each component is a positive sum tensor products of completely positive maps; (2) coarse-graining preserves separability; (3) an instrument that is LOCC-linked to a separable one is again separable; (4) separability is preserved under limits; and finally (5) if $(\Phi_i)_{i \in I}$ (acting on AB) is separable as an instrument, $(\Phi_i(u_A \otimes u_B))_{i \in I}$ is separable as a measurement.

Remark. Let us stress here that *LOCC is the only locally constrained set of measurements that depends explicitly on the choice of the positive cone C_{AB} of the bipartite system.* In fact, it is easy to realise that only the structure of the local GPTs appears in (2.21), (2.22), (2.23), and (2.25). Following the discussion before Definition 2.8, we see that this dependence is hidden inside the concept of completely positive map, in turn necessary to define local instruments.

2.3 Examples

Until now, we have been developing the mathematical machinery of the GPT framework in full generality. However, time has come to see it in action with specific examples. Apart from the obvious goal of familiarising the reader with the formalism, the purpose of doing this is twofold. First, we will convince ourselves that the realm of GPTs is sufficiently rich, as to encompass not only all physically relevant theories, but also nearly all physically conceivable examples. This is to be expected, as the assumptions we made so far are very minimal. Secondly, we will also set the stage for more elaborate constructions that refer only to particular examples of GPTs or classes of GPTs. These constructions obviously come at the price of a loss of generality, but on the one hand they can still give valuable insights into general behaviours, and on the other hand they can tell us something about models we care about particularly because of their physical relevance.

2.3.1 Classical probability theory

The simplest instance of a GPT is undoubtedly classical probability theory on a finite alphabet. Although it is trivial in some respects, its understanding is crucial in appreciating more complicated examples to be treated throughout the rest of this section.

The state space of classical probability theory is the set of probability distributions over a finite alphabet. The corresponding GPT can be defined as the triple

$$\text{Cl}_d := (\mathbb{R}^d, \mathbb{R}_+^d, u), \quad (2.27)$$

where $\mathbb{R}_+^d := \{x \in \mathbb{R}^d : x_i \geq 0 \forall i = 1, \dots, d\}$ is the positive octant, and the unit effect acts as $\langle u, y \rangle = \sum_{i=1}^d y_i$ for all $y \in (\mathbb{R}^d)^* \simeq \mathbb{R}^d$. The base norm associated with classical probability theory is easily seen to be the l_1 -norm $|x|_1 = \sum_{i=1}^d |x_i|$.

Of course, what we have described so far is the standard form of a classical theory. Other GPTs can be obtained from it via the applications of linear isomorphisms. Since we will be interested in these models in the following, it is worth it to give the following definition.

Definition 2.9. A *simplicial cone* in a real vector space V of finite dimension d is by definition the set of non-negative linear combinations of d linearly independent vectors of V . A GPT (V, C, u) is called **classical** if the cone of states C is simplicial, i.e. if the ordered vector spaces (V, C) and $(\mathbb{R}^d, \mathbb{R}_+^d)$ (where $d = \dim V$) are linear- and order-isomorphic (see Subsection 1.7.9 for a definition of order isomorphism).

Let us devote a moment to comment on some elementary properties of simplicial cones. First of all, simplicial cones are closed and spanning by definition. Moreover, the dual cone C^* of a closed and spanning cone C is simplicial iff C is itself simplicial. Finally, if the dimension of the host space is low enough, namely $d \leq 2$, then every closed cone is simplicial [103, §2.1, Exercise 12].

We now come back to the standard probability theory as given by (2.27). Until now, we have dealt with a single system. As for bipartite systems, let us start by observing that if $A = \text{Cl}_{d_A}$ and $B = \text{Cl}_{d_B}$ the vector space pertaining to the joint system AB , given by the rule (2.12), is easily seen to be

$$\mathbb{R}^{d_A} \otimes \mathbb{R}^{d_B} = \mathbb{R}^{d_A d_B} . \quad (2.28)$$

Moreover, it can be seen by direct inspection that the lower and upper bound in (2.18) coincide, hence the composition rule is trivial: if $A = \text{Cl}_{d_A}$ or $B = \text{Cl}_{d_B}$, then necessarily

$$C_{AB} = C_A \otimes_{\min} C_B = C_A \otimes_{\max} C_B = \mathbb{R}_+^{d_A d_B} , \quad (2.29)$$

or in other words

$$AB = A \otimes_{\min} B = A \otimes_{\max} B = \text{Cl}_{d_A d_B} . \quad (2.30)$$

In Section 2.5 we will see that this is a particular example of a more general fact: *as soon as at least one of the two GPTs is a classical probability theory, the minimal and maximal tensor products always coincide* (Lemma 2.27)

2.3.2 Quantum mechanics

We now turn our attention to describing finite-dimensional quantum systems from the point of view of GPTs. It goes without saying, quantum mechanics is the most important example of GPT and the archetypical example of a non-classical theory. In what follows, we will describe a system with n -level. Vectors in \mathbb{C}^n ('pure states') will be represented in Dirac notation as $|\alpha\rangle, |\beta\rangle$ and so on. The canonical hermitian product is given by $\langle\alpha|\beta\rangle = \sum_{i=1}^n \alpha_i^* \beta_i$, and according to the physicists' convention it is antilinear in the first entry and linear in the second.

As is well-known, the quantum mechanical cone of states (mixed states) is composed of the positive semidefinite $n \times n$ matrices (collectively denoted by PSD_n), embedded in the real space of hermitian matrices (called \mathcal{H}_n) whose real dimension is $d = n^2$. Since density matrices, which represent quantum mechanical states, are positive matrices with trace one, the unit effect is easily seen to coincide with the trace. Therefore, we will write symbolically

$$\text{QM}_n := (\mathcal{H}_n, \text{PSD}_n, \text{Tr}), \quad (2.31)$$

remembering that

$$\dim \text{QM}_n = n^2. \quad (2.32)$$

The extremal points in the set of quantum mechanical states are exactly the (normalised) rank-one orthogonal projectors $|\psi\rangle\langle\psi|$, called **pure states**.

Observe that the positive semidefinite cone is strongly self-dual, i.e. $\text{PSD}_n^* = \text{PSD}_n$, when one endows \mathcal{H}_n with the natural **Hilbert-Schmidt scalar product** given by $\langle X, Y \rangle := \text{Tr } XY$. The base norm in quantum mechanics can be proven to coincide with the **trace norm** $\|X\|_1 = \text{Tr } |X| = \sum_{i=1}^n |\lambda_i(X)|$, where $\lambda_i(X)$ are the eigenvalues of $X \in \mathcal{H}_n$ and the last equality holds because X is hermitian.

Now, let us discuss the composition rules for bipartite systems. First of all, it is easy to see that the real vector spaces of hermitian matrices combine according to a very simple rule under tensor products (as prescribed by (2.12)):

$$\mathcal{H}_{n_A} \otimes \mathcal{H}_{n_B} = \mathcal{H}_{n_A n_B}. \quad (2.33)$$

The fact that there is equality above is really a special feature of hermitian matrices, and comes directly from a computation of the dimensions. This is not the case, for instance, for real symmetric matrices, for which we have a strict inclusion of the form $\mathcal{S}_{n_A} \otimes \mathcal{S}_{n_B} \subsetneq \mathcal{S}_{n_A n_B}$. To see this, it suffices to compute the dimensions: one gets $\frac{n_A(n_A+1)}{2} \frac{n_B(n_B+1)}{2}$ for the left-hand side and $\frac{n_A n_B (n_A n_B + 1)}{2}$ (which is strictly larger as soon as $n_A, n_B > 1$) for the right-hand side. This is the reason why real quantum mechanics, where all density matrices are prescribed to be real symmetric rather than complex hermitian, does not respect the local tomography principle (Axiom 6) and thus does not satisfy (2.12).

As for the positive cone of a bipartite quantum system, using the definitions (2.14) and (2.15), we see that

$$\text{PSD}_{n_A} \otimes_{\min} \text{PSD}_{n_B} = \left\{ \sum_{i \in I} P_i \otimes Q_i : I \text{ finite}, \right. \\ \left. P_i \in \text{PSD}_{n_A}, Q_i \in \text{PSD}_{n_B} \forall i \in I \right\}, \quad (2.34)$$

$$\text{PSD}_{n_A} \otimes_{\max} \text{PSD}_{n_B} = \left\{ W \in \mathcal{H}_{n_A n_B} : \text{Tr}[P \otimes Q W] \geq 0 \right. \\ \left. \forall P \in \text{PSD}_{n_A}, Q \in \text{PSD}_{n_B} \right\}. \quad (2.35)$$

In quantum information, states belonging to (2.34) are called **separable** (originally, **classically correlated** [91]), while elements of (2.35) are variously called **entanglement witnesses**, **separability witnesses** or **block-positive operators**. This latter name comes from the fact that in (2.35) we can restrict P and Q to be pure states (i.e. rank-one projectors), and the defining condition for belonging to the set amounts to impose the positivity of the diagonal blocks in all product bases. Interestingly enough, Nature has a preferred choice for the cones of bipartite systems, which is neither the maximal nor the minimal tensor product. Instead, if $A = \text{QM}_{n_A}$ and $B = \text{QM}_{n_B}$ then $AB = \text{QM}_{n_A n_B}$, i.e.

$$\text{PSD}_{n_A} \otimes_{\min} \text{PSD}_{n_B} \subsetneq C_{AB} = \text{PSD}_{n_A n_B} \subsetneq \text{PSD}_{n_A} \otimes_{\max} \text{PSD}_{n_B}. \quad (2.36)$$

The fact that all the three above inclusions are strict is an elementary result in entanglement theory, whose proof we do not report here. We refer the interested reader to [109] or to the monograph [110]. Let us stress here that quantum mechanics is our first example of a GPT whose composition rules prescribe the bipartite cone to be neither the minimal nor the maximal tensor product of the local cones.

In fact, we can let ourselves inspire by the above observation and use it to define a new composition rule that allows only separable bipartite states. Such a construction is mentioned in [50, §IV.E] and considered at a purely hypothetical level. We take the trouble to write out the definition because we will find this example useful and instructive later.

Definition 2.10. *We call **W-theory** the class of GPT models that can be obtained from the family $\{\text{QM}_n\}_{n \in \mathbb{N}}$ of single quantum systems by taking minimal tensor products according to the rule (2.19).*

In *W-theory*, the only allowed states of a multipartite system are fully separable, while the set of possible effects contains all entanglement witnesses (equivalently, all elements in the cone (2.35)). Thus, the base norm of a hermitian operator $X_{AB} \in \mathcal{H}_{n_A n_B}$ will be given by

$$\begin{aligned} \|X\|_W &:= \|X\|_{\text{QM}_{n_A} \otimes_{\min} \text{QM}_{n_B}} \\ &= \max \left\{ \text{Tr} XY : |\langle \alpha \beta | Y | \alpha \beta \rangle| \leq \langle \alpha | \alpha \rangle \langle \beta | \beta \rangle \quad \forall |\alpha\rangle \in \mathbb{C}^{n_A}, |\beta\rangle \in \mathbb{C}^{n_B} \right\}, \end{aligned} \quad (2.37)$$

where the last line corresponds to the formula (2.9).

2.3.3 Centrally symmetric models

Until now, we have examined standard examples of GPTs that have been previously considered in the literature. Throughout this subsection, we want to introduce a novel class of models whose special status has been somehow neglected in the past. This class of so-called ‘centrally symmetric’ theories are

characterised by their state spaces Ω being invariant under ‘inversion’. We will see what this means precisely in a moment.

Restricting the analysis to centrally symmetric models, as we will do under some circumstances in Section 2.5, has undoubtedly some drawbacks, the most serious one being that while doing so we are excluding by default physically relevant examples such as classical probability theory and quantum mechanics (Subsections 2.3.1 and 2.3.2, respectively). However, there are also lots of reasons why this is both instructive and meaningful to do this. First, we still have at our disposal an ample family of theories that can be used to test conjectures, because they are theoretically simpler, and optimality of certain bounds, because they are computationally convenient. We will see examples of both approaches in action in Section 2.5 and Chapter 3. Secondly, centrally symmetric models uncover a surprising connection between GPTs and a certain branch of functional analysis that deals with tensor product of Banach spaces. In some cases this connection can be exploited in full generality, and this idea will be crucial in finding the main results of Chapter 3. Let us start with the following definition.

Definition 2.11 (Centrally symmetric models). *A GPT of the form (\mathbb{R}^d, C, u) such that $u = (1, 0, \dots, 0)^T$ is said to be **centrally symmetric** if there exists a norm $|\cdot|$ on \mathbb{R}^{d-1} such that*

$$C = \{(x_0, \bar{x}) \in \mathbb{R} \oplus \mathbb{R}^{d-1} : x_0 \geq |\bar{x}|\}. \quad (2.38)$$

Note. It is somewhat unusual to denote a vector norm by $|\cdot|$. This choice is made to distinguish this latter object, which is a norm on \mathbb{R}^{d-1} , from the base norm $\|\cdot\|$, which has the global space \mathbb{R}^d as its domain.⁶

As can be easily verified, the dual cone to a ‘centrally symmetric’ cone as in (2.38) shares the same structure, being given by

$$C^* = \{(y_0, \bar{y}) \in \mathbb{R} \oplus \mathbb{R}^{d-1} : y_0 \geq |\bar{y}|_*\}, \quad (2.39)$$

where $|\cdot|_*$ is the dual to the norm $|\cdot|$. With a straightforward calculation via (2.9), the base norm of a centrally symmetric GPT can be easily seen to be given by

$$\|x\| = \|(x_0, \bar{x})\| = \max\{|x_0|, |\bar{x}|\}. \quad (2.40)$$

Another peculiarity of centrally symmetric models is the existence of a privileged state, denoted by $u_* := (1, 0, \dots, 0)^T$. The vector space can then be written as

$$\mathbb{R}^d = \mathbb{R}u_* \oplus \mathbb{R}^{d-1}, \quad (2.41)$$

the second component being sometimes referred to as the **section** of the host vector space \mathbb{R}^d . We will keep calling x_0, \bar{x} the two components of $x \in \mathbb{R}^d$ according to the above decomposition. Along the same lines, for $v \in \mathbb{R}^{d-1}$ we

⁶Some mathematicians would rather introduce a third bar, preferring the notation $|||\cdot|||$, which to us looks however too outrageous to be considered.

will write $\widehat{v} := 0 \oplus v$. A remarkable feature of centrally symmetric models is the existence of a simple linear map $T : \mathbb{R}^d \rightarrow \mathbb{R}^d$, given by

$$T := 1 \oplus (-I_{d-1}) \quad (2.42)$$

according to the decomposition (2.41) and with I_{d-1} denoting the identity on \mathbb{R}^{d-1} . Observe that T is an isomorphism and in fact also an order isomorphism (i.e. it is such that $T(C) = C$). Consequently, for all $x \in \mathbb{R}^d$ one has $\|x\| = \|T(x)\|$ (also obvious from the explicit formula for the base norm given above). Now, let us turn to bipartite systems. From (2.41) we infer the natural decomposition

$$\begin{aligned} V_{AB} &= \mathbb{R}^{d_A} \otimes \mathbb{R}^{d_B} \\ &= (\mathbb{R} u_{A*} \otimes u_{B*}) \oplus (u_{A*} \otimes \mathbb{R}^{d_B-1}) \oplus (\mathbb{R}^{d_A-1} \otimes u_{B*}) \oplus (\mathbb{R}^{d_A-1} \otimes \mathbb{R}^{d_B-1}). \end{aligned} \quad (2.43)$$

Tensors belonging to the bipartite vector space $\mathbb{R}^{d_A} \otimes \mathbb{R}^{d_B}$ (or to its dual) can be thought of as $d_A \times d_B$ matrices $Z \in \mathbb{R}^{d_A \times d_B}$. We shall find useful to denote by \bar{Z} the $(d_A - 1) \times (d_B - 1)$ submatrix of $Z \in \mathbb{R}^{d_A \times d_B}$ which is obtained by cutting off the zeroth components Z_{i0}, Z_{0j} of the latter. Complementarily, given $M \in \mathbb{R}^{(d_A-1) \times (d_B-1)}$ we call \widehat{M} the $d_A \times d_B$ ‘lifted’ matrix whose entries are

$$\widehat{M}_{ij} := \begin{cases} 0 & \text{if } i = 0 \text{ or } j = 0, \\ M_{ij} & \text{if } i, j \geq 1. \end{cases}$$

In what follows, the tensor product of the two unit functionals will be denoted by $U := u_A \otimes u_B$. Also the composite system hosts a ‘privileged state’ inherited by the two local spaces, namely $U_* := u_{A*} \otimes u_{B*}$.

As for the cone C_{AB} of bipartite states, there is no a priori preferred rule to compose two centrally symmetric models, so we will use different ones depending on the circumstances.

Now, there is a natural example of centrally symmetric model to which we want to devote special attention. Our starting point is the following elementary observation: as is well-known, the state space of QM_2 is identifiable with a 3-dimensional ball (Bloch sphere). We can let ourselves inspire by this for defining a hypothetical class of physical models whose state space is a Euclidean ball of arbitrary dimension. These GPTs have been considered recently in connections to attempts of reconstructing quantum mechanics starting from few physically motivated axioms [67, 74, 70]. This is to be expected in light of a famous classification theorem by Koecher and Vinberg [111, 112], implying among other things that spherical models are one of the few classes of models that are strongly self-dual, the most notable of which is exactly quantum mechanics [41].

In the following, given $x \in \mathbb{R}^d$ we call x_i its i th entry ($i = 0, \dots, d-1$), while \bar{x} will denote the column vector obtained from x by eliminating the zeroth component. The GPT corresponding to this **spherical model** has the form

$$\text{Sph}_d := (\mathbb{R}^d, C_d, u). \quad (2.44)$$

Here, C_d is the Lorentz or ‘ice-cream’ cone

$$C_d := \{x \in \mathbb{R}^d : |\bar{x}|_2 \leq x_0\}, \quad (2.45)$$

where $|\cdot|_2$ is the standard Euclidean norm in \mathbb{R}^{d-1} , and u is the functional defined by $\langle u, y \rangle := y_0$ for all $y \in (\mathbb{R}^d)^*$. In light of the canonical identification $(\mathbb{R}^d)^* \simeq \mathbb{R}^d$, we find it convenient to adopt from now on a column notation for the dual as well as for the primal space. Within this convention, we shall write unambiguously $u = (1, 0, \dots, 0)^T$. This simplification of the notation is going to pay off because just like quantum mechanics, also the spherical model is strongly self-dual, i.e. with above conventions $C_d = C_d^*$. As for the base norm in Sph_d , we can specialise (2.40) to our particular case and obtain

$$\|x\|_{\text{Sph}_d} = \|(x_0, \bar{x})\|_{\text{Sph}_d} = \max\{|x_0|, |\bar{x}|_2\}. \quad (2.46)$$

2.3.4 Generalised non-signalling theories

In this subsection, we will discuss standard examples of GPTs known as generalised non-signalling theories [50]. The usefulness of these models is that every experimental setting can be used to map an arbitrary physical state into a state of such a theory. Therefore, in some sense, every experimentally accessible feature of a physical theory should be analysable via these mappings. And in fact the class of generalised non-signalling theories is sufficiently rich so that this is typically what happens.

There is another good reason to care about these models. Namely, concepts like Bell inequalities and non-locality admit a very natural formulation within the formalism of GPTs applied to the specific case of generalised non-signalling theories. We especially like this formulation because of the crystal clear mathematical representation of the phenomenon of nonlocality that it gives, and not least because of the compactness of the notation, free from complicated sums running over many indices. We hope that this work will contribute to spreading the word and making the usage of this framework more common among quantum information theorists.⁷

A generalised bit, or **gbit** for short [50], is one of the simplest example of a non-classical system. As a GPT, we can define it as $G_2 := (\mathbb{R}^3, C_\square, u)$, where

$$C_\square := \{(x, y, z) \in \mathbb{R}^3 : |x| + |y| \leq z\} \quad (2.47)$$

⁷Almost all the time I read a paper on nonlocality, I personally really feel the need of a more compact notation. The situation somewhat reminds me of the times when linear algebra was a great deal a matter of writing down and manipulating linear systems with spelled-out coefficients. The gradual introduction, throughout the 19th century, of objects called matrices, not only simplified the notation considerably, but shed also new light on the conceptual meaning of linear systems and linear substitutions. My favourite instance of this phenomenon that one may call of notation compactification occurred however in electromagnetism. At the dawn of the modern theory the equations of the electromagnetic field were written *in components*, which makes them significantly harder to read, visualise, and ultimately manipulate. Maxwell’s paper ‘*A dynamical theory of the electromagnetic field*’, for instance, features long lists of scalar equations instead of the compact vector equations we all know and love. Of course, these secondary issues did not stop Maxwell from gaining a full command of the subtleties behind said equations, but this is another story.

and $\langle u, (x, y, z) \rangle := z$, or equivalently $u = (0, 0, 1)^T$. As is easily seen, the state space of G_2 is a two-dimensional square. Thus, a straightforward generalisation of the above concept takes as state space a Cartesian product of n identical k -simplices, in symbols $\Omega_n^k := \Delta_k^{\times n}$ [50]. Here, a k -simplex is just the convex space of all probability distributions on an alphabet of k symbols (its linear dimension is thus $k - 1$, the missing degree of freedom coming from the normalisation condition). The corresponding GPT will be denoted by G_n^k , and the associated cone by C_n^k . The underlying vector space is naturally identified with a subspace of $\mathbb{R}^{n \times k}$ of dimension $\dim G_n^k = n(k - 1) + 1$. States of G_n^k can thus be identified with $n \times k$ matrices of probabilities $P = (p_{ij})_{i=1, \dots, n}^{j=1, \dots, k} \in \mathbb{R}^{n \times k}$. With this notation, $(p_{i0}, \dots, p_{ik})^T$ is nothing but the component of $P \in \Delta_k^{\times n}$ corresponding to the i -th copy of the simplex Δ_k .

This family of GPTs is particularly useful because any state of an arbitrary GPT $A = (V, C, u)$ can be used to generate a state of G_n^k with the help of a *measurement apparatus*. We think of any such apparatus as a measuring device with a selectable setting, indexed by $i = 1, \dots, n$. Any particular value i of the setting index performs a given n -outcome measurement (e_{i0}, \dots, e_{ik}) on an input state $\omega \in \Omega$ of A . We can then give the following definition.

Definition 2.12 (Apparatuses). *Given a GPT $A = (V, C, u)$ and two positive integers n, k , an (n, k) -**apparatus** on A is a linear map of the form*

$$\begin{aligned} \Gamma : V &\longrightarrow \mathbb{R}^{n \times k} \\ x &\longmapsto ((e_{ij}, x))_{i=1, \dots, n}^{j=1, \dots, k}, \end{aligned} \quad (2.48)$$

where $(e_{ij})_j \in \mathbf{M}$ is a measurement on A for all $i = 1, \dots, n$. Observe that Γ maps states of A to states of G_n^k .

Particularly appealing for its simplicity is the case $k = 2$, with the corresponding state space being an n -dimensional hypercube. We will call this latter example $G_n^2 =: G_n$ **cubic model**. We remark here that *cubic models are centrally symmetric in the sense of Subsection 2.3.3*, hence all the theory we developed and notation we fixed there applies also in this case. In particular, the base norm in G_n can be easily deduced from (2.40), and becomes

$$\|x\|_{G_n} = \|(x_0, \bar{x})\|_{G_n} = \max\{|x_0|, |\bar{x}|_\infty\}, \quad (2.49)$$

where for $v \in \mathbb{R}^n$ one has $|v|_\infty := \max_{1 \leq i \leq n} |v_i|$.

The main reason for introducing these toy models is that they provide a useful tool in understanding nonlocal correlations. In fact, the celebrated Popescu-Rohrlich (PR) box [48] can be seen as states of the maximal tensor product $G_2 \otimes_{\max} G_2$. In general, GPTs of the class G_n^k equipped with the maximal tensor product composition rule are called ‘generalized non-signalling theories’ (GNST) in [50]. Within this framework, nonlocality can be formulated somewhat more easily. Let us give the following definition.

Definition 2.13 (Bell inequalities). A **Bell inequality** with n measurements having k outcomes each is a linear inequality that is obeyed by all states in $G_n^k \otimes_{\min} G_n^k$ but violated by some states in $G_n^k \otimes_{\max} G_n^k$.

The above definition makes sense because probability distributions in the minimal tensor product are exactly those that are reproducible via a so-called *local hidden variable model*. Thus a Bell inequality separates local hidden variable models from probability patterns that are signatures of genuine nonlocal phenomena. Providing a summary of the extensive literature on the subject is almost impossible, so we refer the interested reader to the excellent review in [113].

Naturally, in a realistic scenario, correlations are produced by means of physical states. Therefore, we should have a way to refer the concept of violation of a Bell inequality to states rather than to abstract state spaces such as G_n^k . This goal is accomplished by the following definition, with which we close this subsection.

Definition 2.14. Given two local GPTs $A = (V_A, C_A, u_A)$ and $B = (V_B, C_B, u_B)$ and a way to combine them into a bipartite GPT $AB = (V_{AB}, C_{AB}, u_{AB})$, where C_{AB} obeys (2.18), we say that AB **violates a Bell inequality** if for some integers $n, k, n', k' \geq 2$ there are an (n, k) -apparatus Γ_A on A and an (n', k') -apparatus Γ_B on B such that

$$(\Gamma_A \otimes \Gamma_B)(C_{AB}) \not\subseteq C_n^k \otimes_{\min} C_{n'}^{k'}.$$

It is not difficult to observe that if $AB = A \otimes_{\min} B$ (i.e. only separable bipartite states are allowed) then no violation of a Bell inequality can occur. In particular, the forthcoming Lemma 2.27 ensures that if either A or B is a classical theory according to Definition 2.9, then AB does not violate any Bell inequality.

2.4 Separability problem

Throughout this section, we will explore some feature of the separability problem in general probabilistic theories. This corresponds to point (iv) of the discussion in Subsection 2.1.1. While the core of the present chapter is mainly formed by the forthcoming Sections 2.5 and 2.6, and in some respects this section could be seen as a mathematical divertissement, we will see that some of the tools we develop here will turn out useful several times throughout the thesis. In fact, we chose to present this material before delving into the more fundamental problems discussed in the forthcoming sections exactly because some of the questions that we shall rise later in the chapter have a lot to do with the separability problem. Furthermore, some of the results we present here will be instrumental in Chapter 3.

We remind the reader that in quantum information the separability problem [110] consists in establishing computationally convenient criteria to decide

whether a given state is separable or entangled. Analogously, in GPTs we will have to decide whether a given vector in the maximal tensor product of two cones belongs also to the minimal tensor product or not.

For quantum states there is a wealth of available criteria, even if the problem is known to be NP-hard in general (under certain assumptions on the required error) [114, 115]. However, many of those can not be expected to carry over to the GPT realm, because they depend on features of the problem that are specific to the quantum case. For instance, the positive partial transpose (PPT) criterion [116, 117], as well as the reduction criterion [118] depend on the existence of a special map that preserves the positive cone, while the reshuffling criterion [119, 120, 121] makes use of the fact that each local vector space is a space of linear operators (hermitian matrices), and thus itself a tensor product. The same is true at the global level for the range criterion [122], which indeed involves the notion of support and thus refers to the density matrices as operators and not merely as vectors in a real space. This is common to all the conditions that are related to the spectra, like the majorization criterion [123], the separable ball construction [124], and the minimal eigenvalue sufficient condition [125].

Of all the mainstream entanglement criteria, we are left with three: what we could call (following the terminology of [124]) the ‘Woronowicz condition’ [94], and the complete extendibility criterion [126]. Furthermore, there is a less known sufficient condition based on the tensor rank of the density matrix itself [95, 127, 96]. The rest of this section is dedicated to investigating these conditions in the GPT framework.

2.4.1 Woronowicz and complete extendibility criteria

We start with the Woronowicz criterion, since it is the simplest. For the quantum separability problem, it states roughly that a state is separable if and only if it has a positive expectation value on all entanglement witnesses, where an entanglement witness is an observable that in turn has positive expectation value on all separable states. As one sees immediately, this has nothing to do with specific features of quantum theory, and is rather an instance of the Hahn-Banach separation theorem (Theorem 1.7), which allows us to identify a closed convex set with the intersection of all the closed half-spaces containing it. An equivalent way of recasting the Woronowicz condition is via positive maps: a bipartite quantum state ω_{AB} is separable iff $(T_{A \rightarrow \tilde{A}} \otimes I)(\omega_{AB})$ is still a positive operator for all positive maps $T_{A \rightarrow \tilde{A}} : \mathcal{H}_{n_A} \rightarrow \mathcal{H}_{n_{\tilde{A}}}$, where positive means that $T_{A \rightarrow \tilde{A}}(\text{PSD}_{n_A}) \subseteq \text{PSD}_{n_{\tilde{A}}}$.

This latter notion admits an obvious generalisation to GPTs. Given two ordered vector spaces V_1, V_2 , a map $T : V_1 \rightarrow V_2$ is said to be **positive** if it sends positive elements to positive elements, i.e. if $T(x) \geq 0$ whenever $x \geq 0$. This is a different notion than that of ‘complete positivity’ we investigated in Subsection 2.2.4. Every completely positive map must be positive, but the converse is generally not true. In quantum mechanics, the transposition provides a classical example of a positive yet not completely positive map [116].

We can rephrase the positivity requirement for maps within the language we have been developing so far if we remember that we can equivalently see a map $T : V_1 \rightarrow V_2$ as a tensor $W_T \in V_1^* \otimes V_2$. This identification is made possible by the canonical isomorphism $T \leftrightarrow W_T$ defined by the identity

$$\langle W_T, Z \rangle = \text{tr} [(T \otimes I)(Z)] \quad \forall Z \in V_1 \otimes V_2^*, \quad (2.50)$$

where for an arbitrary vector space V one defines the **trace function**

$$\begin{aligned} \text{tr} : V \otimes V^* &\longrightarrow \mathbb{R} \\ x \otimes \varphi &\longmapsto \langle \varphi, x \rangle. \end{aligned} \quad (2.51)$$

It is very simple to verify that T is a positive map iff $W_T \in C_{1 \max}^* \otimes C_2^*$, where C_1 and C_2 are the positive cones of V_1 and V_2 , respectively.

Equipped with these tools, we can formulate a generalisation of Woronowicz condition to the GPT setting.

Proposition 2.15 (Woronowicz condition for GPTs). *Let V_A, V_B be finite-dimensional ordered vector spaces with positive cones C_A, C_B . For a tensor $Z \in C_A \otimes_{\max} C_B$, the following are equivalent:*

- (a) $Z \in C_A \otimes_{\min} C_B$;
- (b) for all $W \in C_A^* \otimes_{\max} C_B^*$ one has $\langle W, Z \rangle \geq 0$;
- (c) for all positive maps $T : V_A \rightarrow V_B^*$, one has $\text{tr} [(T \otimes I)(Z)] \geq 0$.

Proof. The equivalence between (a) and (b) is a consequence of a duality relation analogous to (2.17), which reads

$$\left(C_A^* \otimes_{\max} C_B^* \right)^* = C_A \otimes_{\min} C_B. \quad (2.52)$$

Moreover, the existence of the canonical isomorphism in (2.50) implies that (c) is just a reformulation of (b). \square

Until now we have discussed only the Woronowicz criterion, which depends basically only on the convex structure of the set of separable states and therefore admits an immediate generalisation to the GPT setting. Let us turn our attention to another separability criterion that has found extensive application in quantum information theory, namely the complete extendibility criterion [126], which states that every separable density matrix ρ_{AB} admits a symmetric extension of arbitrarily high order, i.e. for all $k = 1, 2, \dots$ there is a legitimate (i.e. positive) quantum state that: (i) extends ρ_{AB_1} on $AB_1 \dots B_k$, where the systems B_i are identical copies of B , and (ii) is symmetric under any permutation of the B_i systems. More than a single condition, this is instead a whole hierarchy of increasingly stringent conditions. Each level of the hierarchy

can be decided in a computationally efficient fashion, since it is nothing but a semidefinite program, and furthermore the hierarchy is complete, meaning that if a symmetric extension is found for all $k = 1, 2, \dots$, then the original state was separable.

In the GPT setting, we do not have a unique notion of positivity at our disposal, but we can still observe that if an arbitrary tensor $Z \in C_A \otimes_{\max} C_B$ is separable, i.e. $Z \in C_A \otimes_{\min} C_B$, then as before for all positive integers k there is a symmetric extension $Z_{AB_1 \dots B_k}$ that belongs to $Z \in C_A \otimes_{\max} C_{B_1} \dots \otimes_{\max} C_{B_k}$. In fact, if $Z = \sum_i x_i \otimes y_i$ with $x_i, y_i \geq 0$, and we assume without loss of generality that $\langle u, y_i \rangle = 1$, a symmetric extension can be constructed as $Z_{AB_1 \dots B_k} = \sum_i x_i \otimes y_i^{\otimes k}$. It is clear that $Z_{AB_1 \dots B_k} \in C_A \otimes_{\min} C_{B_1} \dots \otimes_{\min} C_{B_k} \subseteq C_A \otimes_{\max} C_{B_1} \dots \otimes_{\max} C_{B_k}$.

Whether the complete extendibility condition is also sufficient in general, it does not seem transparent. We limit ourselves to illustrating why the standard way of proving the completeness of the hierarchy in quantum mechanics fails already in W -theory, as given by Definition 2.10. We remind the reader that states in W -theory are represented by entanglement witnesses, which are in general non-positive. This lack of positivity deprives us of an important tool, i.e. the existence of purifications. Most standard proofs of the quantum de Finetti theorem (see for instance [128] or [129, §9]) seem to rest crucially on the possibility of purifying a symmetric system while retaining the symmetry. Thus, we are forced to leave this problem open, hoping that further investigations will provide either a proof of the universality of the complete extendibility criterion or an appropriate counterexample.

2.4.2 Separability under symmetry

Our main focus in this subsection is on how to simplify the separability problem when the states under examination possess further symmetries. As it turns out, these simplifications can be crucial in making the problem analytically solvable for specific classes of states, as we shall see in Section 4.5. In the case of quantum theory, what follows has become part of the common knowledge in the entanglement theory community [130, 131].

First of all, let us fix some notation. Consider two GPTs $A = (V_A, C_A, u_A)$ and $B = (V_B, C_B, u_B)$, and take a vector subspace $\mathcal{V} \subseteq V_{AB} = V_A \otimes V_B$ of the global vector space. Naturally, we will want to consider subspaces that contain some nontrivial states, i.e. such that $\mathcal{V} \cap (C_A \otimes_{\max} C_B) \neq \{0\}$.

The problem we address here consists in deciding separability of states that belong to \mathcal{V} . In particular, we want to identify what properties of \mathcal{V} can simplify the analysis significantly. In what follows, we denote by \mathcal{C}_{\min} and \mathcal{C}_{\max} the cones on \mathcal{V} obtained by intersecting it with $C_A \otimes_{\min} C_B$ and $C_A \otimes_{\max} C_B$,

respectively. In formulae,

$$\mathcal{E}_{\min} := \mathcal{V} \cap \left(C_{A_{\min}} \otimes C_B \right), \quad (2.53)$$

$$\mathcal{E}_{\max} := \mathcal{V} \cap \left(C_{A_{\max}} \otimes C_B \right). \quad (2.54)$$

Naturally, one can think of \mathcal{V}^* as ordered by any of the two dual cones $\mathcal{E}_{\min}^*, \mathcal{E}_{\max}^* \subset \mathcal{V}^*$. For a totally generic subspace \mathcal{V} , a very basic statement concerning the characterisation of separability is obtained by ‘projecting’ the Hahn-Banach separation theorem onto \mathcal{V} , as expressed by the following result.

Proposition 2.16. *For $Z \in \mathcal{V}$, we have that $Z \in \mathcal{E}_{\min}$ iff*

$$\langle W, Z \rangle \geq 0 \quad \forall W \in \mathcal{E}_{\min}^*. \quad (2.55)$$

The problem with the above statement is that \mathcal{E}_{\min}^* is in general not easy to characterise. In fact, even though the extremal rays of $C_{A_{\min}} \otimes C_B$ have a mathematically convenient form, once the intersection with \mathcal{V} is carried out it is no longer clear how to identify the new extremal rays, i.e. those of \mathcal{E}_{\min} . Consequently, imposing $W \in \mathcal{E}_{\min}^*$, i.e. $\langle W, Z \rangle \geq 0$ for all vectors Z that generate extremal rays of \mathcal{E}_{\min} , can easily be analytically unfeasible.

From the above discussion it is clear that the solution of the separability problem would be greatly simplified if one had in hand an efficient way of characterising \mathcal{E}_{\min}^* . This is possible for a relevant class of subspaces \mathcal{V} that we are now set to define. We remind the reader that a **projection** onto a subspace V_2 of a vector space V_1 is a linear map $\Pi : V_1 \rightarrow V_1$ such that: (i) $\Pi^2 = \Pi$; and (ii) $\Pi(V_1) = V_2$.

Definition 2.17. *A subspace $\mathcal{V} \subseteq V_{AB}$ of a bipartite system is called a **central section** (of $C_{A_{\min}} \otimes C_B$) if there exists a projection Π of V_{AB} onto \mathcal{V} that is also separability-preserving, i.e. such that*

$$\Pi \left(C_{A_{\min}} \otimes C_B \right) \subseteq \mathcal{V} \cap \left(C_{A_{\min}} \otimes C_B \right) = \mathcal{E}_{\min}. \quad (2.56)$$

Before proceeding further, let us stop for a second and look into a particular class of example of central sections that are of prominent importance in quantum information.

Example 2.18. An interesting particular case of separability-preserving projection is of the form

$$\Pi_G := \int_G dg \zeta_1(g) \otimes \zeta_2(g), \quad (2.57)$$

where $\zeta_i : G \rightarrow GL(V_i)$ are (real) representations of a compact group G on V_1, V_2 , and $\int_G dg$ is the Haar integral. In that case, \mathcal{V} must be taken to be the set of fixed points of the representation $\zeta_1 \otimes \zeta_2$, that coincides (up to the Choi isomorphism in an orthogonal basis for a G -invariant product) with the set of

G -commuting isomorphisms $V_1 \rightarrow V_2$, and therefore can be easily determined with the aid of representation theory. What we just described constitutes the typical application of this framework [91, 130]. We will stumble upon it ourselves in Section 4.5.

Returning to the general framework, we see that the existence of a (for now generic) projection Π of V_{AB} onto \mathcal{V} , which can be equivalently seen as an application $\Pi : V_{AB} \rightarrow \mathcal{V}$, enables us to give some more structure to the problem. In fact, we can construct also the adjoint projection $\Pi^* : \mathcal{V}^* \rightarrow V_{AB}^*$ and identify $\mathcal{V}^* \simeq \Pi^*(V_{AB}^*)$. This amounts to thinking of \mathcal{V}^* as a subspace of V_{AB}^* by saying that $\varphi \in \mathcal{V}^*$ can be extended to $\tilde{\varphi} \in V_{AB}^*$ through $\langle \tilde{\varphi}, \cdot \rangle := \langle \varphi, \Pi(\cdot) \rangle$. Therefore, from now on we can safely let $\varphi \in \mathcal{V}^*$ act on global states.

Once we fix a projection onto \mathcal{V} , so that $\mathcal{V}^* \subseteq V_{AB}^*$, we can define two other cones in the dual space:

$$\mathcal{E}_{*\min} := \mathcal{V}^* \cap \left(C_{A\min}^* \otimes C_B^* \right), \quad (2.58)$$

$$\mathcal{E}_{*\max} := \mathcal{V}^* \cap \left(C_{A\max}^* \otimes C_B^* \right). \quad (2.59)$$

Together with (2.53) and (2.54), we then have four possible cones in the dual space \mathcal{V}^* . However, not all of these cones are independent of each other. For instance, one can easily show that

$$\mathcal{E}_{*\min} \subseteq \mathcal{E}_{*\max}^*, \quad (2.60)$$

$$\mathcal{E}_{*\max} \subseteq \mathcal{E}_{*\min}^*, \quad (2.61)$$

and consequently that

$$\mathcal{E}_{\max} \subseteq \mathcal{E}_{*\min}^*, \quad (2.62)$$

$$\mathcal{E}_{\min} \subseteq \mathcal{E}_{*\max}^*. \quad (2.63)$$

Now, remember that we are looking for a convenient mathematical characterisation of the cone $\mathcal{E}_{*\min}^*$. Such a characterisation, in fact, would allow us to improve the efficiency of the criterion in Proposition 2.16. The importance of central sections in this context rests primarily on the following observation.

Lemma 2.19. *Let $\mathcal{V} \subseteq V_{AB}$ be a central section. Then the cones $\mathcal{E}_{*\min}$, $\mathcal{E}_{*\max}$ associated with it through (2.53) and (2.59) satisfy*

$$\mathcal{E}_{*\max} = \mathcal{E}_{*\min}^*, \quad (2.64)$$

i.e. the bound in (2.61) (equivalently, that in (2.63)) is saturated.

Proof. Take a $W \in \mathcal{E}_{*\min} \subseteq \mathcal{V}^* \subseteq V_{AB}^*$, and let us show that $W \in \mathcal{E}_{*\max}$, indeed. Because of the definition (2.59), all we have to prove is that $W \in C_{A\max}^* \otimes C_B^*$. Thanks to the duality relation (2.52), this translates to

$$\langle W, Z \rangle \geq 0 \quad \forall Z \in C_{A\min} \otimes C_B.$$

Now, the way the inclusion of $\mathcal{V}^* \subseteq V_{AB}^*$ is designed implies that

$$\langle W, Z \rangle = \langle W, \Pi(Z) \rangle \geq 0,$$

because $\Pi(Z) \in \Pi(C_A \otimes_{\min} C_B) \subseteq \mathcal{C}_{\min}$ as follows from (2.56). \square

As an immediate consequence, we obtain an improved version of Proposition 2.16.

Proposition 2.20. *For \mathcal{V} central section and $Z \in \mathcal{V}$, we have that $Z \in \mathcal{C}_{\min}$ iff*

$$\langle W, Z \rangle \geq 0 \quad \forall W \in \mathcal{C}_{*\min}. \quad (2.65)$$

Proof. The necessity of (2.65) is obvious, while the converse statement follows by putting together Proposition 2.16 and Lemma 2.19. \square

Rephrasing the claim of Theorem 2.20, we could say that whenever \mathcal{V} is a central section, in order to decide whether $Z \in \mathcal{V}$ is separable or not it suffices to test the inequality $\langle W, Z \rangle \geq 0$ on all functionals $W \in \mathcal{V}^* \subseteq V_{AB}^*$ that happen to be global entanglement witnesses.

2.4.3 Tensor rank and separability

In what follows, we are going to adopt the following notation. For a finite-dimensional ordered vector space V , a vector that is internal to the positive cone, in symbols $x \in \text{int}(V_+)$, will be called strictly positive, and we will write $x > 0$.

Lemma 2.21. *Let V_A, V_B be finite-dimensional ordered vector spaces, whose positive cone we denoted by C_A, C_B . If $a \in V_A$ and $b \in V_B$ are both strictly positive, i.e. $x, y > 0$, then $a \otimes b \in \text{int}(C_A \otimes_{\min} C_B)$, i.e. $a \otimes b > 0$ with respect to the ordering determined on $V_A \otimes V_B$ by the minimal tensor product $C_A \otimes_{\min} C_B$.*

Proof. Fix two bases $\{v_i\}_{i=1}^n$ on V_A and $\{w_j\}_{j=1}^m$ on V_B . For all vectors $\delta \in V_A$ expressed uniquely as $\delta = \sum_{i=1}^n \alpha_i v_i$, define the norms $|\delta|_\infty := \max_i |\alpha_i|$, $|\delta|_1 := \sum_i |\alpha_i|$, and analogously for V_B . The same constructions can be applied to the tensor product $V_A \otimes V_B$ equipped with the basis $\{v_i \otimes w_j\}_{i,j=1}^{nm}$. Now, take $\epsilon > 0$ such that $|\delta|_\infty \leq \epsilon$ implies $a + \delta \geq 0$ for all $\delta \in V_A$ and $b + \delta \geq 0$ for all $\delta \in V_B$. Then we claim that for all $\Delta \in V_A \otimes V_B$ we have that $|\Delta|_1 \leq \epsilon^2$ implies $a \otimes b + \Delta \in C_A \otimes_{\min} C_B$. In fact, taking $\Delta = \sum_{ij} \alpha_{ij} v_i \otimes w_j$ such that

$|\Delta|_1 \leq \epsilon^2$, we have

$$\begin{aligned}
a \otimes b + \Delta &= a \otimes b + \sum_{i,j} \alpha_{ij} v_i \otimes w_j \\
&= \left(1 - \frac{|\Delta|_1}{\epsilon^2}\right) a \otimes b + \frac{1}{\epsilon^2} \sum_{ij} |\alpha_{ij}| (a \otimes b + s_{ij} \epsilon^2 v_i \otimes w_j) \\
&= \left(1 - \frac{|\Delta|_1}{\epsilon^2}\right) a \otimes b + \frac{1}{\epsilon^2} \sum_{ij} |\alpha_{ij}| (a \otimes b + s_{ij} \epsilon^2 v_i \otimes w_j) \\
&= \left(1 - \frac{|\Delta|_1}{\epsilon^2}\right) a \otimes b \\
&\quad + \sum_{ij} |\alpha_{ij}| ((a + \epsilon v_i) \otimes (b + s_{ij} \epsilon w_j) + \\
&\quad + (a - \epsilon v_i) \otimes (b - s_{ij} \epsilon w_j)),
\end{aligned}$$

where $s_{ij} = \text{sign}(\alpha_{ij})$ are the signs of the coefficients α_{ij} . The above representation is explicitly separable, since the choice of $\epsilon > 0$ makes

$$a \pm \epsilon v_i \geq 0, \quad b \pm \epsilon w_j \geq 0$$

for all i and j . We thus have $a \otimes b + \Delta \in C_A \otimes_{\min} C_B$, as claimed. \square

Lemma 2.22. *Let V be a finite-dimensional ordered vector space V . For $a > 0$ and $v \in V$ with $v \not\geq 0$, define $t_0 := \max\{t \geq 0 \in \mathbb{R} : a + tv \geq 0\}$. Then there exists a positive functional φ_0 that: (a) defines an extremal ray of the dual positive cone; (b) is such that $\langle \varphi, a + t_0 v \rangle = 0$; and (c) satisfies $\langle \varphi, a \rangle > 0$.*

Proof. We denote by C the positive cone of V . A little thought reveals that by construction $a + t_0 v$ must lie on the boundary of C . This implies that there is a positive functional $\varphi_0 \in C^*$ such that $\langle \varphi_0, a + t_0 v \rangle = 0$, meeting (b). In fact, if this were not the case, we could take a norm on V , denoted by $\|\cdot\|$, then consider

$$\epsilon := \inf \{ \langle \varphi, a + t_0 v \rangle : \varphi \in C^*, \|\varphi\|_* = 1 \} > 0,$$

where the last inequality follows by compactness, and note that the vector $a + \left(t_0 + \frac{\epsilon}{\|v\|}\right) v$ belongs to $C = C^{**}$ since it has a positive expectation value on all functionals $\varphi \in C^*$. This would be in contradiction with the definition of t_0 .

We can of course decompose $\varphi_0 \in C^*$ as a convex combination of functionals lying on extreme rays of C^* , and naturally all of those functionals vanish when evaluated on $a + t_0 v$. It follows that we can safely assume that φ_0 determines an extremal ray itself, meeting requirement (a). Finally, since a is internal to C and $\varphi_0 \in C^*$, we must have $\langle \varphi, a \rangle > 0$. \square

The first part of the proof follows closely that given for the quantum case in [95] (see also [127]). However, we will see that we are forced to depart from it at some point.

Theorem 2.23. *Let V_A, V_B be finite-dimensional ordered vector spaces with positive cones C_A, C_B . Whenever $Z \in V_A \otimes V_B$ has tensor rank 2 or less, $Z \in C_{A \max} \otimes C_B$ if and only if $Z \in C_{A \min} \otimes C_B$.*

Proof. Clearly, we have just to show that every vector $Z \in C_{A \max} \otimes C_B$ of tensor rank 2 belongs also to $C_{A \min} \otimes C_B$, as the inclusion $C_{A \min} \otimes C_B \subseteq C_{A \max} \otimes C_B$ always holds. From now on, we will consider V_A, V_B endowed with the partial orders given by C_A, C_B . We write

$$Z = a \otimes b + x \otimes y. \quad (2.66)$$

Let us break down the proof into several steps.

1. First of all, we want to show that we can suppose without loss of generality $a, b \geq 0$ (each of them in its own space). To see this, pick strictly positive functionals $u_A \in \text{int}(C_A^*)$ and $u_B \in \text{int}(C_B^*)$. Since $u_A \otimes u_B$ is again internal to $C_{A \min}^* \otimes C_B^*$ thanks to Lemma 2.21, applying (2.17) we conclude that $\langle u_A \otimes u_B, Z \rangle > 0$. Therefore, without loss of generality we can assume that

$$\langle u_A \otimes u_B, Z \rangle = \langle u_A, a \rangle \langle u_B, b \rangle + \langle u_A, x \rangle \langle u_B, y \rangle = 1.$$

Now, let us write

$$\begin{aligned} Z &= (\langle u_B, b \rangle a + \langle u_B, y \rangle x) \otimes (\langle u_A, a \rangle b + \langle u_A, x \rangle y) \\ &\quad + (\langle u_A, x \rangle a - \langle u_A, a \rangle x) \otimes (\langle u_B, y \rangle b - \langle u_B, b \rangle y). \end{aligned}$$

In the above equation, which has the same form as (2.66), we see that $\langle u_B, b \rangle a + \langle u_B, y \rangle x = \langle u_B, Z \rangle \geq 0$, where $\langle u_B, Z \rangle$ corresponds to a partial evaluation of the functional u_B on the tensor Z , and is therefore a positive vector since $Z \in C_{A \max} \otimes C_B$. Analogously, $\langle u_A, a \rangle b + \langle u_A, x \rangle y = \langle u_A, Z \rangle \geq 0$.

2. In fact, we can do better, and suppose in (2.66) that $a, b > 0$. This is because for all $a, b \geq 0$ there are always sequences a_n, b_n such that $a_n > a, b_n > b$ for all $n \in \mathbb{N}$, with equality in the limit $n \rightarrow \infty$. To obtain such sequences, it suffices to add to a, b asymptotically small multiples of a strictly positive vector. The new tensor $Z_n \equiv a_n \otimes b_n + x \otimes y$ still belongs to $C_{A \max} \otimes C_B$, as it can be verified that the application of a nonzero product functional always yields a larger result than that one would obtain with Z . Since $\lim_{n \rightarrow \infty} Z_n = Z$ and the minimal tensor product is closed, it suffices to show that $Z_n \in C_{A \min} \otimes C_B$ for all n .

3. At this point Cariello [95] operates a local filtering operation to assume that a, b are both the identity matrix. This is obviously not possible in the GPT setting, so we are forced to take a different route. The originality of our proof lies mainly in the rest of the argument.

We refer again to (2.66) but now we take $a, b > 0$. We claim that we can suppose without loss of generality that there are two positive functionals $\varphi \in C_A^*$, $\lambda \in C_B^*$ such that $\langle \lambda, a \rangle \langle \eta, b \rangle = -\langle \lambda, x \rangle \langle \eta, y \rangle > 0$. In fact, defining

$$k := \max \left\{ t \geq 0 : a \otimes b + kx \otimes y \in C_{A_{\max}} \otimes C_B \right\} \geq 1,$$

we see that

$$Z = \left(1 - \frac{1}{k}\right) a \otimes b + \frac{1}{k} (a \otimes b + kx \otimes y). \quad (2.67)$$

Clearly, proving that $a \otimes b + kx \otimes y \in C_{A_{\min}} \otimes C_B$ would allow us to conclude. Observe that by Lemma 2.22 and identity (2.17) there is a functional $\Phi \in (C_{A_{\max}} \otimes C_B)^* = C_{A_{\min}}^* \otimes C_B^*$ corresponding to an extremal ray and such that $\langle \Phi, a \otimes b + kx \otimes y \rangle = 0$ and $\langle \Phi, a \otimes b \rangle > 0$. It is elementary to verify that the extremal rays in the minimal tensor product are just tensor product of positive functionals, i.e. $\Phi = \lambda \otimes \eta$ with $\lambda \in C_A^*$, $\eta \in C_B^*$. Renaming $kx \mapsto x$ yields the claim.

4. Consider the two positive functionals λ, η we constructed above. According to step 3, we now assume that Z written as in (2.66) satisfies $\langle \lambda, a \rangle \langle \eta, b \rangle = -\langle \lambda, x \rangle \langle \eta, y \rangle > 0$. Applying them on one side or the other of Z we get

$$\begin{aligned} \langle \eta, Z \rangle &= \langle \eta, b \rangle a + \langle \eta, y \rangle x =: c \geq 0, \\ \langle \lambda, Z \rangle &= \langle \lambda, a \rangle b + \langle \lambda, x \rangle y =: d \geq 0. \end{aligned}$$

Since $\langle \lambda, a \rangle \langle \eta, b \rangle > 0$, proving that $\langle \lambda, a \rangle \langle \eta, b \rangle Z \in C_{A_{\min}} \otimes C_B$ will suffice.

We obtain

$$\begin{aligned} \langle \lambda, a \rangle \langle \eta, b \rangle Z &= \langle \eta, b \rangle a \otimes \langle \lambda, a \rangle b + \langle \lambda, a \rangle \langle \eta, b \rangle x \otimes y \\ &= (c - \langle \eta, y \rangle x) \otimes (d - \langle \lambda, x \rangle y) + \langle \lambda, a \rangle \langle \eta, b \rangle x \otimes y \\ &= c \otimes d - \langle \eta, y \rangle x \otimes d - \langle \lambda, x \rangle c \otimes y \\ &=: W. \end{aligned} \quad (2.68)$$

5. By the same reasoning as in step 2, we can suppose that $c, d > 0$. Exactly as in step 3, we can again assume without loss of generality that there are two positive functionals $\mu \in C_A^*$, $\nu \in C_B^*$ satisfying $\langle \mu \otimes \nu, W \rangle = 0$

and $\langle \mu, c \rangle \langle \nu, d \rangle > 0$. Then, it suffices to prove that $\langle \mu, c \rangle \langle \nu, d \rangle W \in C_A \otimes_{\min} C_B$. Applying μ and ν on W written as in (2.68) we get

$$\begin{aligned}\langle \nu, W \rangle &= (\langle \nu, d \rangle - \langle \lambda, x \rangle \langle \nu, y \rangle) c - \langle \eta, y \rangle \langle \nu, d \rangle x =: p \geq 0, \\ \langle \mu, W \rangle &= (\langle \mu, c \rangle - \langle \eta, y \rangle \langle \mu, x \rangle) d - \langle \lambda, x \rangle \langle \mu, c \rangle y =: q \geq 0.\end{aligned}$$

Finally,

$$\begin{aligned}\langle \mu, c \rangle \langle \nu, d \rangle W &= \langle \mu, c \rangle \langle \nu, d \rangle c \otimes d - \langle \eta, y \rangle \langle \nu, d \rangle x \otimes \langle \mu, c \rangle d - \langle \nu, d \rangle c \otimes \langle \lambda, x \rangle \langle \mu, c \rangle y \\ &= \langle \mu, c \rangle \langle \nu, d \rangle c \otimes d + (p - (\langle \nu, d \rangle - \langle \lambda, x \rangle \langle \nu, y \rangle) c) \otimes \langle \mu, c \rangle d \\ &\quad + \langle \nu, d \rangle c \otimes (q - (\langle \mu, c \rangle - \langle \eta, y \rangle \langle \mu, x \rangle) d) \\ &= (\langle \mu, c \rangle \langle \nu, d \rangle - \langle \mu, c \rangle (\langle \nu, d \rangle - \langle \lambda, x \rangle \langle \nu, y \rangle) \\ &\quad - \langle \nu, d \rangle (\langle \mu, c \rangle - \langle \eta, y \rangle \langle \mu, x \rangle)) c \otimes d \\ &\quad + \langle \mu, c \rangle p \otimes d + \langle \nu, d \rangle c \otimes q \\ &= (-\langle \mu, c \rangle \langle \nu, d \rangle + \langle \mu, c \rangle \langle \lambda, x \rangle \langle \nu, y \rangle + \langle \nu, d \rangle \langle \eta, y \rangle \langle \mu, x \rangle) c \otimes d \\ &\quad + \langle \mu, c \rangle p \otimes d + \langle \nu, d \rangle c \otimes q \\ &= -\langle \mu \otimes \nu, W \rangle c \otimes d + \langle \mu, c \rangle p \otimes d + \langle \nu, d \rangle c \otimes q \\ &= \langle \mu, c \rangle p \otimes d + \langle \nu, d \rangle c \otimes q \\ &\in C_A \otimes_{\min} C_B,\end{aligned}$$

which allows us to conclude. □

Remark. A little of thought reveals that we have also proven that any $Z \in C_A \otimes_{\max} C_B$ of tensor rank 2 can be always written as $Z = \sum_{i=1}^4 a_i \otimes b_i$, with $a_i, b_i \geq 0$.

When restricted to quantum mechanics, Theorem 2.23 tells us the following.

Corollary 2.24. *Every entanglement witness W as in (2.35) whose tensor rank is 2 or less is necessarily separable, i.e. it belongs to the cone (2.34).*

Proof. Straightforward application of Theorem 2.23 to quantum mechanics as discussed in Subsection 2.3.2. □

The above Corollary 2.24 generalises the result obtained by Carriello [95, 127], which states that any *positive* operator of tensor rank 2 or less is separable. Things can not be pushed further, as in [96] it is proven that there exist two-qutrit entangled states whose tensor rank is 3.

2.4.4 Separability problem for centrally symmetric models

We now turn our attention to the class of centrally symmetric models (Definition 2.11). The reason why we should care about these models will become clear in Chapter 3. For now, as we said at the beginning of the section, let us see this as a mathematical divertissement. For notation and conventions, we refer the reader to Subsection 2.3.3.

For the sake of the presentation, let us start by providing a concise explanation of the notation to be used in what follows. The vector space hosting a centrally symmetric GPT is a Euclidean space \mathbb{R}^d . Vectors belonging to this space are denoted by $x = (x_0, x_1, \dots, x_{d-1})^T$, the 0-th component being the normalisation of x . The space thus decomposes naturally as $\mathbb{R}^d = \mathbb{R}u_* \oplus \mathbb{R}^{d-1}$, where $u_* = (1, 0, \dots, 0)^T$ is a distinguished normalised state, and the second ‘vector component’ is endowed with a norm that we denote with $|\cdot|$. A vector $v \in \mathbb{R}^{d-1}$ can be ‘lifted’ to $\widehat{v} = (0, v) \in \mathbb{R}^d$, and conversely any $x = (x_0, \bar{x}) \in \mathbb{R}^d$ can be ‘lowered’ to obtain $\bar{x} \in \mathbb{R}^{d-1}$.

We now consider two centrally symmetric GPTs of dimensions d_A and d_B . Every state of the bipartite system AB will be represented in a canonical way by a matrix Z of size $d_A \times d_B$. The distinguished subspace of $(\mathbb{R} \oplus \mathbb{R}^{d_A-1}) \otimes (\mathbb{R} \oplus \mathbb{R}^{d_B-1})$ formed by the product $\mathbb{R}^{d_B-1} \otimes \mathbb{R}^{d_B-1}$ of the two vector components occupies the bottom right corner of Z . Retaining only this corner of Z yields another matrix $\bar{Z} \in \mathbb{R}^{(d_A-1) \times (d_B-1)}$. One can conversely lift matrices of this latter kind to larger ones by adding a row and a column of zeros. The lifted version of $M \in \mathbb{R}^{(d_A-1) \times (d_B-1)}$ will be denoted by \widehat{M} .

It is important to realise that bipartite systems made of local centrally symmetric theories possess a privileged state $U_* = u_{A*} \otimes u_{B*} = u_{A*} u_{B*}^T$. We are going to solve the separability problem for states of the particular form $Z = U_* + \widehat{M}$, where $M \in \mathbb{R}^{(d_A-1) \times (d_B-1)}$. As we shall see in a moment, the solution involves the theory of tensor norms as we discussed it in Subsection 2.2.1.

Proposition 2.25. *Let Z be a state of a bipartite theory whose local systems are represented by centrally symmetric GPTs. Then*

$$Z \in C_A \underset{\min}{\otimes} C_B \implies |\bar{Z}|_\pi \leq Z_{00}, \quad (2.69)$$

$$Z \in C_A \underset{\max}{\otimes} C_B \implies |\bar{Z}|_\varepsilon \leq Z_{00}, \quad (2.70)$$

where $|\cdot|_\pi, |\cdot|_\varepsilon$ are the projective and injective tensor norms constructed out of the local norms $|\cdot|$. Furthermore, the above conditions are also sufficient when all cross terms Z_{i0}, Z_{0j} ($i, j \geq 1$) vanish.

Proof. We start by proving (2.69). If $Z = \sum_k x_k y_k^T$ with $|\bar{x}_k| \leq x_{k0}, |\bar{y}_k| \leq y_{k0}$

for all k , then

$$\begin{aligned} |\bar{Z}|_\pi &= \left| \sum_k \bar{x}_k \bar{y}_k^T \right|_\pi \\ &\leq \sum_k |\bar{x}_k| |\bar{y}_k| \\ &\leq \sum_k x_{k0} y_{k0} \\ &= Z_{00}. \end{aligned}$$

Now, let us turn our attention to the proof of the sufficiency claim. Suppose that $Z_{ij} = 0$ whenever $i = 0, j \geq 1$ or $i \geq 1, j = 0$, i.e. that $Z = Z_{00}U_* + \widehat{Z}$, and further assume that $|\bar{Z}|_\pi \leq Z_{00}$. Then by definition of projective norm (2.2) there are vectors $v_k \in \mathbb{R}^{d_A-1}$ and $w_k \in \mathbb{R}^{d_B-1}$ such that $\bar{Z} = \sum_k v_k w_k^T$ and $|\bar{Z}|_\pi = \sum_k |v_k| |w_k|$. Define vectors $x_k^\pm := |v_k| u_A \pm \hat{v}_k \in \mathbb{R}^{d_A}$, i.e.

$$(x_k^\pm)_i = \begin{cases} |v_k| & \text{if } i = 0, \\ \pm (v_k)_i & \text{if } i \geq 1, \end{cases}$$

and analogously for $y_k^\pm := |w_k| u_B \pm \hat{w}_k \in \mathbb{R}^{d_B}$. Observe that the definition of centrally symmetric model (2.38) tells us that $x_k^\pm \in C_A$ and $y_k^\pm \in C_B$. Then it is easy to see that

$$Z = (Z_{00} - |\bar{Z}|_\pi) u_A u_B^T + \frac{1}{2} \sum_k \left(x_k^+ (y_k^+)^T + x_k^- (y_k^-)^T \right),$$

which shows that Z is separable, as claimed. Alternatively, one can show separability by exploiting the ‘Woronowicz condition’ (Proposition 2.15(b)).

We can perform an analogous reasoning at the level of the dual, which shares the same centrally symmetric structure, as apparent from (2.38). We obtain immediately that $W \in C_A^* \otimes_{\min} C_B^*$ implies that $|\bar{W}|_{*\pi} \leq W_{00}$. An arbitrary $Z \in C_A \otimes_{\max} C_B$ will satisfy $\langle W, Z \rangle \geq 0$ for all such W . Optimising this condition yields $Z_{00} - |\bar{Z}|_{*\pi} \geq 0$, which becomes $|\bar{Z}|_\varepsilon \leq Z_{00}$ after employing the duality formulae (2.3). This concludes the proof of (2.70). Exploiting again duality in the form of (2.16), showing that (2.70) expresses a sufficient condition when $Z = Z_{00}U_* + \widehat{Z}$ is a simple exercise that is left to the reader. \square

Of course, the above Proposition 2.25 constitutes just a rephrasing of the problem, if we do not know how to compute injective and projective norms in the specific case we want to analyse. And in fact in general the optimisations problems in (2.1) and (2.2) are hard to solve explicitly. However, as we saw in Subsection 2.2.1 there is an important special case where some sort of closed expression is indeed available, and this is the case of two Euclidean spaces. At the level of GPTs, this is naturally the same as considering two spherical models, as defined in (2.44), (2.45). We obtain the following.

Corollary 2.26. *Let Z be a state of a bipartite theory whose local systems are modelled by spherical models (2.44). Then*

$$Z \in C_{d_A \min} \otimes C_{d_B} \implies \|\bar{Z}\|_1 \leq Z_{00}, \quad (2.71)$$

$$Z \in C_{d_A \max} \otimes C_{d_B} \implies \|\bar{Z}\|_\infty \leq Z_{00}, \quad (2.72)$$

where the cones are defined by (2.45), and $|\cdot|_1, |\cdot|_\infty$ denote the trace and operator norm, respectively. Furthermore, the above conditions are also sufficient when all cross terms Z_{i0}, Z_{0j} ($i, j \geq 1$) vanish.

Proof. Follows by putting together Proposition 2.25 and Example 2.2. \square

2.5 Minimal and maximal tensor product

2.5.1 Generalities

In Subsection 2.2.3, we have discussed two constructions that are fundamental in defining the composition rules of GPTs, i.e. minimal and maximal tensor products of cones. As we mentioned, these constructions were originally developed in [106, 107] in connection with the theory of tensor products of ordered Banach spaces. Although certainly meaningful from the mathematical point of view, the reader might wonder why we should care about them here, since this thesis is primarily about mathematical *physics*. The reason is simple: *bipartite states that are in the maximal but not in the minimal tensor product are physically entangled*, meaning that they can not be prepared with LOCC operations (2.24). Therefore, the question (i) discussed in Subsection 2.1.1 can be rephrased in the GPT setting as follows: *given two non-classical GPTs A and B , is it always the case that $C_{A \min} \otimes C_B \neq C_{A \max} \otimes C_B$?* This happens for quantum mechanical systems as we saw in Subsection 2.3.2, but we suspect that this is indeed the case in full generality. With the help of Definition 2.9, we formulate this conjecture as follows.

Conjecture 2.1. *If C_A and C_B are two non-simplicial cones, then*

$$C_{A \min} \otimes C_B \neq C_{A \max} \otimes C_B. \quad (2.73)$$

When the positive cones of two GPTs $A = (V_A, C_A, u_A)$ and $B = (V_B, C_B, u_B)$ satisfy $C_{A \min} \otimes C_B \neq C_{A \max} \otimes C_B$, we say that A and B are **entangleable**. Despite its outstanding physical and conceptual relevance, the question in Conjecture 2.1 seems to have been considerably overlooked. In fact, we could not find any clear statement of it anywhere in the rich literature on the subject of nonlocality in GPTs [132, 133, 134, 135, 136, 137, 138, 63]. The first to have investigated and solved a closely related question have been Namioka and Phelps. In their paper [8], they prove a weaker form of Conjecture 2.1 whose validity was apparently suggested to them by E. Effros. Namely, they show

that there is a particularly simple choice of a non-simplicial cone C_B such that for any other cone C_A , the equality in (2.73) is possible iff C_A is simplicial. This ‘test’ cone C_B can be taken as the gbit defined by (2.47). For details, we refer the reader to Subsection 2.5.5, where we will state and prove Namioka-Phelps’ theorem.

Before we proceed further, we want to discuss a slightly less general scenario than that treated in Conjecture 2.1. Namely, we can consider the special case of two copies of the same model, and ask whether they can be entangled. We thus state the following.

Conjecture 2.2. *If C is a non-simplicial cone, then*

$$C \otimes_{\min} C \neq C \otimes_{\max} C. \quad (2.74)$$

This is clearly weaker than Conjecture 2.1. The reason to consider this question is two-fold. First, we believe it captures all the essential features of the general problem, and a complete solution to it could very well lead to solving Conjecture 2.1. Second, it is technically a bit easier to work with, for a variety of reasons. For instance, we will see in Subsection 2.6.2 that we can show an explicit construction of an entangled state in the case of a single centrally symmetric model (thus establishing the validity of Conjecture 2.2 in this setting), but that the same construction for two different centrally symmetric models does not quite work.

Before we delve into the discussion of Conjecture 2.1 and of its cousin Conjecture 2.2, let us start with almost trivial sanity check. Namely, we want to verify that the assumption that both C_A and C_B are non-simplicial is essential for the claim of Conjecture 2.1 to hold. The following is a well-known result that confirms this intuition.

Lemma 2.27. [8]. *If either C_A or C_B is a simplicial cone, then*

$$C_A \otimes_{\min} C_B = C_A \otimes_{\max} C_B. \quad (2.75)$$

Proof. Since we work in finite dimension, a direct proof of the result that differs slightly from the one presented in [8] is available. Let us present this rather intuitive proof. We make no claim of originality, since the argument is extremely simple and part of the folklore of the field. Assume that C_A is a simplicial cone. According to Definition 2.27, this amounts to saying that

$$C_A = \text{co} \{v_1, \dots, v_d\}$$

for some basis $\{v_1, \dots, v_d\}$ of V_A . Let $\{v_1^*, \dots, v_d^*\}$ be the corresponding dual base of V_A^* . It is not difficult to realise that $v_i^* \in C_A^*$ for all i , since writing any $x \in C_A$ as $x = \sum_j \alpha_j v_j \in C_A$ for some coefficients $\alpha_j \geq 0$ ensures that $\langle v_i^*, x \rangle = \alpha_i \geq 0$.

Since the minimal tensor product is always included in the maximal one, we have to show only the opposite inclusion. For an arbitrary tensor $Z \in$

$C_A \otimes_{\max} C_B$ expressed as $Z = \sum_{i=1}^d v_i \otimes y_i$ for some vectors $y_i \in V_B$, the defining property of the maximal tensor product (2.15) guarantees that

$$0 \leq \langle v_i^* \otimes \varphi, Z \rangle = \langle \varphi, y_i \rangle$$

for all $\varphi \in C_B^*$. This implies that $y_i \in C_B^{**} = C_B$, thus ensuring that $Z \in \text{co}(C_A \otimes C_B) = C_A \otimes_{\min} C_B$. \square

In order to make some progress toward establishing Conjecture 2.1 at least in some special cases, we need to gain a better understanding of classical systems. The the purpose of the following Subsections 2.5.2, 2.5.3 and 2.5.4 is to introduce some basic notions of *lattice theory*, which treats ordered sets (in our case, ordered vector spaces) enjoying the additional property that every pair of elements admits an *infimum*, i.e. a minimal common upper bound, and a *supremum*, i.e. a maximal common lower bound. As we will see, this seemingly innocent property is a strong requirement to impose on an ordered vector space, in fact so strong that in finite dimension vector lattices corresponds exactly to simplicial cones as given by Definition 2.9. This alternative characterisation is heavily exploited in the proof of Namioka-Phelps' theorem (Theorem 2.37) in Subsection 2.5.5, which is the ultimate reason why we are about to devote some time reviewing the basics of lattice theory.

2.5.2 Lattices and vector lattices

Throughout this subsection, we discuss some properties of abstract mathematical structures called lattices, with special attention devoted to the case of vector lattices. The theory we will review here is covered by standard textbooks like [139] (for lattice theory) and [103, 100] (for vector lattices). A brief introduction to vector lattices can also be found in [140, §10]. Our exposition follows mostly [103, I]. Let us start with the following general definition.

Definition 2.28. A *lattice* is a partially ordered set (X, \leq) in which for every two elements $x, y \in X$ there exists the minimum simultaneous upper bound $x \vee y := \min \{z \in X : z \geq x \text{ and } z \geq y\}$ (supremum) and the maximum simultaneous lower bound $x \wedge y := \max \{z \in X : z \leq x \text{ and } z \leq y\}$ (infimum).

The above notion of lattice is fully general, and it will be helpful to have it at hand. However, in our case we will be mostly concerned with lattices on vector spaces. The relevant definition is as follows.

Definition 2.29. A *vector lattice* (also called a **Riesz space**) is an ordered vector space that is also a lattice (with respect to the same order relation).

It can be easily proven that if V is a vector lattice then its positive cone V_+ is necessarily generating. In fact, any $x \in V$ can be written as

$$x = (x - x \wedge 0) - (-x \wedge 0) \in V_+ - V_+.$$

Here we do not list all the elementary properties of vector lattices, for which we refer the reader to [139, I] or to [103, Theorem 1.17]. However, there are few simple observations we want to make because they will turn out to be useful for what follows.

Proposition 2.30. *Let V be a vector lattice. Then the infimum ε satisfies the following identities:*

(a) *translational invariance: $(x + z) \wedge (y + z) = x \wedge y + z$ for all $x, y, z \in V$;*

(b) *for all $x, y, z \geq 0$ one has $(x + y) \wedge z \leq x \wedge z + y \wedge z$.*

Analogous relations can be obtained for the supremum ε with the help of the identity $x\varepsilon y = -((-x) \wedge (-y))$.

Proof.

(a) On the one hand, $x \wedge y \leq x, y$ implies $x \wedge y + z \leq x + z, y + z$, from which we deduce $x \wedge y + z \leq (x + z) \wedge (y + z)$. Applying this latter relation with the substitutions $x \mapsto x + z, y \mapsto y + z$ and $z \mapsto -z$ gives also the opposite inequality, so that equality is established.

(b) Define $w := (x + y) \wedge z$ and observe that: (1) $w = (x + y) \wedge (x + z - x) = x + y \wedge (z - x) \leq x + y \wedge z$, because $b \leq c \Rightarrow a \wedge b \leq a \wedge c$; (2) $w \leq z \leq z + y \wedge z$. Since w is therefore a lower bound for both $x + y \wedge z$ and $z + y \wedge z$, we conclude that $w \leq (x + y \wedge z) \wedge (z + y \wedge z) = x \wedge z + y \wedge z$.

□

2.5.3 Decomposition and interpolation properties

Let us now turn our attention to the study of two important features that an ordered vector space can exhibit, namely the so-called decomposition and interpolation properties. Here we define them separately and then prove that they are equivalent, and in finite dimension even equivalent to the underlying ordered vector space being a vector lattice, provided that the positive cone is closed.

Definition 2.31. [103, Definition 1.50]. *An ordered vector space is said to have the **decomposition property** if for any finite families of positive vectors $x_i, y_j \geq 0$ ($i \in I, j \in J$ finite sets) satisfying $\sum_i x_i = \sum_j y_j$, there exists a matrix of positive vectors $z_{ij} \geq 0$ over $I \times J$ such that*

$$x_i = \sum_j z_{ij} \quad \forall i \in I, \quad (2.76)$$

$$y_j = \sum_i z_{ij} \quad \forall j \in J. \quad (2.77)$$

We will say that the space has the binary decomposition property if said positive decomposition can always be found provided that $|I|, |J| \leq 2$.

Definition 2.32. [103, Definition 1.52]. *An ordered vector space is said to have the interpolation property if for all $a_1, a_2 \leq b_1, b_2$ (all the four inequalities satisfied), there exists a vector c such that (with the same notation) $a_1, a_2 \leq c \leq b_1, b_2$.*

Remark. We did not distinguish a finite and a binary interpolation properties, because the two are elementarily seen to be equivalent.

Interestingly enough, the three aforementioned properties turn out to be equivalent in a fully general setting, as the next lemma due to Riesz [141] will show.

Lemma 2.33. [103, Theorem 1.54]. *Let V be an ordered vector space. The following are equivalent:*

- (a) V has the decomposition property;
- (b) V has the binary decomposition property;
- (c) V has the interpolation property.

Proof.

(a) \Rightarrow (b). Obvious.

(b) \Rightarrow (a). We show how to reconstruct the generic case of finite $|I|, |J|$ by separate induction on the two cardinalities. We start by looking at $|I| = 2$ and performing an induction on $|J| = m > 2$.

Consider $x_1, x_2, y_1, \dots, y_m \geq 0$ such that $x_1 + x_2 = \sum_{j=1}^m y_j$; we want to find an appropriate set of z vectors that decomposes this set. Since we assume that the claim has been established for $m - 1$, we can find $z'_{ij} \geq 0$ ($i = 1, 2, j = 1, \dots, m - 1$) that satisfy

$$\sum_{j=1}^{m-1} z'_{ij} = x_i,$$

$$z'_{1j} + z'_{2j} = \begin{cases} y_{m-1} + y_m & \text{if } j = m - 1, \\ y_j & \text{otherwise.} \end{cases}$$

The construction of the requested vectors z_{ij} ($i = 1, 2, j = 1, \dots, m$) is in two steps. If $j \leq m - 2$, we simply identify $z_{ij} := z'_{ij}$. In order to construct the remaining four elements, we can exploit the equality $z'_{1,m-1} + z'_{2,m-1} = y_{m-1} + y_m$ and the $|I| = |J| = 2$ decomposition property in order to find $z_{1,m-1}, z_{2,m-1}, z_{1m}, z_{2m} \geq 0$ such that

$$\begin{aligned} z'_{1,m-1} &= z_{1,m-1} + z_{1m}, \\ z'_{2,m-1} &= z_{2,m-1} + z_{2m}, \\ y_{m-1} &= z_{1,m-1} + z_{2,m-1}, \\ y_m &= z_{1m} + z_{2m}. \end{aligned}$$

It is very easy to verify that one has $x_i = \sum_{j=1}^m z_{ij}$ and $y_j = z_{1j} + z_{2j}$ for all $i = 1, 2$ and $j = 1, \dots, m$, respectively.

Finally, the induction on $|I| = n$ can be performed in a totally analogous way by adding one column after the other to the basic case $n = 2$. These additions are made possible by the above $2 \times m$ decomposition property.

(b) \Rightarrow (c). Given $a_1, a_2 \leq b_1, b_2$, define $x_1 := b_1 - a_1$, $x_2 := b_2 - a_2$, $y_1 := b_1 - a_2$, $y_2 := b_2 - a_1$. By hypothesis $x_1, x_2, y_1, y_2 \geq 0$, and moreover $x_1 + x_2 = b_1 + b_2 - a_1 - a_2 = y_1 + y_2$. Applying the binary decomposition property yields four positive vectors $z_{11}, z_{12}, z_{21}, z_{22} \geq 0$ such that

$$\begin{aligned} x_1 &= z_{11} + z_{12}, \\ x_2 &= z_{21} + z_{22}, \\ y_1 &= z_{11} + z_{21}, \\ y_2 &= z_{12} + z_{22}. \end{aligned}$$

We now define $c := z_{12} + a_1 = z_{21} + a_2$, that clearly satisfies $c \geq a_1, a_2$. Furthermore, $c \leq z_{12} + z_{11} + a_1 = b_1$ and analogously $c \leq b_2$, so the construction is complete.

(c) \Rightarrow (b). Consider arbitrary positive vectors $x_1, x_2, y_1, y_2 \geq 0$ such that $x_1 + x_2 = y_1 + y_2$. Since it is easy to verify that the four inequalities $0, x_1 - y_1 \leq x_1, y_2$ hold, by the interpolation property we can find $z_{12} \geq 0$ such that $0, x_1 - y_1 \leq z_{12} \leq x_1, y_2$. The other three elements of the z matrix are now fixed by the equalities in Definition 2.31:

$$\begin{aligned} z_{11} &:= x_1 - z_{12} \geq 0, \\ z_{22} &:= y_2 - z_{12} \geq 0, \\ z_{21} &:= x_2 - y_2 + z_{12} = y_1 - x_1 + z_{12} \geq 0. \end{aligned}$$

By construction, these four z vectors satisfy the sum constraints required by the decomposition property.

□

2.5.4 Decomposition lemma and converses thereof

If we spent so much time discussing the above properties is because they lie at the heart of the concept of vector lattice. In fact, as we mentioned, they are even equivalent to the concept of lattice for closed cones in finite dimension. Let us start with a general result, that holds in full generality, not only for finite-dimensional spaces.

Lemma 2.34 (Decomposition lemma). [103, Corollary 1.55]. *All vector lattices have the decomposition property.*

Proof. Thanks to the above discussion, it suffices to observe that a vector lattice has the interpolation property, since $a_1 \leq b_1, b_2$ implies $a_1 \leq b_1 \wedge b_2$, and analogously $a_2 \leq b_1 \wedge b_2$, so that $c := b_1 \wedge b_2$ is a valid choice. \square

Alternative proof (direct). We just prove the binary version of the decomposition property. Given $x_1, x_2, y_1, y_2 \geq 0$ such that $x_1 + x_2 = y_1 + y_2$, the equations to be satisfied are

$$\begin{aligned} x_1 &= z_{11} + z_{12}, \\ x_2 &= z_{21} + z_{22}, \\ y_1 &= z_{11} + z_{21}, \\ y_2 &= z_{12} + z_{22}. \end{aligned}$$

We define in the first place $z_{11} := x_1 \wedge y_1$, so that necessarily $z_{12} := x_1 - x_1 \wedge y_1$, $z_{21} := y_1 - x_1 \wedge y_1$ and finally $z_{22} := y_2 - x_1 + x_1 \wedge y_1 = x_2 - y_1 + x_1 \wedge y_1$ (where for the last equality we used $x_1 + x_2 = y_1 + y_2$). We have to prove that the expression we gave for z_{22} identifies indeed a positive vector. This can be done by applying the lower bounding technique in Proposition 2.30(b):

$$\begin{aligned} z_{22} &= x_2 - y_1 + x_1 \wedge y_1 \\ &\geq x_2 - y_1 + (x_1 + x_2) \wedge y_1 - x_2 \wedge y_1 \\ &= x_2 - y_1 + (y_1 + y_2) \wedge y_1 - x_2 \wedge y_1 \\ &= x_2 - y_1 + y_1 - x_2 \wedge y_1 \\ &= x_2 - x_2 \wedge y_1 \\ &\geq 0. \end{aligned}$$

Since it is apparent from the definitions that also $z_{11}, z_{12}, z_{21} \geq 0$, and these z vectors satisfy the required identities by construction, the proof is complete. \square

The connection between decomposition property and lattice axioms is so deep that in finite dimension the two are in fact equivalent under the only assumption of topological closeness of the positive cone. Moreover, finite-dimensional closed cones that define a lattice order are exactly the simplicial cones considered in Definition 2.9, a result that goes back to Yudin [142].

Theorem 2.35 (Yudin). [103, Theorem 3.21] or [140, Proposition 10.10]. *Let V be a finite-dimensional ordered vector space whose positive cone V_+ is closed and spanning. Then the following are equivalent:*

- (a) V is a vector lattice;
- (b) V has the decomposition property;
- (c) V_+ is simplicial.

Proof. We give on purpose a redundant proof of this important result. In what follows, $d = \dim V < \infty$.

(a) \Rightarrow (b). See Lemma 2.34.

(b) \Rightarrow (a). We have to prove that the decomposition property ensures that V is a lattice. Thanks to Lemma 2.33, we can instead assume that the interpolation property is satisfied. Consider $x, y \in V$ and let us prove that there exists a maximum simultaneous lower bound (the reasoning for the minimum simultaneous upper bound is totally analogous). First of all, denoting with $x = x_+ - x_-$ and $y = y_+ - y_-$ two decompositions such that $x_+, x_-, y_+, y_- \geq 0$ (which exist because V_+ is generating), we note that $x \wedge y$ (if it exists) can be sought inside the interval

$$K_0 := [-x_- - y_-, x_+ + y_+] = \{z \in V : -x_- - y_- \leq z \leq x_+ + y_+\}.$$

This is a consequence of the interpolation property: whenever $w \leq x, y$, taking into account that also $-x_- - y_- \leq x, y$, we can construct w' such that $-x_- - y_-, w \leq w' \leq x, y \leq x_+ + y_+$.

We now observe that any interval $[a, b]$ in a finite-dimensional ordered vector space is necessarily bounded. To show this fact, it suffices to consider a basis $\{\varphi_1, \dots, \varphi_d\}$ of the dual space V^* , and to prove that $\langle \varphi_i, z \rangle$ is bounded when $z \in [a, b]$, for all $i = 1, \dots, d$. Since the dual cone V_+^* is generating, we can take a basis in which each φ_i is a positive functional. Thanks to the positivity, from $a \leq z \leq b$ the relation $\langle \varphi_i, a \rangle \leq \langle \varphi_i, z \rangle \leq \langle \varphi_i, b \rangle$ follows immediately. In particular, $\langle \varphi_i, z \rangle$ is bounded for all $i = 1, \dots, d$ and for all $z \in [a, b]$.

Since V_+ is assumed to be closed, the intervals $[a, b]$ are also closed and thus compact (because in finite-dimensional spaces a set is compact iff it is closed and bounded). The (non-empty) set $K := \{z \in K_0 : z \leq x, y\}$ is compact, too, because it is a closed subset of a compact set. It can be shown that compact sets have always a maximal element in an ordered topological space, and in particular in a vector space ordered by a closed cone. This intuitive fact is usually referred to as Wallace's lemma [143].

Therefore, there must exist a maximal element z_0 of K . Thanks to the interpolation property, z_0 has to be indeed the maximum, because for any other $z \in K$ in the set the inequalities $z, z_0 \leq x, y$ guarantee the existence of z'_0 such that $z, z_0 \leq z'_0 \leq x, y$, which in turn implies $z'_0 = z_0$ and therefore $z_0 \geq z$.

(a) \Rightarrow (c). By Corollary 2.5, V_+ is the set of all finite sums of its extremal vectors. We now prove that there can not be more than d extremal rays in V_+ if this has to be a lattice. Since V_+ is also spanning, this will suffice to prove that their number is in fact exactly d . Suppose by contradiction that $x_0, \dots, x_d \geq 0$ are all extremal (and none of them is a multiple of another one). There must be a linear dependence relation $\sum_{j=0}^d \alpha_j x_j = 0$, so that for all $i = 0, \dots, d$ thanks to Proposition 2.30(b)

one obtains

$$0 = x_i \wedge \sum_{j=0}^n \alpha_j x_j \leq \sum_{j=0}^n \alpha_j (x_i \wedge x_j) = \alpha_i,$$

because if x_i and x_j are both extremal $x_i \wedge x_j \leq x_i, x_j$ must lie at the same time in the ray generated by x_i and in the one generated by x_j , and therefore has to be zero. We are left with $\sum_{i=0}^d \alpha_i x_i = 0$ and $\alpha_i \geq 0$ for all $i = 0, \dots, d$, which is possible only if $\alpha_i \equiv 0$ identically. In fact, since $x_i \geq 0$ and $x_i \neq 0$ for all i , if $\alpha_i > 0$ were true for some i we would obtain $\sum_{i=0}^d \alpha_i x_i \neq 0$, absurd.

(c) \Rightarrow (a). The order induced by a simplex-based cone $V_+ = \{\alpha_1 e_1 + \dots + \alpha_n e_n : \alpha_i \geq 0\}$ (where the e_i s constitute a basis) makes V a lattice. For instance, if $x = \sum_{i=1}^n \alpha_i e_i$ and $y = \sum_{i=1}^n \beta_i e_i$, it is easy to verify that $x \wedge y := \sum_{i=1}^n \min\{\alpha_i, \beta_i\} e_i$ is the least simultaneous upper bound of x and y .

(b) \Rightarrow (c). The following classic argument is a modification of the one at the very end of [140, §10]. Suppose by contradiction that $x_0, \dots, x_d \geq 0$ are all extremal vectors (and none of them is a multiple of another one). There will exist a linear dependence relation $\sum_{i=0}^d \alpha_i x_i = 0$ among said vectors. Let us rewrite it as $\sum_{i \in I} a_i = \sum_{j \in J} b_j$, where $I = \{0 \leq i \leq d : \alpha_i > 0\}$, $J = \{0 \leq j \leq d : \alpha_j < 0\}$ are both non-empty, $v_i := \alpha_i x_i \neq 0$ and $w_j := -\alpha_j x_j \neq 0$. Apply now the decomposition property in order to construct $z_{ij} \geq 0$ such that $\sum_j z_{ij} = v_i$ for all i and $\sum_i z_{ij} = w_j$ for all j . Since all the vectors v_i are extremal, we deduce immediately $z_{ij} \propto v_i$ and analogously $z_{ij} \propto w_j$, so that the existence of a single $z_{ij} \neq 0$ leads to the absurd conclusion $v_i \propto w_j$.

(c) \Rightarrow (b). The binary decomposition property for simplicial cones reduces immediately to the same property for the ordered space $(\mathbb{R}, \mathbb{R}_+)$, whose proof is an elementary exercise.

□

The equivalence expressed by the above Theorem 2.35 breaks down in infinite dimension. Ultimately, this happens because there is no converse to the decomposition lemma (Lemma 2.34) for generic ordered Banach spaces. For explicit examples showing this phenomenon in action, we refer the interested reader to [103, Examples 1.56-1.58]. A great deal of the existent literature on Banach lattices is devoted to finding suitable replacements for the crucial result expressed in Theorem 2.35. It is hopeless to make a complete list of references here, so we limit ourselves to mentioning the cornerstone result known as Ando's theorem. In brief, this states that under certain regularity assumptions on the cone (that it is spanning and *normal*, see Definition 1.17), the decomposition property at the level of the primal space is equivalent to that

on the dual space. For details, see the original paper by Ando [144] or the excellent presentation in [103, Theorem 2.47].

Let us now discuss some consequences of Theorem 2.35. As an immediate corollary, we find the following result.

Corollary 2.36. [103, Theorem 3.9]. *A finite-dimensional ordered vector space V whose positive cone is closed and generating is a vector lattice iff its dual V^* is a vector lattice.*

Proof. Follows by putting together Theorem 2.35 with the observation that a closed spanning cone is simplicial iff its dual is so, as discussed in Subsection 2.3.1. \square

The above consequence admits an appealing physical interpretation. Let us consider a GPT (V, C, u) . We remind the reader that two measurements $(e_i)_i, (f_j)_j \in \mathbf{M}$ are said to be **jointly measurable** (or simply **compatible**) if there is a third measurement $(g_{ij})_{ij} \in \mathbf{M}$ such that

$$e_i = \sum_j g_{ij} \quad \forall i, \quad (2.78)$$

$$f_j = \sum_i g_{ij} \quad \forall j. \quad (2.79)$$

This is probably not the most intuitive representation of what ‘compatibility’ of two measurements means. One could formulate an alternative notion of **coexistent** measurements, by requiring the existence of a third measurement $(g_k)_{k \in K}$ such that for all $i \in I$ and $j \in J$ there are subsets $K_i, K^j \subseteq K$ with the property that $\sum_{k \in K_i} g_k = e_i$ and $\sum_{k \in K^j} g_k = f_j$. While it turns out that joint measurability and coexistence are equivalent concepts for binary measurements [145, 146, 147], this equivalence does no longer hold already when $|I| = 2$ and $|J| = 3$. In what follows, we will have to deal mostly with binary measurements.

Equations (2.78) and (2.79) have the same form as those appearing in the definition of the decomposition property, (2.76) and (2.77). There is however a small difference, namely, that the sum $\sum_i e_i = \sum_j f_j = u$ is fixed to a specific value (i.e. the order unit), and can not be changed. A consequence of this fact is that we can not use Theorem 2.35 directly to conclude that every non-classical GPT admits two binary measurements that are not jointly measurable. This is however the case, as proven in [137].⁸

2.5.5 Namioka-Phelps theorem

Now that we have at hand the powerful Theorem 2.35, which provides an alternate characterisation of classical GPTs as defined in Definition 2.9, we

⁸However, from Theorem 2.35 it *does* follow – by applying a suitable rescaling – that in every non-classical theory there are measurements with at most 3 outcomes that are not jointly measurable, which is exactly the claim in the title of [137].

can turn our attention to reviewing the proof of a result by Namioka and Phelps [8, Theorem 1.4] that confirms Conjecture 2.1 at least in an important special case, namely when one of the two cones in question has a square base, i.e. is of the form (2.47).

Theorem 2.37 (Namioka-Phelps). *Let C be a closed spanning cone in a finite-dimensional real space V . Then the following are equivalent:*

(i) C is simplicial;

(ii) $C \otimes_{\min} \tilde{C} = C \otimes_{\max} \tilde{C}$ for all closed spanning finite-dimensional cones \tilde{C} ;

(iii) $C \otimes_{\min} C_{\square} = C \otimes_{\max} C_{\square}$, with C_{\square} being given by (2.47).

Proof. In light of Lemma 2.27, the only nontrivial implication is (iii) \Rightarrow (i). Employing also Theorem 2.35, it suffices to prove that if $C \otimes_{\min} C_{\square} = C \otimes_{\max} C_{\square}$ then C has the binary decomposition property. For simplicity, we will consider V as a vector space ordered by C .

Let us start with some notation. Denote the vectors generating the four extremal rays of C_{\square} by

$$\begin{aligned} v_1 &= (1, 1, 1), & v_2 &= (1, -1, 1), \\ v_3 &= (-1, 1, 1), & v_4 &= (-1, -1, 1). \end{aligned} \quad (2.80)$$

Observe that $v_4 = v_2 + v_3 - v_1$. The extremal rays of the dual cone C_{\square}^* are generated by

$$\begin{aligned} \varphi_1 &= (1, 0, 1) & \varphi_2 &= (-1, 0, 1), \\ \varphi_3 &= (0, 1, 1), & \varphi_4 &= (0, -1, 1). \end{aligned} \quad (2.81)$$

Now, take three arbitrary positive vectors $x_1, x_2, y_1 \geq 0$ in V such that $y_1 \leq x_1 + x_2$, and define the **Namioka-Phelps state**

$$Z_{NP} := (x_1 - y_1) \otimes v_1 + y_1 \otimes v_2 + x_2 \otimes v_3. \quad (2.82)$$

We want to check that $Z_{NP} \in C \otimes_{\max} C_{\square}$. This can be easily done by applying systematically all the extremal functionals φ_i ($i = 1, 2, 3, 4$) on the second subsystem, and by verifying that the resulting ‘partially evaluated’ vectors on the first subsystem belong to C . A simple graphical picture will help us understanding why this is the case.

With reference to Figure 2.1, we see that evaluating any of the functionals φ_i on Z_{NP} yields (up to a constant factor) a vector in V that is the sum of the

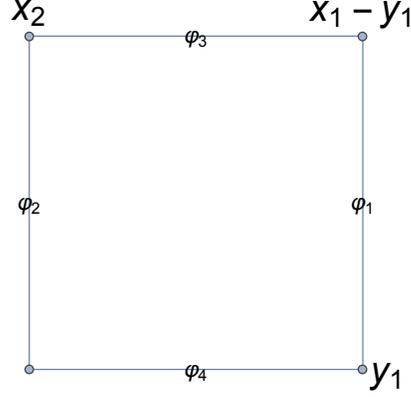


Figure 2.1: Schematic representation of the Namioka-Phelps state in (2.82). The vertices of the graph represent, in clockwise order from the upper right corner, v_1, v_2, v_4, v_3 . The vertices are labelled by the V vector they are associated with via tensor product. Edges are labelled by the functionals among those in (2.81) that take the maximal value on the corresponding straight lines.

labels of the two vertices contained in the edge identified by φ_i . In formulae,

$$\begin{aligned} \frac{1}{2} \langle \varphi_1^B, Z_{NP}^{AB} \rangle &= x_1^A \geq 0, \\ \frac{1}{2} \langle \varphi_2^B, Z_{NP}^{AB} \rangle &= x_2^A \geq 0, \\ \frac{1}{2} \langle \varphi_3^B, Z_{NP}^{AB} \rangle &= x_1^A - y_1^A + x_2^A = y_2^A \geq 0, \\ \frac{1}{2} \langle \varphi_4^B, Z_{NP}^{AB} \rangle &= y_1^A \geq 0. \end{aligned}$$

This is enough to show that $Z_{NP} \in C_{\max}^{\otimes} C_{\square}$. In fact, a consequence of the calculations we performed is that $\langle \lambda^A \otimes \varphi_i^B, Z_{NP}^{AB} \rangle \geq 0$ for all $i = 1, 2, 3, 4$ and for all $\lambda^A \in C^*$. Since every $\varphi^B \in C_{\square}^*$ is a non-negative combination of the extremal functionals φ_i (by Corollary 2.5), this shows that

$$\langle \lambda^A \otimes \varphi_i^B, Z_{NP}^{AB} \rangle \geq 0 \quad \forall \lambda^A \in C^*, \varphi^B \in C_{\square}^*,$$

implying that $Z_{NP} \in C_{\max}^{\otimes} C_{\square}$ by the very definition of maximal tensor product (2.15).

We now proceed to show that if $Z_{NP} \notin C_{\min}^{\otimes} C_{\square}$ then x_1, x_2, y_1, y_2 obey the binary decomposition property. To this purpose, we use Corollary 2.5 to write any tensor $Z \in C_{\min}^{\otimes} C_{\square}$ in the minimal tensor product as $Z = \sum_{i=1}^4 z_i \otimes v_i$,

where $z_i \geq 0$ for all $i = 1, 2, 3, 4$. In the case of Z_{NP} we would obtain the equality

$$\begin{aligned} Z_{NP} &= (x_1 - y_1) \otimes v_1 + y_1 \otimes v_2 + x_2 \otimes v_3 \\ &= z_1 \otimes v_1 + z_2 \otimes v_2 + z_3 \otimes v_3 + z_4 \otimes v_4 \\ &= z_1 \otimes v_1 + z_2 \otimes v_2 + z_3 \otimes v_3 + z_4 \otimes (v_2 + v_3 - v_1) \\ &= (z_1 - z_4) \otimes v_1 + (z_2 + z_4) \otimes v_2 + (z_3 + z_4) \otimes v_3. \end{aligned}$$

Since v_1, v_2, v_3 are linearly independent, this is possible only if

$$\begin{aligned} x_1 - y_1 &= z_1 - z_4, \\ y_1 &= z_2 + z_4, \\ x_2 &= z_3 + z_4. \end{aligned}$$

Adding the first two equations and deducing the identity for $y_2 = x_1 + x_2 - y_1$ yields

$$\begin{aligned} x_1 &= z_1 + z_2, \\ x_2 &= z_3 + z_4, \\ y_1 &= z_2 + z_4, \\ y_2 &= z_1 + z_3. \end{aligned}$$

Up to relabelling, this are the equations that define the binary decomposition property. We are then led to conclude that if $Z_{NP} \in C_{\min}^{\otimes} C_{\square}$ for all $x_1, x_2, y_1, y_2 \geq 0$ such that $x_1 + x_2 = y_1 + y_2$, then the cone C satisfies the binary decomposition property, and thus (thanks to Theorem 2.35) is simplicial. \square

From our point of view, the above results is important primarily because it confirms Conjecture 2.1 when one of the two cones is the simplest non-classical theory, i.e. the gbit discussed in Subsection 2.3.4. We state this observation as a corollary.

Corollary 2.38. *Conjecture 2.1 holds when one of the two cones has a square basis (2.47).*

However, the importance of Theorem 2.37 lies also in the proof technique itself. In fact, the Namioka-Phelps state (2.82) is to the best of our knowledge the first and most general example of an entangled state. Its generality stems from the fact that the construction can be applied to all cones C_A , provided that C_B is fixed to be the square-based cone in (2.47). In this sense there is a lot of work to be done yet, in order to extend the construction (if at all possible) to the case of an arbitrary C_B .

Let us stress here that (2.82) anyway encompasses as special cases other examples of entangled states that have been discovered much later. For instance, the so-called *PR-box* [48] can be seen as the state of $G_2^{\otimes} G_2$ obtained from (2.82) by taking $C = C_{\square}$, $x_1 = \frac{1}{2}v_2$, $x_2 = \frac{1}{2}v_3$, $y_1 = \frac{1}{2}v_1$ and $y_2 = \frac{1}{2}v_4$.

2.6 Beyond Namioka-Phelps

Throughout this section, we explore some possible generalisations and extensions of the Namioka-Phelps theorem, Theorem 2.37. In doing so, we will review and formalise many of the questions listed in Subsection 2.1.1. To the best of my knowledge, most of what I will present here constitutes an original contribution.

2.6.1 Nonlocality in maximal tensor products

As we mentioned in Subsection 2.1.1, point (i), there is a stronger version of Conjecture 2.1 that is perhaps more operationally motivated. Namely, instead of asking whether two non-classical GPTs are entangleable, we could rather look at the presence of nonlocal correlations. With the language of Definition 2.14, we are thus asking the following question.

Question 2.3. *Given two non-classical GPTs $A = (V_A, C_A, u_A)$ and $B = (V_B, C_B, u_B)$, is it always the case that $A \otimes_{\max} B$ violates a Bell inequality according to Definition 2.14?*

We did not make this question a conjecture because even if Conjecture 2.1 holds true, it may well be the case that the entanglement contained in certain exotic models can not be ‘extracted’ via measurements and transformed into nonlocal correlations. This is, however, pure speculation, as to the present day we do not have evidence of such behaviour in any explicit example.

As the reader may know, the simplest Bell inequality of all is the so-called **CHSH inequality** [148]. In the terminology we fixed in Definition 2.13, CHSH separates point of $G_2^2 \otimes_{\max} G_2^2$ from local hidden variable models in $G_2^2 \otimes_{\min} G_2^2$. In particular, it deals with the case of two (2,2)-apparatuses acting on the two shares of a bipartite state. We change slightly the notation of Definition 2.12 to adapt it to the most common conventions.

Let us denote the physical state used to generate the correlations as $\omega \in C_A \otimes_{\max} C_B$, with $\langle u_A \otimes u_B, \omega_{AB} \rangle = 1$. In most literature on the subject, the settings of the apparatuses are labelled by $\mathbf{x} = 0, 1$ and $\mathbf{y} = 0, 1$ on the A and B side, respectively. The outcomes of the measurements are instead indexed by $\mathbf{a} = \pm 1$ and $\mathbf{b} = \pm 1$. With this conventions, the CHSH inequality takes the form

$$\langle \mathbb{B}, \omega \rangle := \sum_{\mathbf{a}, \mathbf{b}, \mathbf{x}, \mathbf{y}} \mathbf{a}\mathbf{b}(1 - 2\mathbf{x}\mathbf{y}) \langle e_{\mathbf{x}\mathbf{a}} \otimes f_{\mathbf{y}\mathbf{b}}, \omega \rangle \leq 2. \quad (2.83)$$

The functional acting on ω on the left-hand side of the above inequality is called **Bell functional**. A state ω of $A \otimes_{\max} B$ for which (2.83) is not obeyed is said to **violate CHSH**. In particular, any such state violates a Bell inequality according to Definition 2.14, but the converse is not a priori true, since it is well-known that for higher values of n, k, n', k' there are many more Bell inequalities coming into play [113].

We have seen that the state (2.82) considered by Namioka and Phelps is basically the only explicit construction of an entangled state that works for all non-classical GPTs on the A system, provided that $B = G_2$ is the gbit discussed in Subsection 2.3.4. It is then natural to wonder whether such states violates any Bell inequality. The first original result we present here is the observation that *the Namioka-Phelps state indeed violates CHSH, for all non-classical GPTs A .*

Before we state and prove this curious result, let us discuss some consequences of the non-classicality of a GPT A . By the results in [137], we know that any such theory admits two binary measurements that are not jointly measurable (see the discussion at the end of Subsection 2.5.4). The concept of incompatibility is important in this context because it is not difficult to realise that *compatible measurements can not violate any Bell inequality* [149]. There is an intuitive way of *quantifying* how much two binary measurements fail to be jointly measurable via a convex optimisation program. This idea was introduced in [150]⁹, and subsequently translated into the GPT framework [133, 134, 135, 137, 63].

Given two binary measurements $(e, u - e)$ and $(f, u - f)$ on a GPT $A = (V, C, u)$, by direct inspection of (2.78) and (2.79) we see that they are jointly measurable iff there is $g \in V^*$ such that

$$0, e + f - u \leq g \leq e, f. \quad (2.84)$$

Therefore, one can consider the convex optimisation program [150, 135]

$$\begin{aligned} t[e, f] &:= \inf_{t, g} t \\ \text{s.t.} \quad &0, e + f - u \leq g \leq e + tu, f + tu. \end{aligned} \quad (2.85)$$

In what follows, $t[e, f]$ will be called the **incompatibility degree** or simply **incompatibility**.¹⁰ By construction, we see that the two measurements defined by e and f are jointly measurable iff $t[e, f] \leq 0$. In general, the range of $t[e, f]$ is the interval $[-1/2, 1/2]$. It is indeed very easy to see that (t, g) has a chance of being a feasible point for the program in (2.85) only when $t \geq -1/2$, and that $t = 1/2, g = \frac{1}{2}(e + f)$ is always feasible. The extremal value $-1/2$ is achieved for instance by $e = f = u/2$, while $1/2$ is achieved by the two measurements on a gbit identified, in the notation of (2.81), by $e = \frac{1}{2}\varphi_1$ and $f = \frac{1}{2}\varphi_3$ [135].

As shown in [150, 135], $t[e, f]$ has an interpretation in terms of the minimal amount of ‘noise’ that is needed in order to make the two measurements $(e, u - e)$ and $(f, u - f)$ jointly measurable. Namely, the maximal λ such that $\lambda e + \frac{1-\lambda}{2}u$ and $\lambda f + \frac{1-\lambda}{2}u$ identify two jointly measurable measurements is $\lambda[e, f] = \frac{1}{1+2t[e, f]}$.

⁹In this case, the convex program becomes a semidefinite program, or SDP.

¹⁰Other authors variously call closely related measures ‘unsharp parameter’ [133], ‘degree of incompatibility’ [137], and so on. We follow the terminology of [63].

Exactly as their cousins SDPs, also convex optimisation programs come with a dual. In our case, the dual to (2.85) can be taken in exactly the same way as in [150]. We obtain

$$\begin{aligned} t[e, f] &= \sup_{x, y, z} \langle e + f - u, z \rangle - \langle e, x \rangle - \langle f, y \rangle \\ \text{s.t.} \quad &x, y, z \geq 0, \\ &z \leq x + y, \\ &\langle u, x + y \rangle = 1. \end{aligned} \tag{2.86}$$

Observe that since the dual program is strictly feasible, its value coincides with that of the primal [105].

Since every non-classical GPT A admits two binary measurements that are not jointly measurable, it is clear that the **global incompatibility**

$$T[A] := \sup_{e, f \in [0, u]} t[e, f] \tag{2.87}$$

satisfies $T[A] > 0$ iff A is non-classical. Using the dual program (2.86), we can rewrite

$$\begin{aligned} T[A] &= \sup_{e, f \in [0, u]} t[e, f] \\ &= \sup_{e, f \in [0, u]} \sup_{\substack{x, y, z \geq 0 \\ z \leq x + y \\ \langle u, x + y \rangle = 1}} \{ \langle e, z - x \rangle + \langle f, z - y \rangle - \langle u, z \rangle \} \\ &\stackrel{(1)}{=} \sup_{\substack{x, y, z \geq 0 \\ z \leq x + y \\ \langle u, x + y \rangle = 1}} \sup_{e, f \in [0, u]} \{ \langle e, z - x \rangle + \langle f, z - y \rangle - \langle u, z \rangle \} \\ &\stackrel{(2)}{=} \sup_{\substack{x, y, z \geq 0 \\ z \leq x + y \\ \langle u, x + y \rangle = 1}} \sup_{\varphi, \eta \in [-u, u]} \frac{1}{2} \{ \langle \varphi, z - x \rangle + \langle \eta, z - y \rangle - 1 \} \\ &\stackrel{(3)}{=} \sup_{\substack{x, y, z \geq 0 \\ z \leq x + y \\ \langle u, x + y \rangle = 1}} \frac{1}{2} \{ \|z - x\| + \|z - y\| - 1 \}. \end{aligned}$$

The justification of the above calculation is as follows: (1) exchanging two suprema is always allowed; (2) we substituted $\varphi := 2e - u$ and $\eta := 2f - u$; (3) we employed the expression (2.9) for the base norm. In conclusion, we have found that the global incompatibility measure can be written in the form

$$T[A] = \max_{\substack{x, y, z \geq 0 \\ z \leq x + y \\ \langle u, x + y \rangle = 1}} \frac{1}{2} \{ \|z - x\| + \|z - y\| - 1 \}, \tag{2.88}$$

where we exchanged the supremum with a maximum by virtue of the compactness of the domain. Upon completion of this chapter, we realised that this expression for $T[A]$ was independently found in [63, Proposition 6].

We now proceed to show that for all GPTs A , the incompatibility $T[A]$ quantifies exactly the violation of the CHSH inequality (2.83) exhibited by the best Namioka-Phelps state (and indeed by any state in the composite system).

Theorem 2.39. *For A an arbitrary GPT, consider the composite system $A \otimes_{\max} G_2$. The following holds.*

- (a) *The value of the Bell functional in (2.83) does not exceed $2(1 + 2T[A])$, for all bipartite states ω [133].*
- (b) *There is a normalised state Z_{NP} of Namioka-Phelps kind (2.82) such that $\langle \mathbb{B}, Z_{NP} \rangle = 2(1 + 2T[A])$.*
- (c) *In particular, all non-classical theories A are such that $A \otimes_{\max} G_2$ violates CHSH. The answer to Question 2.3 is thus affirmative when $B = G_2$.*

Proof. Let us prove the various claims one by one.

- (a) In the GPT setting, this claim was originally proven in [133], the main idea coming once again from [150]. Let us repeat the argument here for the convenience of the reader. Denote by $\mathbb{B}_{e,f}$ the Bell functional (2.83) corresponding to some measurements $(e, u - e)$ and $(f, u - f)$ on the first subsystem, and some unspecified but fixed measurements on the second subsystem. One observes that under the transformation rule $e' := \lambda e + \frac{1-\lambda}{2}u$, $f' := \lambda f + \frac{1-\lambda}{2}u$ the Bell functional behaves according to

$$\mathbb{B}_{e',f'} = \lambda \mathbb{B}_{e,f}.$$

Since we saw that for $\lambda = \lambda[e, f] = \frac{1}{1+2\lambda[e, f]}$ the transformed effects e', f' give rise to compatible measurements [135], and that compatible measurements can never lead to a violation of a Bell inequality, we deduce that

$$\frac{B_{e,f}}{1 + 2\lambda[e, f]} \leq 2,$$

which yields the claim upon maximisation over $e, f \in [0, u]$.

- (b) For arbitrary $x_1, x_2, y_1, y_2 \geq 0$ such that $x_1 + x_2 = y_1 + y_2$ is a normalised state, i.e. such that $\langle u, x_1 + x_2 \rangle = 1$, construct the Namioka-Phelps state Z_{NP} as in (2.82). We want to choose appropriate measurements on A and $B = G_2$ such that the associated Bell functional \mathbb{B} of (2.83) evaluates to $\langle \mathbb{B}, Z_{NP} \rangle = 2(1 + 2T[A])$. To this purpose, the measurements on the gbit that constitutes the second subsystem are chosen as follows. Adopting the notation of (2.83) and (2.81), we take

$$\begin{aligned} f_{0+} &= \frac{1}{2}\varphi_1, & f_{0-} &= \frac{1}{2}\varphi_2, \\ f_{1+} &= \frac{1}{2}\varphi_4, & f_{1-} &= \frac{1}{2}\varphi_3. \end{aligned} \tag{2.89}$$

As in the proof of Theorem 2.37, one verifies that partial evaluation of these functionals on Z_{NP} yields

$$\begin{aligned} \langle f_{0+}^B, Z_{NP}^{AB} \rangle &= x_1^A, & \langle f_{0-}^B, Z_{NP}^{AB} \rangle &= x_2^A, \\ \langle f_{1+}^B, Z_{NP}^{AB} \rangle &= y_1^A, & \langle f_{1-}^B, Z_{NP}^{AB} \rangle &= y_2^A. \end{aligned} \quad (2.90)$$

Equipped with these relations, we can evaluate all the expectation values $\langle e_{\mathbf{xa}} \otimes f_{\mathbf{yb}}, Z_{NP} \rangle$, for all $\mathbf{x}, \mathbf{y} = 0, 1$ and $\mathbf{a}, \mathbf{b} = \pm$. These calculations are reported in Table 2.1.

$\mathbf{ab} \backslash \mathbf{xy}$	++	+-	-+	--
00	$\langle e_{0+}, x_1 \rangle$	$\langle e_{0+}, x_2 \rangle$	$\langle e_{0-}, x_1 \rangle$	$\langle e_{0-}, x_2 \rangle$
01	$\langle e_{0+}, y_1 \rangle$	$\langle e_{0+}, y_2 \rangle$	$\langle e_{0-}, y_1 \rangle$	$\langle e_{0-}, y_2 \rangle$
10	$\langle e_{1+}, x_1 \rangle$	$\langle e_{1+}, x_2 \rangle$	$\langle e_{1-}, x_1 \rangle$	$\langle e_{1-}, x_2 \rangle$
11	$\langle e_{1+}, y_1 \rangle$	$\langle e_{1+}, y_2 \rangle$	$\langle e_{1-}, y_1 \rangle$	$\langle e_{1-}, y_2 \rangle$

Table 2.1: Evaluation of all the terms $\langle e_{\mathbf{xa}} \otimes f_{\mathbf{yb}}, Z_{NP} \rangle$ appearing in (2.83). Here, Z_{NP} is the Namioka-Phelps state in (2.82), while the measurements on the second subsystem are chosen according to (2.89).

With the notation of (2.83), we deduce that

$$\begin{aligned} \langle \mathbb{B}, Z_{NP} \rangle &= \sum_{\mathbf{a}, \mathbf{b}, \mathbf{x}, \mathbf{y}} \mathbf{ab}(1 - 2\mathbf{xy}) \langle e_{\mathbf{xa}} \otimes f_{\mathbf{yb}}, Z_{NP} \rangle \\ &= \langle e_{0+} - e_{0-}, x_1 - x_2 + y_1 - y_2 \rangle \\ &\quad + \langle e_{1+} - e_{1-}, x_1 - x_2 - y_1 + y_2 \rangle \\ &\leq \|x_1 - x_2 + y_1 - y_2\| + \|x_1 - x_2 - y_1 + y_2\|. \end{aligned} \quad (2.91)$$

Observe that the upper bound in the last line is achievable by a careful choice of $e_{0+}, e_{1+} \in [0, u]$, since for instance $e_{0+} - e_{0-} \in [-u, u]$, and (2.9) holds. Now, let x, y, z be the vectors that achieve the maximum in (2.86). We have $x, y, z \geq 0$, $z \leq x + y$, and $\langle u, x + y \rangle = 1$, i.e. $x + y$ is a legitimate state. Construct the Namioka-Phelps state in (2.82) with $x_1 := x$, $x_2 := y$, $y_1 := z$, and $y_2 := x + y - z$, so that $x_1 + x_2 = y_1 + y_2$ is a legitimate state as required. The upper bound in (2.91) then reads

$$\begin{aligned} \|x_1 - x_2 + y_1 - y_2\| + \|x_1 - x_2 - y_1 + y_2\| &= 2\|z - y\| + 2\|z - x\| \\ &= 2(1 + 2T[A]), \end{aligned}$$

which concludes the proof of claim (b).

- (c) Thanks to the results of [137], we know that any non-classical GPT admits two incompatible binary measurements. Therefore, $T[A] > 0$ for all

non-classical theories A . By claim (b), this implies that there is a Namioka-Phelps state Z of $A \otimes_{\max} G_2$ that violates CHSH (2.83). This proves Question 2.3 when $B = G_2$ is a gbit.

□

The above Theorem 2.39 provides good evidence that the answer to Question 2.3 might be affirmative in general, after all. Further research into this central problem will be conducted elsewhere.

2.6.2 Universal entangleability of centrally symmetric models

This subsection is concerned with proving Conjecture 2.2 in the special case when both $A = B$ is a centrally symmetric model in the sense of Definition 2.11. After some time spent in introducing the necessary mathematical machinery, this goal is accomplished by Theorem 2.40. In our view, pursuing this programme is important for several good reasons. First, this will provide us with some evidence that Conjecture 2.1 or at least its weaker version Conjecture 2.2 holds also for high-dimensional GPTs, besides the simplest case treated in Theorem 2.37, in which the dimension of the second subsystem is fixed to the smallest value that does not trivialise the problem, i.e. $\dim B = 3$. Second, while investigating this question we will develop the only other construction of an entangled state that we are aware of, besides that of Namioka and Phelps. Such construction is semi-universal, in the sense that works for all centrally symmetric models, but it is not obvious how to extend it to the general setting. Last, but not least, in spite of the fact that our proof technique uses tools that are tailored to the centrally symmetric case, we will see in Chapter 3 that the same tools – for some a priori unclear reason – will turn out to be extremely useful in tackling general problems that have nothing to do with the special class of centrally symmetric models.

Furthermore, as a by-product of the proof we will obtain some novel mathematical insights into the theory of tensor norms. In fact, to the extent of our knowledge the content of Theorem 2.40 was not known to the functional analysis community. Building upon this, in Chapter 3 we will go a bit more in depth and prove a complementary result (Theorem 3.18) that might also be of independent interest in functional analysis.

For notation and conventions on centrally symmetric models, we refer the reader to Subsection 2.3.3. Recall that if the local systems A and B are modelled by two centrally symmetric GPTs of dimensions $n + 1$ and $m + 1$, states of the bipartite system AB can be represented via $(n + 1) \times (m + 1)$ matrices. A particularly important class of bipartite states are of those of the form $\omega = U_* + \widehat{Z}$, where $U = u_{A_*} u_{B_*}^T$ with $u_* = (1, 0, \dots, 0)^T$ (the length of the vector depending on the dimension of the system), and $Z \in \mathbb{R}^{n \times m}$ occupies the bottom right corner of ω .

The separability problem in centrally symmetric models has already been faced in Subsection 2.4.4, and our starting point is Proposition 2.25. A remarkable consequence of this result is that a state of the form $\omega = U_* + \widehat{Z}$ satisfies $\omega \in C_{A \max} \otimes C_B$ iff $|Z|_\varepsilon \leq 1$, and is separable iff $|Z|_\pi \leq 1$. Therefore, to construct an entangled state, which is our ultimate goal if we want to show that Conjecture 2.1 holds, it suffices to provide an example of a tensor Z satisfying $|Z|_\varepsilon < |Z|_\pi$. Once such a Z has been constructed, and we normalise it as to take $|Z|_\varepsilon \leq 1$, it is clear that $\omega = U_* + \widehat{Z}$ will be an entangled state. Of course, the problem is nontrivial only when $n, m \geq 2$ (that is, when Z is a matrix of size at least 2×2).

Note. Since the problem is now completely independent from the framework of centrally symmetric GPTs, we will denote the norms on the two sections \mathbb{R}^n and \mathbb{R}^m with the usual symbol $\|\cdot\|$, instead of $|\cdot|$, which is reserved for the restriction of the base norm of a centrally symmetric model to the section $\{(0, \bar{x}) : x \in \mathbb{R}^{d-1}\}$, see (2.40).

It does not take long to convince ourselves that tensors Z satisfying $\|Z\|_\varepsilon < \|Z\|_\pi$ should always exist, as it seems really implausible that in (2.1) one can always find functionals φ, λ achieving $\langle \varphi, x_i \rangle = \|x_i\|$ and $\langle \lambda, y_i \rangle = \|y_i\|$ for all vectors x_i, y_i appearing in some decomposition of Z as in (2.2). A simple example of this phenomenon is provided by the paradigmatic example of two Euclidean spaces (Example 2.2). In that case, the injective norm coincides with the matrix operator norm (i.e. the largest singular value), while the projective norm is nothing but the trace norm (i.e. the sum of all singular values). Apart from those of rank-one, all other matrices exhibit a gap between the two, the maximal such gap being displayed by the identity. Our intuition was thus correct in this case.

One might be tempted to try to generalise this reasoning straight away for the case of two arbitrary norms. However, a little of thought shows that the reason why we were able to identify a gap between the two norms in the Euclidean case is that we knew what their expressions in terms of the list of singular values were. That is, that proof strategy rests crucially on the singular value decomposition theorem (and on the Euclidean nature of the local norms, of course). Despite belonging to the field of standard linear algebra, this is in turn a highly nontrivial result. This is, in a nutshell, the reason why the same statement $\|\cdot\|_\varepsilon \neq \|\cdot\|_\pi$ (which is the content of Theorem 2.40) requires a so much more careful (and involved) proof in the case of arbitrary Banach spaces.

In light of the above discussion, the main goal of this subsection is the following: *given two arbitrary Banach spaces V, W of dimensions $n, m \geq 2$, show that there exists a tensor $Z \in V \otimes W$ with the property that $\|Z\|_\varepsilon < \|Z\|_\pi$.*¹¹ Following the discussion in Subsection 2.1.1(ii), we can make this

¹¹Of course for specific tensors (e.g. product ones) it may happen that the two norms are the same.

question quantitative and define the universal constants

$$R(n, m) := \inf_{\substack{\dim V_A = n \\ \dim V_B = m}} \sup_{0 \neq Z \in V_A \otimes V_B} \frac{\|Z\|_\pi}{\|Z\|_\varepsilon}, \quad (2.92)$$

$$S(n) := \inf_{\dim V = n} \sup_{0 \neq Z \in V \otimes V} \frac{\|Z\|_\pi}{\|Z\|_\varepsilon}. \quad (2.93)$$

Here, the infima are taken over all Banach spaces V_A, V_B, V of fixed dimensions. We dub the above functions **inseparability ratios**. These quantities are universal in the sense that they characterise the theory of Banach spaces itself rather than specific examples.

A more careful analysis of the inseparability ratios will be conducted elsewhere, as now we want to focus on the question of establishing the validity of Conjectures 2.1 and 2.2 for centrally symmetric models. In the language we have been developing so far, our central question becomes the following.

Question 2.4. *Do the universal constants defined in (2.92) satisfy $R(n, m) > 1$ for all $n, m \geq 2$?*

As we mentioned earlier, we will only deal with the simplest case $A = B$ and $V_A = V_B = V$. This corresponds to focusing on Conjecture 2.2 instead of its stronger version Conjecture 2.1. At the level of inseparability ratios, we are thus looking at Question 2.4 with $S(n)$ instead of $R(n, m)$. We believe it possible to extend the proof strategy we will discuss to this latter case, however doing so does not boil down to a simple change of notation, and instead requires a more careful manipulation of the technical tools we use. We are now ready to state and prove the main result of this section.

Theorem 2.40. *There exists a universal constant $c > 1$ such that for all $n \geq 2$ one has $S(n) \geq c$. A possible choice is $c = 1 + \frac{1}{29+12\sqrt{6}}$.*

Proof. Let V be two Banach space of dimension n . In what follows, we will denote with τ_{ij} ($i, j = 1, 2$) the four entries of the matrix

$$\tau := \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}. \quad (2.94)$$

For arbitrary vectors $x_1, x_2 \in V$, consider the tensor

$$Z[x_1, x_2] := \sum_{i,j=1,2} \tau_{ij} x_i \otimes x_j \in V \otimes V. \quad (2.95)$$

An analogous construction can be performed at the level of the duals, in which case we adopt a different notation, that is, for $\varphi_1, \varphi_2 \in V^*$ we define

$$H[\varphi_1, \varphi_2] := \sum_{i,j=1,2} \tau_{ij} \varphi_i \otimes \varphi_j \in V^* \otimes V^*. \quad (2.96)$$

It is easy to show that $\|Z[x_1, x_2]\|_\varepsilon \leq 2$ whenever $\|x_i\| \leq 1$ for $i = 1, 2$. This is an immediate consequence of the fact that $|\alpha_i|, |\beta_j| \leq 1$ ensures that $\sum_{ij} \tau_{ij} \alpha_i \beta_j \leq 2$. Analogously, $\|\varphi_i\|_* \leq 1$ implies that

$$\|H[\varphi_1, \varphi_2]\|_{\pi*} = \|H[\varphi_1, \varphi_2]\|_{*\varepsilon} \leq 2, \quad (2.97)$$

where we employed also (2.3).

Now, we use Auerbach's lemma (Lemma 2.1) to construct an Auerbach basis $\{v_1, \dots, v_n\}$ on V . This means the following:

$$\langle v_i^*, v_j \rangle = \delta_{ij}, \quad \|v_i^*\|_*, \|v_j\| \leq 1 \quad \forall i = 1, \dots, n, j = 1, \dots, m. \quad (2.98)$$

Denote by P the Auerbach-orthogonal projector onto $\text{Span}\{v_1^*, v_2^*\}$, whose action is defined by

$$P\varphi := v_1^* \langle \varphi, v_1 \rangle + v_2^* \langle \varphi, v_2 \rangle. \quad (2.99)$$

As is easy to see, for all $\varphi \in V^*$ and $x \in \text{span}\{v_1, v_2\}$ one has $\langle \varphi, x \rangle = \langle P\varphi, x \rangle$. Define

$$1 + k := \max\{|\alpha_1| + |\alpha_2| : \exists \varphi \in V^* : \|\varphi\|_* \leq 1, P\varphi = \alpha_1 v_1^* + \alpha_2 v_2^*\}, \quad (2.100)$$

so that $0 \leq k_V \leq 1$. Intuitively, we are looking at the convex body given by the projection of the global unit ball of the dual norm $\|\cdot\|_*$ onto the subspace spanned by v_1^*, v_2^* , and determining the size of the smallest 'tilted square' containing such a body. Up to changing sign to v_1^* or v_2^* (and to the corresponding primal vector) and up to swapping them, we can suppose that the α_1, α_2 achieving the maximum in (2.100) are non-negative and that $\alpha_1 \leq \alpha_2$.

We now construct the Z tensor as in (2.95) by taking $x_1 = v_1, x_2 = v_2$. Lemma 2.42 below implies that

$$\|Z[v_1, v_2]\|_\varepsilon \leq 2 - (1 - k)^2. \quad (2.101)$$

To see why this is the case, consider an arbitrary product functional $\eta \otimes \xi$ entering the maximisation that defines the injective norm as in (2.1), so that $\|\eta\|_*, \|\xi\|_* \leq 1$. We can expand both η and ξ in the dual Auerbach basis, obtaining $\eta = \sum_{i=1}^n \alpha_i v_i^*$ and $\xi = \sum_{j=1}^n \beta_j v_j^*$. Clearly, we have $|\alpha_i|, |\beta_j| \leq 1$ for all i and j , because for instance $|\alpha_i| = |\langle \eta, v_i \rangle| \leq \|\eta\|_* \|v_i\| \leq 1$. Furthermore, if P is defined as in (2.99) one has $P\eta = \alpha_1 v_1^* + \alpha_2 v_2^*$. By definition of k , this implies that $|\alpha_1| + |\alpha_2| \leq 1 + k$. The same is naturally true on the second space, so that $|\beta_1| + |\beta_2| \leq 1 + k$ holds as well. Thanks to Lemma 2.42, one obtains

$$\langle \varphi \otimes \lambda, Z[v_1, v_2] \rangle = \alpha_1(\beta_1 + \beta_2) + \alpha_2(\beta_1 - \beta_2) \leq f(k) = 2 - (1 - k)^2.$$

Since this holds for all functionals η, ξ such that $\|\eta\|_*, \|\xi\|_* \leq 1$, we conclude that $\|Z[v_1, v_2]\|_\varepsilon \leq 2 - (1 - k)^2$, as claimed.

Next, consider the tensor $H[v_1^*, v_2^*]$ constructed out of the first two functionals of the dual Auerbach basis as in (2.96). Since (2.97) holds, we can

estimate the projective norm of $Z[v_1, v_2]$ by plugging the ansatz $\frac{1}{2}H[v_1^*, v_2^*]$ into the generic formula (1.1). This yields

$$\begin{aligned} \|Z[v_1, v_2]\|_\pi &\geq \frac{1}{2} \|H[v_1^*, v_2^*]\|_{\pi^*} \|Z[v_1, v_2]\|_\pi \\ &\geq \frac{1}{2} \langle H[v_1^*, v_2^*], Z[v_1, v_2] \rangle \\ &= 2, \end{aligned} \tag{2.102}$$

where we used the definitions (2.95) and (2.96) of Z and H together with the properties (2.98) of the Auerbach basis.

Now that we have an upper bound for the injective norm (2.101) and a lower bound for the projective one (2.102), we can hope to prove the existence of a gap between the two. Unfortunately, what we have in hand right now is not enough to conclude, because the right-hand side of (2.101) can coincide with that of (2.102) when $k = 1$. We then look for a new lower estimate of the projective norm $\|Z[v_1, v_2]\|_\pi$. In order to do so, we construct another witness of the form $H[\lambda_1, \lambda_2]$, with $\lambda_1, \lambda_2 \in V^*$. Since $\|\lambda_i\|_* \leq 1$, we obtain as in (2.97) that $H[\lambda_1, \lambda_2]_{*\varepsilon} = H[\lambda_1, \lambda_2]_{\pi^*} \leq 2$.

The trick now is to focus on the values of k corresponding to the ‘worst case scenario’. As we saw, when $k \approx 1$ the right-hand side of (2.101) approaches 2, making (2.102) useless for the purpose of establishing a gap between projective and injective norm. However, we can exploit the very definition of k to try to say something about this case, too. Namely, we can invoke the existence of a new functional λ_1 that saturates the maximisation in (2.100). Such functional will be obtained by ‘tilting’ the pair (v_1^*, v_2^*) .

More precisely, one can find $\lambda_1 \in V^*$ with $\|\lambda_1\|_* \leq 1$ and such that

$$P\lambda_1 = \alpha_1 v_1^* + \alpha_2 v_2^*, \tag{2.103}$$

with $0 \leq \alpha_1 \leq \alpha_2 \leq 1$ and $\alpha_1 + \alpha_2 = 1 + k$. To construct λ_2 we simply appeal to the triangular inequality. Namely, for $|s_1| + |s_2| \leq 1$ we notice that $\lambda_2 := s_1 v_1^* + s_2 v_2^*$ obeys $\|\lambda_2\|_* \leq 1$. A quick computation yields

$$\begin{aligned} &\langle H[\lambda_1, \lambda_2], Z[v_1, v_2] \rangle \\ &= \langle H[\alpha_1 v_1^* + \alpha_2 v_2^*, s_1 v_1^* + s_2 v_2^*], Z[v_1, v_2] \rangle \\ &= \alpha_1^2 + 2\alpha_1 \alpha_2 - \alpha_2^2 + 2(\alpha_1 + \alpha_2)s_1 + 2(\alpha_1 - \alpha_2)s_2 - s_1^2 - 2s_1 s_2 + s_2^2. \end{aligned}$$

We are free to maximise over s_1, s_2 subjected to the constraint $|s_1| + |s_2| \leq 1$. Using Lemma 2.43 below, we obtain

$$\begin{aligned} \|Z[v_1, v_2]\|_\pi &\geq \max_{|s_1| + |s_2| \leq 1} \frac{1}{2} \langle H[\lambda_1, \lambda_2], Z[v_1, v_2] \rangle \\ &= \frac{3}{2} \alpha_1^2 + \alpha_1(\alpha_2 - 1) - \frac{1}{2}(\alpha_2 - 1)^2 + 1. \end{aligned}$$

Now, we do not know what the actual values of α_1, α_2 are, but we know that they are subjected to the constraints $0 \leq \alpha_1 \leq \alpha_2 \leq 1$ and $\alpha_1 + \alpha_2 = 1 + k \geq 1$.

Therefore, we can estimate the above lower bound by taking the minimum over all such pairs α_1, α_2 . This can be done thanks to Lemma 2.44 below, and the result is

$$\|Z[v_1, v_2]\|_\pi \geq 1 + \frac{3}{2}k^2. \quad (2.104)$$

Putting together (2.102) and (2.104) yields

$$\|Z[v_1, v_2]\|_\pi \geq \max \left\{ 2, 1 + \frac{3}{2}k^2 \right\}. \quad (2.105)$$

Finally, we see that

$$\frac{\|Z[v_1, v_2]\|_\pi}{\|Z[v_1, v_2]\|_\varepsilon} \geq \frac{\max \{2, 1 + \frac{3}{2}k^2\}}{2 - (1 - k)^2} \geq 1 + \frac{1}{29 + 12\sqrt{6}}, \quad (2.106)$$

where for the last step we used Lemma 2.45 below. \square

Remark. A by-product of the proof is that it is always possible to achieve $\|Z\|_\pi / \|Z\|_\varepsilon \geq c$ with a tensor Z whose tensor rank is 2.

Remark. The constant we found is most likely prone to substantial improvement. We know we must have $c \leq 2$, though, because as we saw in Example 2.2 the two-dimensional Euclidean norm satisfies $\|Z\|_\pi = \|Z\|_1 \leq 2\|Z\|_\infty = 2\|Z\|_\varepsilon$, where we thought of Z as a 2×2 real matrices, $\|\cdot\|_1, \|\cdot\|_\infty$ denote the sum of the singular values and the largest singular value, respectively, and thus 2 is the tightest possible constant in the above inequality.

By examining tensor products of simple two-dimensional Banach spaces like l_2^2 (whose unit ball is the unit circle) or l_1^2 (whose unit ball is the unit square), one might be tempted to conjecture that the best constant c satisfying $\|Z\|_\pi \leq c\|Z\|_\varepsilon$ can be taken as large as 2. However, this turns out not to be the case, as there are examples of ‘highly local’ Banach spaces such that the largest c satisfying Theorem 2.40 is $\sqrt{3}$, as we show in Appendix D.

Corollary 2.41. *For all centrally symmetric cones C defined as in (2.38), one has*

$$C \underset{\min}{\otimes} C \neq C \underset{\max}{\otimes} C. \quad (2.107)$$

Let us repeat here that we think of the above Corollary 2.41 as some strong evidence in favour of Conjecture 2.1. Indeed, to the extent of our knowledge, it constitutes the first attempt to prove such conjecture for models of arbitrarily high dimension (even if subjected to a further symmetry constraint). The other major partial result, Namioka-Phelps theorem (Theorem 2.37), in fact, covers only the case of one of the two local theories being a particular, low-dimensional GPT. Furthermore, the proof technique is relatively original and might give some hints on how to prove Conjecture 2.1 in general. Last, but not least, Theorem 2.40 may be of independent interest in functional analysis, since it deals with universal behaviours of Banach spaces.

Below we state and prove the four lemmas we use in the proof of Theorem 2.40.

Lemma 2.42. *Let $\alpha_1, \alpha_2, \beta_1, \beta_2 \in [-1, 1]$ be real numbers. Let $k \in [-1, 1]$ be such that*

$$1 + k \geq \max\{|\alpha_1| + |\alpha_2|, |\beta_1| + |\beta_2|\}. \quad (2.108)$$

For $t \in [-1, 1]$, consider the monotonically increasing function

$$f(t) := \begin{cases} (1+t)^2 & -1 \leq t \leq 0 \\ 2 - (1-t)^2 & 0 \leq t \leq 1 \end{cases} \quad (2.109)$$

Then

$$\alpha_1(\beta_1 + \beta_2) + \alpha_2(\beta_1 - \beta_2) \leq f(k). \quad (2.110)$$

Proof. We have to prove that the function $g : [0, 1] \rightarrow \mathbb{R}$ defined as

$$g(k) := \max \left\{ \sum_{i,j=1,2} \tau_{ij} \alpha_i \beta_j : |\alpha_i|, |\beta_j| \leq 1, \sum_{i=1,2} |\alpha_i|, \sum_{j=1,2} |\beta_j| \leq 1 + k \right\}$$

coincides with that given in (2.109). Let us start by assuming $0 < k \leq 1$. The maximisation over α_1, α_2 can be carried out with the help of the general formula

$$\begin{aligned} & \max \{ \alpha_1 r_1 + \alpha_2 r_2 : |\alpha_1|, |\alpha_2| \leq 1, |\alpha_1| + |\alpha_2| \leq 1 + k \} \\ &= \max \{ k|r_1| + |r_2|, |r_1| + k|r_2| \} \\ &= k(|r_1| + |r_2|) + (1 - k) \max\{|r_1|, |r_2|\}, \end{aligned}$$

which is an easy consequence of the linearity of the function inside the maximum on the left-hand side and is valid provided that $0 \leq k \leq 1$. We are left with

$$\begin{aligned} g(k) &= \max \{ k(|\beta_1 + \beta_2| + |\beta_1 - \beta_2|) + (1 - k) \max\{|\beta_1 + \beta_2|, |\beta_1 - \beta_2|\} : \\ & \quad |\beta_1|, |\beta_2| \leq 1, |\beta_1| + |\beta_2| \leq 1 + k \} \\ &= \max \{ 2k \max\{|\beta_1|, |\beta_2|\} + (1 - k)(|\beta_1| + |\beta_2|) : \\ & \quad |\beta_1|, |\beta_2| \leq 1, |\beta_1| + |\beta_2| \leq 1 + k \} \\ &= 2k + (1 - k)(1 - k) \\ &= 2 - (1 - k)^2 \\ &= f(k). \end{aligned}$$

If $-1 \leq k \leq 0$ we can perform exactly the same steps, which become even easier since the geometric shapes involved in the maximisations are squares and not octagons. For instance, one uses

$$\max\{\alpha_1 r_1 + \alpha_2 r_2 : |\alpha_1|, |\alpha_2| \leq 1, |\alpha_1| + |\alpha_2| \leq 1 + k\} = (1 + k) \max\{|r_1|, |r_2|\},$$

to eliminate α_1, α_2 , so that

$$\begin{aligned} g(k) &= \max \left\{ (1+k) \max\{|\beta_1 + \beta_2|, |\beta_1 - \beta_2|\} : \right. \\ &\quad \left. |\beta_1|, |\beta_2| \leq 1, |\beta_1| + |\beta_2| \leq 1+k \right\} \\ &= (1+k)^2 \\ &= f(k), \end{aligned}$$

concluding the proof. \square

Lemma 2.43. *For $\alpha_2 \geq \alpha_1 \geq 0$ such that $\alpha_1 + \alpha_2 \geq 1$ we have*

$$\begin{aligned} &\max \left\{ 2(\alpha_1 + \alpha_2)s_1 + 2(\alpha_1 - \alpha_2)s_2 - s_1^2 - 2s_1s_2 + s_2^2 : |s_1| + |s_2| \leq 1 \right\} \\ &= 2\alpha_1^2 - 2\alpha_1 + 2\alpha_2 + 1. \end{aligned}$$

Proof. Call $w(s_1, s_2) := 2(\alpha_1 + \alpha_2)s_1 + 2(\alpha_1 - \alpha_2)s_2 - s_1^2 - 2s_1s_2 + s_2^2$. It is easily seen that $\partial_{s_1} w = 0 = \partial_{s_2} w$ happens only for $s_1 = \alpha_1, s_2 = \alpha_2$. However, the constraint $\alpha_1 + \alpha_2 \geq 1$ imposes that this point is not inside the tilted square $|s_1| + |s_2| \leq 1$. Therefore, in order to find the maximum we have to look at the values of w on the border.

Rewriting $-s_1^2 - 2s_1s_2 + s_2^2 = (s_1 - s_2)^2 - 2s_1^2 = -(s_1 + s_2)^2 + 2s_2^2$ we see that w is convex on the two sides of the square where $s_1 + s_2 = \pm 1$ and concave on the other two where $s_1 - s_2 = \pm 1$. Therefore, these latter two sides (together with their vertices) are all that matters.

Now, on the side identified by $s_2 = s_1 - 1$ ($0 \leq s_1 \leq 1$) we obtain

$$w(s_1, s_1 - 1) = 2(\alpha_1 + \alpha_2)s_1 + 2(\alpha_1 - \alpha_2)(s_1 - 1) + 1 - 2s_1^2,$$

which achieves his maximum for $s_1 = \alpha_1$:

$$\max_{0 \leq s_1 \leq 1} w(s_1, s_1 - 1) = 2\alpha_1^2 - 2\alpha_1 + 2\alpha_2 + 1. \quad (2.111)$$

On the other side $s_2 = s_1 + 1$ ($-1 \leq s_1 \leq 0$) we have

$$w(s_1, s_1 + 1) = 2(\alpha_1 + \alpha_2)s_1 + 2(\alpha_1 - \alpha_2)(s_1 + 1) + 1 - 2s_1^2,$$

which is maximum on the extreme points since the derivative is never zero for $s_1 < 0$. Thus,

$$\begin{aligned} \max_{-1 \leq s_1 \leq 0} w(s_1, s_1 + 1) &= \max\{w(-1, 0), w(0, 1)\} \\ &= \max\{-2(\alpha_1 + \alpha_2) - 1, 2(\alpha_1 - \alpha_2) + 1\} \\ &= 2(\alpha_1 - \alpha_2) + 1. \end{aligned} \quad (2.112)$$

Thanks to $\alpha_2 \geq \alpha_1$, the rightmost side of (2.112) is clearly upper bounded by (2.111), which becomes the absolute maximum we are looking for. \square

Lemma 2.44. *For all $0 \leq k \leq 1$, we have*

$$\begin{aligned} & \min \left\{ \frac{3}{2}\alpha_1^2 + \alpha_1(\alpha_2 - 1) - \frac{1}{2}(\alpha_2 - 1)^2 + 1 : \right. \\ & \quad \left. 0 \leq \alpha_1 \leq \alpha_2 \leq 1, \alpha_1 + \alpha_2 \geq 1 + k \right\} \\ & = 1 + \frac{3}{2}k^2. \end{aligned} \quad (2.113)$$

Proof. Call $\nu(\alpha_1, \alpha_2)$ the objective function and $\mu(k)$ the minimum on the left-hand side of (2.113). The region of the plane α_1, α_2 determined by the inequalities $0 \leq \alpha_1 \leq \alpha_2 \leq 1$ and $\alpha_1 + \alpha_2 \geq 1 + k$ is a triangle with vertices $(k, 1)$, $(1, 1)$, $(\frac{1+k}{2}, \frac{1+k}{2})$. For a fixed $\frac{1+k}{2} \leq \alpha_2 \leq 1$, the range of α_1 is easily seen to be $[1 + k - \alpha_2, \alpha_2]$. The function $\nu(\alpha_1, \alpha_2)$ is convex in α_1 for fixed α_2 , but the point $\alpha_1 = \frac{1-\alpha_2}{3}$ in which the derivative $\partial_{\alpha_1}\nu(\alpha_1, \alpha_2)$ is zero is never inside the range, because $\frac{1-\alpha_2}{3} + \alpha_2 = \frac{1+2\alpha_2}{3} \leq 1 \leq 1 + k$. Instead, for the allowed values of α_1 we always have $\partial_{\alpha_1}\nu(\alpha_1, \alpha_2) \geq 0$. Therefore, in order to find the minimum we have to look at the smallest possible α_1 , and we get

$$\begin{aligned} \mu(k) &= \min_{\frac{1+k}{2} \leq \alpha_2 \leq 1} \nu(1 + k - \alpha_2, \alpha_2) \\ &= \min_{\frac{1+k}{2} \leq \alpha_2 \leq 1} \left(1 - 2k(\alpha_2 - 1) + \frac{3}{2}k^2 \right) \\ &= 1 + \frac{3}{2}k^2, \end{aligned}$$

as claimed. \square

Lemma 2.45. *The following holds:*

$$\min_{0 \leq k \leq 1} \frac{\max\{2, 1 + \frac{3}{2}k^2\}}{2 - (1 - k)^2} = 1 + \frac{1}{29 + 12\sqrt{6}}. \quad (2.114)$$

Proof. For $0 \leq k \leq \sqrt{\frac{2}{3}}$ the function to be maximised is $2 / (2 - (1 - k)^2)$, which is decreasing for $0 \leq k \leq 1$. For $\sqrt{\frac{2}{3}} \leq k \leq 1$, we obtain instead $(1 + \frac{3}{2}k^2) / (2 - (1 - k)^2)$, whose derivative is positive outside $[-2, \frac{1}{3}]$. Therefore, the minimum is achieved for $k = \sqrt{\frac{2}{3}}$, and we can conclude. \square

2.7 Conclusions and outlook

Throughout this chapter, we reviewed the basic of the abstract state space formalism of GPTs, and we discussed the problem of the composition of local systems within this framework. We discussed some known examples of GPTs such as classical probability theory, quantum mechanics, or generalised non-signalling theories, and saw how the formalism of GPTs can be useful in rephrasing known concept such as that of Bell nonlocality in a more compact

way. We also introduced new classes of models like the centrally symmetric ones, which will turn out to be instrumental in the forthcoming Chapter 3.

Inspired by its importance (and beauty) in quantum information, we spent some time studying some features of the separability problem in GPTs. We extended some recent results on separability of low-rank tensor to the GPT setting, and uncovered a connection between the theory of tensor norms in functional analysis and the entanglement of centrally symmetric models.

Next, we formulated and motivated some conjectures about composite systems whose solution is in our view central to deepen our understanding of the connections between non-classicality of local theories and the possible existence of entangled states when they are composed to form a single bipartite system. We presented what is known about the problem and strengthened an old argument of Namioka and Phelps to include also a statement about nonlocality, which is a strictly stronger concept than entanglement. Moreover, we answered said conjecture in the affirmative in the case when both local subsystems are described by the same (but otherwise arbitrary) centrally symmetric model. Along the way, we proved a general result on Banach spaces that might be of independent interest in functional analysis.

As the careful reader will have noticed, we (intentionally) evaded the problem of giving substance to point (ii) of the discussion in Subsection 2.1.1. There, we were asking to provide a lower estimate of the entangleability of two local GPTs in terms of their ‘degree of non-classicality’. In order to do this, the following goals must be accomplished.

- (i) We have to find some appropriate measure of non-classicality of GPTs \mathcal{N} , which vanishes for classical theories and somehow quantifies how far from this extremal case we are. An example of such measure could be the global incompatibility quantifier T of (2.87), but many other choices are possible.
- (ii) We have to be able to quantify how much entanglement is contained in a state of $A \otimes_{\max} B$, for A, B GPTs. The maximal entanglement achievable in the maximal tensor product will identify some ‘entangleability measure’ $\mathcal{E}[A, B]$. Perhaps surprisingly, here there is a somewhat natural choice, for the only quantum entanglement measure that carries over to GPTs, since it rests only on general concepts such as convexity, is the entanglement robustness [125, 63].
- (iii) The third, decisive step is then combining the above two quantifiers and prove an inequality of the form

$$\mathcal{E}[A, B] \geq \mathcal{N}[A] \mathcal{N}[B], \quad (2.115)$$

to be obeyed by all GPTs A, B of any dimension.

An inequality like (2.115), if established, would be incredibly suggestive and inspiring. It would tell us that there is a deep, quantitative link between a

totally self-referential feature like non-classicality, which reveals itself when a GPT is considered on its own, and a global phenomenon such as the entanglement. Finally, as we saw in Subsection 2.1.1, one could alternatively consider an analogue of (2.115) with the concept of Bell nonlocality displacing that of entanglement.

As the abundance of conjectures and questions we stated in the chapter suggests, our analysis leaves many open problems, and is prone to substantial improvement and deepening. This is partly due to the fact that our interest in the problems presented here was spurred by the results of the forthcoming Chapter 3, which came chronologically earlier in my PhD, and is thus relatively new. Anyway, we hope that the reader has enjoyed our tour of the vast field of study of nonlocality in general probabilistic theories, and appreciated our sincere enthusiasm for the deep problems we were faced with.

Chapter 3

Ultimate data hiding in quantum mechanics and beyond

3.1 Introduction

While discussing the scientific programme of Chapter 2, in Subsection 2.1.1, we stressed the importance of identifying which ones among the many non-classical phenomena we discovered with the advent of the quantum revolution pertain to all non-classical theories, and which ones on the contrary have more to do with the particular nature of quantum mechanics. In point (iii) of Subsection 2.1.1, we pushed things a bit further, and suggested that this conceptually important question can be made *quantitative* instead of merely *qualitative*. Namely, for any particular non-classical phenomenon, we can ask how strong that specific type of non-classicality is in quantum mechanics compared to its ultimate limits within the broad realm of GPTs. One of the main purposes of the present chapter is to conduct this analysis thoroughly for the particular case of *data hiding*.

The present chapter is organised as follows. The rest of this section is devoted to introducing the main questions we will face and answer in broad terms (Subsection 3.1.1), and to pointing the reader at our original contributions (Subsection 3.1.2). Section 3.2 is devoted to giving precise definitions of what we mean by data hiding in the GPT setting. Then, throughout Section 3.3 we complete the solution of the important special case constituted by quantum mechanical data hiding. In Section 3.4 we present the main result of the chapter, i.e. the determination of the optimal universal upper bound on data hiding ratios against locally constrained sets of measurements (Theorem 3.21). Section 3.5 presents a body of techniques to compute data hiding ratios in specific classes of GPT model satisfying some further assumptions. Finally, Section 3.6 is devoted to the conclusions.

3.1.1 Data hiding

In quantum mechanics, data hiding is usually intended as the existence of pairs of states of a bipartite system that are perfectly distinguishable with global measurements yet almost indistinguishable when only protocols involving local operations and classical communication (LOCC) are allowed [9, 10]. In view of the discussion in Subsection 2.2.4, we can easily extend the relevant definitions to encompass more general form of data hiding in arbitrary general probabilistic theories (Definition 3.5). In this context, the effectiveness of discriminating protocols is measured by the minimal probability of error $P_e^{\mathcal{M}}(\rho, \sigma; p)$ in the task of distinguishing the two states ρ, σ with a priori probabilities $p, 1-p$ respectively, when only operations from the class \mathcal{M} are available. The archetypical example of a pair of states exhibiting data hiding is given by the normalised projectors onto the symmetric and antisymmetric subspaces in $\mathbb{C}^n \otimes \mathbb{C}^n$, denoted by ρ_S and ρ_A , respectively. While $P_e^{\text{ALL}}(\rho_S, \sigma_A, \frac{1}{2}) = 0$, because the two states have orthogonal support, it can be shown that $P_e^{\text{LOCC}}(\rho_S, \sigma_A, \frac{1}{2}) = \frac{2}{n+1}$ [9, 10].

For these discussions, a more convenient quantity is the *distinguishability norm* associated to \mathcal{M} [151], denoted by $\|\cdot\|_{\mathcal{M}}$ and defined by $P_e^{\mathcal{M}}(\rho, \sigma; p) = \frac{1}{2}(1 - \|p\rho - (1-p)\sigma\|_{\mathcal{M}})$, as detailed in Definition 3.2 and Lemma 3.3; it quantifies the advantage of making an observation over pure guessing (i.e. the prior information). It is immediately obvious that higher values of $\|p\rho - (1-p)\sigma\|_{\mathcal{M}}$ correspond to an increased discriminating power of the set \mathcal{M} . Thus, the central object of our investigation is a quantity that we name *data hiding ratio*, which depends on the GPT as well as on the restricted set of measurements we consider, and is given by

$$R(\mathcal{M}) = \max \frac{\|p\rho - (1-p)\sigma\|_{\text{ALL}}}{\|p\rho - (1-p)\sigma\|_{\mathcal{M}}},$$

where $\|\cdot\|_{\text{ALL}}$ denotes the norm associated to the whole set of possible measurements and the maximisation ranges over all pairs of states ρ, σ and a priori probabilities p (Definition 3.5 and Proposition 3.6).

We will look at the quantity $R(\mathcal{M})$ in the case of physically or operationally significant restricted sets of protocols \mathcal{M} . Concretely, we focus on the case when the system under examination is bipartite, and \mathcal{M} is some set of locally constrained measurements (Definition 2.8). Our interest in this kind of questions has been spurred by a result in [151], stating that on a finite-dimensional quantum mechanical system $\mathbb{C}^{n_A} \otimes \mathbb{C}^{n_B}$ the data hiding ratio against LOCC protocols (which constitute the most operationally relevant class) satisfies

$$\Omega(n) \leq R_{\text{QM}}(\text{LOCC}) \leq O(\sqrt{n_A n_B}), \quad (3.1)$$

where $n = \min\{n_A, n_B\}$. In particular, when $n_A = n_B$ one obtains

$$R_{\text{QM}}(\text{LOCC}) = \Theta(n).$$

In this chapter, we will conduct a thorough analysis of $R_{\text{QM}}(\text{LOCC})$ and of analogous quantities constructed in arbitrary GPTs. For details, see the following Subsection 3.1.2.

3.1.2 Our contributions

The material presented in this chapter is part of the homonymous paper [99]:

- L. Lami, C. Palazuelos, and A. Winter. Ultimate data hiding in quantum mechanics and beyond. *Preprint arXiv:1703.03392*, 2017.

Let us now point the reader to our main original contributions. From our review of what is already known about data hiding in quantum mechanics, it is immediately clear that there is a gap between lower and upper bound in (3.1), which can become arbitrarily large for very different values of the local dimensions. Then, we are interested in the determination of the exact scaling of $R_{\text{QM}}(\text{LOCC})$ with n_A and n_B . Our first contribution is an intuitive argument which uses the quantum teleportation protocol to show that $R_{\text{QM}}(\text{LOCC}) = \Theta(n)$ holds in fact for all n_A, n_B (Theorem 3.11). The beauty of the argument lies not only in its crystal clear simplicity, but also in the fact that: (i) it works for all n_A, n_B , thus providing a general solution to the problem of computing $R_{\text{QM}}(\text{LOCC})$; (ii) it is constructive, in the sense that it provides an explicit one-way LOCC protocol to achieve the lower bound on the probability of error; (iii) it yields better constants than the one in [151]; in fact, our bounds will be so close to being tight that they will enable us to compute *exactly* the ‘relaxed’ ratio $R_{\text{QM}}(\text{SEP})$ at least when $n_A = n_B$; (iv) it is potentially generalisable to GPTs other than quantum mechanics, as long as they support teleportation [41].

We will spend some time examining some of the consequences of Theorem 3.11 and of the teleportation argument in standard quantum information (Subsection 3.3.3). In particular, we will exhibit the best known lower bound on the squashed entanglement in terms of the trace norm (Corollary 3.13), an alternate quantum de Finetti theorem (Corollary 3.14), and a modified algorithm that solves the weak membership problem for separability in an amount of time that scales only polynomially in one of the two local dimensions when the other one (and the error) are fixed (Corollary 3.15).

Back to the data hiding problem, we already mentioned that Theorem 3.11 implies that $R_{\text{QM}}(\text{LOCC}) = \Theta(\min\{n_A, n_B\})$. Note that the local real dimensions of the quantum cones of states (in other words, the dimensions of the local GPTs in the sense of Definition 2.6) are $d_A = n_A^2$ and $d_B = n_B^2$. Using these as parameters, one could write the value of the quantum data hiding ratio as $R_{\text{QM}}(\text{LOCC}) = \Theta(\min\{\sqrt{d_A}, \sqrt{d_B}\})$.

Along the same lines of thought, we compute the data hiding ratios against locally constrained sets of measurements for another significant GPT called ‘spherical model’ (Subsection 3.4.1). As the name suggests, the state space of such a model is a Euclidean ball. We prove that for fixed local real dimensions, the data hiding ratio displayed by the spherical model is quadratically larger than the quantum mechanical one, i.e. $R_{\text{Sph}}(\mathcal{M}) = \Theta(\min\{d_A, d_B\})$.

Then, the problem we are faced with is to establish an *optimal, universal upper bound* for $R(\mathcal{M})$ that depends only on the local real dimensions d_A, d_B and *not* on the particular GPTs we choose. Here, \mathcal{M} is a locally constrained

set of measurements, as usual. The answer to this central question is the content of the main result of the present chapter, Theorem 3.21, which states that $R(\mathcal{M}) \leq \min\{d_A, d_B\}$ holds for all bipartite GPTs of local dimensions d_A, d_B . Since we have seen that such a scaling characterises spherical models, we deduce that $\min\{d_A, d_B\}$ is the optimal universal upper bound on the data hiding ratio against locally constrained sets of measurements. This answers our fundamental question on the ultimate effectiveness of data hiding when the local systems have bounded size.

The originality of the proof of Theorem 3.21 is two-fold. On the one hand, it rests crucially on a deep connection between the theory of locally constrained distinguishability norms and that of tensor norms, in the functional analysis sense discussed in Subsection 2.2.1. This connection is uncovered here for the first time, and we believe it may lead to further advances in the understanding of both sides of the problem. On the other hand, in order to arrive at the proof of the central claim we will have to solve a question concerning tensor norms constructed out of arbitrary Banach spaces of fixed dimension. This is done in Proposition 3.18, which constitutes – to the extent of our knowledge – an original mathematical contribution on its own.

The last section of the chapter is less essential to the comprehension of the foundational questions posed here, and is mainly motivated by genuine curiosity. In Section 3.5, we analyse the problem of computing the data hiding ratios for special classes of GPTs. More in detail, we look at centrally symmetric models (Definition 2.11), and reduce the task of computing the data hiding ratio to that of evaluating a certain quantity constructed out of some injective and projective tensor norms (Theorem 3.24). This approach constitutes a generalisation of the aforementioned solution of the spherical model, and represents our third instance (after those described in Theorems 2.40 and 3.21) of applications of the theory of tensor norms to quantum information theoretic problems. Finally, we show how to generalise the concept of data hiding Werner states to arbitrary GPTs that possess an adequate symmetry. It turns out that the maximal effectiveness of data hiding that is achievable using those high-symmetry states depends only on few geometric parameters rather than on the whole structure of the involved GPTs, as detailed in Theorem 3.33.

3.2 Data hiding in general probabilistic theories

Throughout this section, we are going to define the central object of our investigation, i.e. the phenomenon known as data hiding. Since we want to address the problem in full generality, our treatment encompasses the case of an arbitrary GPT. In the next Section 3.3, we will analyse in detail the important case of quantum mechanical data hiding.

3.2.1 State discrimination

In the following, let (V, C, u) be a given GPT in the sense of Definition 2.6. As customary, we will denote by Ω its state space. The binary distinguishability

problem consists of choosing secretly one of the two states $\rho, \sigma \in \Omega$ with known a priori probabilities $p, 1 - p$ and handing it over to an agent, whose task is to discriminate between the two alternatives. Naturally, the larger the set of measurements the agent has at his disposal, the lower the associated probability of error will be (for fixed states and a priori probabilities). In general, it will make sense to consider measurements that are at least *informationally complete*, meaning that from their complete statistics the full state can be reconstructed unambiguously. A formal definition is below. Following the notation developed in Subsection 2.2.2, we remind the reader that a measurement $(e_i)_{i \in I} \in \mathbf{M}$ is a finite collection of effects $e_i \in [0, u]$ such that $\sum_{i \in I} e_i = u$.

Definition 3.1. *Let (V, C, u) be a GPT. Then a measurement $(e_i)_{i \in I} \in \mathbf{M}$ is said to be informationally complete if $\text{span}\{e_i : i \in I\} = V^*$. A set $\{\mu_t\}_{t \in T}$ made of measurements $\mu_t = (e_i^{(t)})_{i \in I_t}$ is deemed informationally complete if $\text{span}\{e_i^{(t)} : t \in T, i \in I_t\} = V^*$.*

If $\mathcal{M} \subseteq \mathbf{M}$ is a set of measurements in an arbitrary GPT, we can define an associated norm by translating to GPTs the analogous definition in [151].

Definition 3.2. *Let $\mathcal{M} \subseteq \mathbf{M}$ be an informationally complete set of measurements in a GPT (V, C, u) . The associated distinguishability norm $\|\cdot\|_{\mathcal{M}}$ is a norm on V given by*

$$\|x\|_{\mathcal{M}} := \sup_{(e_i)_{i \in I} \in \mathcal{M}} \sum_i | \langle e_i, x \rangle | \quad (3.2)$$

for all $x \in V$.

As is easy to see, the function defined in (3.2) is truly a norm on V thanks to the informational completeness of \mathcal{M} . Among its elementary properties, we note three: (i) the identity $\|x\|_{\mathcal{M}} = \langle u, x \rangle$ valid on positive states $x \geq 0$ (independently of \mathcal{M}); (ii) the monotonicity of $\|\cdot\|_{\mathcal{M}}$ in \mathcal{M} , with the partial order defined by the inclusion; and (iii) the fact that $\|\cdot\|_{\mathcal{M}}$ coincides with the base norm $\|\cdot\|$. This latter fact can be seen as follows: on the one hand, from (2.10) it is clear that $\|\cdot\|_{\mathbf{M}} \geq \|\cdot\|$, because $(e, u - e)$ is a legitimate measurement for all $e \in [0, u]$; on the other hand, $\|\cdot\| \geq \|\cdot\|_{\mathbf{M}}$ because if $\mathcal{M} = \mathbf{M}$ then in (3.2) we can restrict ourselves to binary measurements with no loss of generality. In fact, for an arbitrary $(e_i)_{i \in I} \in \mathbf{M}$ one has

$$\begin{aligned} \sum_{i \in I} | \langle e_i, x \rangle | &= \sum_{i \in I_+} \langle e_i, x \rangle - \sum_{i \in I_-} \langle e_i, x \rangle \\ &= \left\langle \sum_{i \in I_+} e_i, x \right\rangle - \left\langle \sum_{i \in I_-} e_i, x \right\rangle \\ &\leq \left| \left\langle \sum_{i \in I_+} e_i, x \right\rangle \right| + \left| \left\langle \sum_{i \in I_-} e_i, x \right\rangle \right|, \end{aligned}$$

where we defined $I_+ := \{i \in I : \langle e_i, x \rangle \geq 0\}$ and $I_- := I \setminus I_+$.

The above Definition 3.2 makes sense by virtue of its link to the operational task of state discrimination as given by the following lemma, totally analogous to [151, Theorem 5].

Lemma 3.3. *Let (V, C, u) be a GPT with state space Ω , and consider a set of measurements $\mathcal{M} \subseteq \mathbf{M}$. Then the lowest probability of error for discriminating between two states $\rho, \sigma \in \Omega$ with a priori probabilities $p, 1 - p$, respectively, is given by*

$$P_e^{\mathcal{M}}(\rho, \sigma; p) = \frac{1}{2} (1 - \|p\rho - (1-p)\sigma\|_{\mathcal{M}}), \quad (3.3)$$

where $\|\cdot\|_{\mathcal{M}}$ is the distinguishability norm given by (3.2).

Proof. This goes in complete analogy with the corresponding argument for quantum mechanics [152, 151, 153], but we explain it here for the sake of completeness. Without loss of generality, we can assume that the protocol consists of measuring the state with a measurement $(e_i)_{i \in I}$ and performing a (possibly probabilistic) post-processing of the classical outcome i . To encompass the possibility of such probabilistic post-processing, assume that the outcome i yields ρ or σ as final answers with probabilities q_i and $1 - q_i$, respectively. Then the probability of error is given by

$$\begin{aligned} P_e &= p \sum_{i \in I} \langle e_i, \rho \rangle (1 - q_i) + (1 - p) \sum_i \langle e_i, \sigma \rangle q_i \\ &= p - \sum_i q_i \langle e_i, p\rho - (1 - p)\sigma \rangle, \end{aligned}$$

where we employed the normalisation relation $\sum_{i \in I} e_i = u$. Minimising over all probabilities q_i one obtains

$$P_e = \frac{1}{2} \left(1 - \sum_{i \in I} |\langle e_i, p\rho - (1 - p)\sigma \rangle| \right),$$

and finally (3.3) after a minimisation over all measurements $(e_i)_{i \in I} \in \mathcal{M}$. \square

The analogy with quantum mechanics goes much beyond this. In fact, all the results of [151, §2] (with the exception of Proposition 8 there) carry over to GPTs. In translating the statements one has just to remember that the quantum mechanical trace norm $\|\cdot\|_1$ becomes the base norm in the GPT framework, and that similarly the identity becomes the unit effect. For details on the interpretation of quantum mechanics as a GPT, we refer the reader to Subsection 2.3.2. Here we limit ourselves to providing a formulation of [151, Theorem 4] for arbitrary GPTs. Remember from Subsection 2.2.2 that for a set of measurements $\mathcal{M} \subseteq \mathbf{M}$ we will denote by $\langle \mathcal{M} \rangle$ the set generated by \mathcal{M} via the procedure of coarse graining (2.8).

Lemma 3.4. *The unit ball of the dual to the distinguishability norm (3.2) is given by*

$$B_{\|\cdot\|_{\mathcal{M},*}} = \text{cl} \left(\text{co} \{ 2e - u : \{e, u - e\} \in \langle \mathcal{M} \rangle \} \right). \quad (3.4)$$

Equivalently, $\|\cdot\|_{\mathcal{M}}$ can be computed as

$$\|x\|_{\mathcal{M}} = \sup \left\{ \langle \varphi, x \rangle : \left(\frac{u + \varphi}{2}, \frac{u - \varphi}{2} \right) \in \langle \mathcal{M} \rangle \right\}. \quad (3.5)$$

Consequently, there is a one-to-one correspondence between distinguishability norms (3.2) and closed symmetric convex bodies K such that $\pm u \in K \subseteq [-u, u]$.

Proof. See [151]. □

Remark. From Lemma 3.4 it follows in particular that $\|\cdot\|_{\mathcal{M}}$ depends only on the set of measurements generated by \mathcal{M} via coarse graining, and in fact only on the right-hand side of (3.4).

At this point, the reader should be familiar enough with the body of techniques we have discussed so far, to be able to work out the translations of the other results in [151, §2] by herself. As for us, we believe it more appropriate to devote the forthcoming Subsection 3.2.2 to discussing the phenomenon of data hiding, first in quantum mechanics (where it was originally discovered), and subsequently in an arbitrary GPT.

3.2.2 Data hiding and statement of the problem

Throughout this section, a generalisation of the concept of data hiding against LOCC measurements in quantum mechanics as originally conceived in [9, 10] is discussed. On the one hand, we will extend this notion to an arbitrary GPT, and on the other hand we will allow for data hiding against an arbitrary set of measurements, without any a priori assumption on its nature. We give the following definition.

Definition 3.5. *Let (V, C, u) be a GPT with state space Ω . For a set of measurements $\mathcal{M} \subseteq \mathbf{M}$, we say that there is data hiding against \mathcal{M} with efficiency $R \geq 1$ if there are two normalised states $\rho, \sigma \in \Omega$ and a real number $p \in [0, 1]$ such that the probability of error defined in (3.3) satisfies*

$$P_e^{\mathbf{M}}(\rho, \sigma; p) = 0, \quad P_e^{\mathcal{M}}(\rho, \sigma; p) = \frac{1}{2} \left(1 - \frac{1}{R} \right). \quad (3.6)$$

The highest data hiding efficiency against \mathcal{M} is called data hiding ratio against \mathcal{M} and will be denoted by $R(\mathcal{M})$.

It is clear from the above definition and from the comparison with (3.3) that if a set of measurements exhibits data hiding with high efficiency then the norm defined via (3.2) will be very different from the base norm. The converse of this statement is the content of the following result.

Proposition 3.6. *For an informationally complete set of measurements $\mathcal{M} \subseteq \mathbf{M}$ in an arbitrary GPT (V, C, u) , the data hiding ratio $R(\mathcal{M})$ is given by*

$$R(\mathcal{M}) = \max_{0 \neq x \in V} \frac{\|x\|}{\|x\|_{\mathcal{M}}}. \quad (3.7)$$

Adopting the terminology of [151], we can rephrase (3.7) by saying that $R(\mathcal{M})$ is the constant of domination of $\|\cdot\|_{\mathcal{M}}$ on $\|\cdot\|$, i.e. the smallest $k \in \mathbb{R}$ such that $\|\cdot\| \leq k\|\cdot\|_{\mathcal{M}}$.

Proof. Let the GPT (V, C, u) have state space Ω . Then, from Definition 3.5 and from (3.3) we see that

$$R(\mathcal{M}) = \max \left\{ \|p\rho - (1-p)\sigma\|_{\mathcal{M}}^{-1} : \rho, \sigma \in \Omega, 0 \leq p \leq 1, \|p\rho - (1-p)\sigma\| = 1 \right\}.$$

Now, the crucial observation is that the set K of vectors $x \in V$ that can be represented as $x = p\rho - (1-p)\sigma$ for appropriate $\rho, \sigma \in \Omega$ and $p \in [0, 1]$ coincides with the unit ball of the base norm, i.e. $K = B_{\|\cdot\|}$. This is an easy consequence of (2.6) and of (2.5), something that we exploited already in writing down (1.21). Thanks to this observation, we rewrite the above representation of $R(\mathcal{M})$ as

$$\begin{aligned} R(\mathcal{M}) &= \max \{ \|x\|_{\mathcal{M}}^{-1} : \|x\| = 1 \} \\ &= \max_{x \neq 0} \frac{\|x\|}{\|x\|_{\mathcal{M}}}, \end{aligned}$$

as claimed. □

Some elementary properties of the data hiding ratio are as follows.

Lemma 3.7. *Let \mathcal{M} be an informationally complete set of measurements in an arbitrary GPT. Then:*

- *the data hiding ratio given in Definition 3.5 satisfies $R(\mathcal{M}) \geq 1$, with equality iff the set on the left hand side of (3.4) coincides with the full interval $[-u, u]$;*
- *$R(\mathcal{M})$ is monotonically non-increasing as a function of \mathcal{M} , where the partial order on sets of measurements is the one given by inclusion.*

As expected, not much can be said about data hiding ratios for a single system and when the sets of measurements are completely arbitrary. In fact, it is easy to show that already in a classical GPT (Subsection 2.3.1) such as $(\mathbb{R}^2, \mathbb{R}_+^2, u)$ with $u = (1, 1)$, so that $\langle u, (x, y) \rangle = x + y$, for the particular case when \mathcal{M} is made of just one measurement $((\varepsilon, 0), (1 - \varepsilon, 1))$ we have $R(\mathcal{M}) = (2 - \varepsilon)/\varepsilon$, so that the data hiding ratio can even be unbounded in a system of fixed dimension.

The situation changes dramatically when we consider bipartite systems (Subsection 2.2.3) and data hiding against locally constrained sets of measurements (Definition 2.8). This follows closely as originally done in the context of data hiding, which was initially defined against LOCC protocols in bipartite quantum systems [9, 10]. An important feature of locally constrained sets of measurements, which makes them suitable for the study of data hiding phenomena, is informational completeness. This is a consequence of the local tomography principle (Axiom 6), which we assumed to hold in general when we combine local GPTs to describe a bipartite system. We formulate this elementary observation as follows.

Lemma 3.8. *All the locally constrained sets of measurements (Definition 2.8) are informationally complete according to Definition 3.1.*

Proof. Because of (2.26), we just need to show that LO is informationally complete. In order to prove this elementary fact, consider a local measurement $(e_i \otimes f_j)_{(i,j) \in I \times J} \in \text{LO}$ such that $\text{span}\{e_i\}_{i \in I} = V_A^*$ and $\text{span}\{f_j\}_{j \in J} = V_B^*$. Then, obviously, $\text{span}\{e_i \otimes f_j\}_{(i,j) \in I \times J} = V_A^* \otimes V_B^* = V_{AB}^*$, which yields the claim. Here, we made use of the dual of the identity (2.12), which rests crucially on Axiom 6. \square

Thanks to (2.26) and Lemma 3.7, we also find

$$R(\text{SEP}) \leq R(\text{LOCC}) \leq R(\text{LOCC}_{\rightarrow}) \leq R(\text{LOSR}) \leq R(\text{LO}) \quad (3.8)$$

for all fixed GPTs. Our primary interest from now on is in data hiding against such operationally restricted sets of measurements. In particular, we want to understand how the best data hiding ratio *scales* with the dimensions of the local GPTs. To be more precise, we give the following definition.

Definition 3.9 (Ultimate data hiding ratio). *For a locally constrained set of measurements \mathcal{M} , the ultimate data hiding ratio against \mathcal{M} for fixed local dimensions, denoted by $R_{\mathcal{M}}(d_A, d_B)$, is the supremum over all data hiding ratios $R(\mathcal{M})$ achieved by composite GPTs that satisfy (2.18) and have local dimensions d_A, d_B .*

With this concept at hand, we are ready to formulate the question lying at the heart of our investigation in the present chapter, namely, *what is the scaling of the ultimate ratio $R_{\mathcal{M}}(d_A, d_B)$ with the local dimensions d_A, d_B ?* Clearly, thanks to the chain of inequalities (3.8), we find immediately

$$R_{\text{SEP}} \leq R_{\text{LOCC}} \leq R_{\text{LOCC}_{\rightarrow}} \leq R_{\text{LOSR}} \leq R_{\text{LO}} \quad (3.9)$$

for all positive integers d_A, d_B (the dependence on which of all the above ratios has been omitted for convenience).

We stress that the supremum in Definition 3.9 has to be taken over *all* local GPTs of the given dimensions, and over all *composition rules* adopted to join the system (i.e. among all the global cones respecting the bounds (2.18)). Now,

we will show that at least this latter maximisation can be carried out explicitly when \mathcal{M} is a locally constrained set of measurements different from LOCC, the optimal composite being always given by the *minimal* tensor product (2.14). To see why, notice that the exclusion of LOCC implies that for a fixed $Z \in V_A \otimes V_B$ only the global base norm $\|Z\|$ depends on the composition rule we chose. On the contrary, the locally constrained norm $\|\cdot\|_{\mathcal{M}}$ ($\mathcal{M} \neq \text{LOCC}$) will depend on the local structure only. Then, maximising the ratio between the former and the latter amounts to maximise the global base norm. In order to do so, a large set of global effects and thus a small set of states are required, and according to (2.18) the smallest possible positive cone in a bipartite system is given by the minimal tensor product.

The above reasoning is perhaps not obvious from Definition 3.5 alone, because restricting the set of available bipartite states gives less freedom in choosing the data hiding pair. However, this restriction plays no role once Proposition 3.6 is available. This way around the problem is made possible by the fact that any difference of two normalised states can be thought of as a positive multiple of the difference of two separable states, the multiplication coefficient being given by the base norm induced by the minimal tensor product. We summarise this whole discussion stating the following.

Proposition 3.10. *Given two local GPTs $A = (V_A, C_A, u_A)$, $B = (V_B, C_B, u_B)$, and a locally constrained set of measurements $\mathcal{M} \neq \text{LOCC}$, the maximal data hiding ratio against \mathcal{M} is achieved when the cone of bipartite states is given by the minimal tensor product, i.e. $C_{AB} = C_A \otimes_{\min} C_B$.*

Remark. In the opposite case, i.e. when $C_{AB} = C_A \otimes_{\max} C_B$, the base norm coincides with the separability norm. This is ensured by the fact that every allowed effect within this theory is automatically separable.

Before proceeding further, let us reassure those readers that might be worried about the exclusion of the LOCC protocols from Proposition 3.10. After all, there are operationally motivated reasons to consider LOCC as the most important among the locally constrained set of measurements. However, this exclusion will turn out to have no effect on our final result, Theorem 3.21. The reason why this is the case is that, for how difficult to characterise the LOCC set can be, we know that it obeys the two-fold inclusion bound (2.26), which translates to (3.9) in our case. Ultimately, this will allow us to capture some features of the behaviour of LOCCs without having to deal with them separately.

3.3 Data hiding in quantum mechanics

Throughout this section, we will apply the mathematical machinery we developed through the above Subsection 3.2.2 to the special case of finite-dimensional quantum mechanics, seen as a GPT as in Subsection 2.3.2. The reason for doing this lies primarily in the outstanding physical importance of this example

of GPT, which – as far as we understand – represents Nature as it happens to be. However, from a purely mathematical point of view, one can see this also as an attempt to obtain a lower bound to the ultimate data hiding ratios by investigating a specific example.

3.3.1 Quantum data hiding ratios

The purpose of this subsection is to determine the scaling of the data hiding ratios against locally constrained sets of measurements in the case of a composite quantum system $\text{QM}_{n_A n_B} = \text{QM}_{n_A} \otimes \text{QM}_{n_B}$, as defined in Subsection 2.3.2. In particular, we will be interested in the quantity $R_{QM}(\text{LOCC})$ as a function of n_A and n_B . Here, ‘determining the scaling’ means ‘computing up to universal constants that do not depend on n_A, n_B ’.

Let us start by recalling in brief how quantum mechanics fits the scheme of GPTs. For details, we refer the reader to Subsection 2.3.2. The cone of states of an n -level quantum mechanical system, called PSD_n , comprises all positive semidefinite $n \times n$ matrices. The host vector space is that of real hermitian matrices (denoted by \mathcal{H}_n), and its dimension is $d = n^2$. Density matrices are the positive matrices with trace one, and therefore the unit effect is nothing but the trace. Therefore, we write $\text{QM}_n = (\mathcal{H}_n, \text{PSD}_n, \text{Tr})$ as in (2.31). The quantum mechanical base norm coincides with the trace norm $\|\cdot\|_1$.

The composition rule for bipartite systems is very simple. If $A = \text{QM}_{n_A}$ and $B = \text{QM}_{n_B}$, then $AB = \text{QM}_{n_A n_B}$. This corresponds to choosing a composite cone which is neither the minimal nor the maximal tensor product of the local cones (2.36). While this is clearly not optimal in the sense of data hiding because of Proposition 3.10, it deserves special attention because of its prime importance in physics. In the forthcoming Subsection 3.3.2, we will also show that modifying the tensor product rule does not lead to any substantial improvement of the data hiding ratio of the theory.

The computation of data hiding ratios in quantum mechanics has been the subject of many papers, whose main results we are now about to summarise. The original example of a data hiding pair involves the normalised projectors onto the symmetric and antisymmetric subspace in $\mathbb{C}^n \otimes \mathbb{C}^n$, denoted by ρ_S and ρ_A , respectively [9, 10]. While $\|\rho_S - \rho_A\|_1 = 2$ because the two states have orthogonal support, it can be shown that $\|\rho_S - \rho_A\|_{\text{LOCC}} = 2/(n+1)$ [9, 10]. The fact that the two states are mixed is crucial for this construction to work, as it can be shown that for pure states trace norm and LOCC norm always coincide [154, 155]. In general, from [156, 151] it is known that

$$\|\cdot\|_2 \leq \sqrt{153} \|\cdot\|_{\text{LO}}, \quad (3.10)$$

where $\|Z\|_2 := \sqrt{\text{Tr } Z^\dagger Z}$ is the Hilbert-Schmidt norm. The above relation, whose proof is surprisingly intricate, is useful and interesting per se, and we will come to that later. Once translated in terms of the trace norm thanks to $\|\cdot\|_1 \leq \sqrt{n_A n_B} \|\cdot\|_2$, it yields the upper bound $R_{QM}(\text{LO}) \leq \sqrt{153 n_A n_B}$. Combining this with the lower bound on $R_{QM}(\text{SEP})$ as described in [151], one

obtains

$$\frac{\min\{n_A, n_B\} + 1}{2} \leq R_{QM}(\text{SEP}) \leq R_{QM}(\text{LOCC}) \leq R_{QM}(\text{LO}) \leq \sqrt{153 n_A n_B}, \quad (3.11)$$

while for $R_{QM}(\text{SEP})$ the tighter bound

$$\frac{\min\{n_A, n_B\} + 1}{2} \leq R_{QM}(\text{SEP}) \leq \sqrt{n_A n_B} \quad (3.12)$$

is available. As shown in [151, Corollary 17], the above relations solve the problem of determining the optimal scaling in n of all the data hiding ratios against locally constrained measurements when the subsystems have equal dimensions $n_A = n_B = n$. On the contrary, a problem arises when n_A and n_B are very different and thus $\min\{n_A, n_B\} \ll \sqrt{n_A n_B}$. In this case, the leftmost and rightmost side of (3.11) are no longer of the same order of magnitude, and we have to design an alternative argument.

Here we provide a simple reasoning that shows that in fact $O(\min\{n_A, n_B\})$ is still an upper bound for $R_{QM}(\text{LOCC})$ (and hence for $R_{QM}(\text{SEP})$, too). Furthermore, our reasoning yields much better constants for both the leftmost and the rightmost side of (3.11) (where $R_{QM}(\text{LO})$ is excluded, though). In fact, these constants are so close to being optimal that we are even able to compute $R_{QM}(\text{SEP})$ *exactly* when $n_A = n_B = n$.

Theorem 3.11 (Teleportation argument). *For a bipartite quantum mechanical system with Hilbert space $\mathbb{C}^{n_A} \otimes \mathbb{C}^{n_B}$, define $n := \min\{n_A, n_B\}$. Then the distinguishability norm against $\text{LOCC}_{\rightarrow}$ protocols satisfies*

$$\|\cdot\|_1 \leq (2n - 1) \|\cdot\|_{\text{LOCC}_{\rightarrow}}. \quad (3.13)$$

In particular, the data hiding ratios are such that

$$n \leq R_{QM}(\text{SEP}) \leq R_{QM}(\text{LOCC}) \leq R_{QM}(\text{LOCC}_{\rightarrow}) \leq 2n - 1, \quad (3.14)$$

where the communication direction in $\text{LOCC}_{\rightarrow}$ is from the smaller to the larger subsystem. Moreover, if $n_A = n_B = n$ then $R_{QM}(\text{SEP}) = n$.

Proof. Let us start by proving the upper bound on $R_{QM}(\text{LOCC}_{\rightarrow})$ in (3.14). We can assume without loss of generality that $n = n_A \leq n_B$, and that the classical communication line goes from A to B . We start by reminding the reader that the maximally entangled state $|\Phi\rangle = \frac{1}{\sqrt{n}} \sum_{i=1}^n |ii\rangle \in \mathbb{C}^n \otimes \mathbb{C}^n$, whose corresponding rank-one projector we denote by Φ , has the property that there is a separable state σ such that $\frac{1}{n}\Phi + \frac{n-1}{n}\sigma$ is again separable (in the language of [125], Φ has entanglement robustness $r(\Phi) = n - 1$). For instance, using classic results on the entanglement of isotropic states [118], it is not difficult to see that $\sigma = \frac{\mathbb{1} - \Phi}{n^2 - 1}$ satisfies all the requirements. Now, since we can always produce any separable state with $\text{LOCC}_{\rightarrow}$ operations (for that matter, even LOSR operations suffice), we are free to evaluate the $\text{LOCC}_{\rightarrow}$

norm on $X_{AB} \otimes \left(\frac{1}{n}\Phi + \frac{n-1}{n}\sigma\right)_{A'B'}$, instead of X_{AB} . Here, the systems A', B' have dimension $n_{A'} = n_{B'} = n_A = n$, and the operations are $\text{LOCC}_{\rightarrow}$ with respect to the splitting $AA'|BB'$.

Now, we are ready to apply the quantum teleportation protocol from A to B [157]. This is an $\text{LOCC}_{\rightarrow}$ operation τ mapping states of the system $AA'BB'$ to states of $B'B$, which can be defined as follows. For $p, q = 0, \dots, n-1$, introduce the unitary matrices

$$\begin{aligned} \mathbf{X}(p) &:= \sum_{k=1}^n |k \oplus p\rangle\langle k|, \\ \mathbf{Z}(q) &:= \sum_{k=1}^n e^{2q\pi i/n} |k\rangle\langle k|, \\ U(p, q) &:= \mathbf{X}(p)\mathbf{Z}(q). \end{aligned} \tag{3.15}$$

Then the teleportation τ is given by

$$\begin{aligned} &\tau(Z_{AA'BB'}) \\ &:= \sum_{p, q=0}^n U(p, q)_{B'} \text{Tr}_{AA'} [Z_{AA'BB'} U(p, q)_A \Phi_{AA'} U(p, q)_A^\dagger] U(p, q)_{B'}^\dagger. \end{aligned} \tag{3.16}$$

Most notably, observe that $\tau(Z_{AB} \otimes \Phi_{A'B'}) = Z_{B'B}$ (meaning that the same operator Z is written in the registers $B' \simeq A$ and B instead of A and B). Now, on the one hand, after the protocol has been performed, the local constraint plays no role any more, and any desired measurement can be applied to $B'B$, showing that $\|Z_{B'B}\|_{\text{LOCC}_{\rightarrow}} = \|Z_{AB}\|_1$. On the other hand, $\tau(Z_{AB} \otimes \sigma_{A'B'})$ is obtained from Z_{AB} via an $\text{LOCC}_{\rightarrow}$ protocol, hence $\|\tau(Z_{AB} \otimes \sigma_{A'B'})\|_{\text{LOCC}_{\rightarrow}} \leq \|Z_{AB}\|_{\text{LOCC}_{\rightarrow}}$. Putting all together, we obtain the following chain of inequalities:

$$\|Z_{AB}\|_{\text{LOCC}_{\rightarrow}} = \left\| Z_{AB} \otimes \left(\frac{1}{n}\Phi + \frac{n-1}{n}\sigma\right)_{A'B'} \right\|_{\text{LOCC}_{\rightarrow}} \tag{3.17}$$

$$\geq \left\| \tau \left(Z_{AB} \otimes \left(\frac{1}{n}\Phi + \frac{n-1}{n}\sigma\right)_{A'B'} \right) \right\|_{\text{LOCC}_{\rightarrow}} \tag{3.18}$$

$$= \left\| \frac{1}{n} Z_{B'B} + \frac{n-1}{n} \tau(Z_{AB} \otimes \sigma_{A'B'}) \right\|_{\text{LOCC}_{\rightarrow}} \tag{3.19}$$

$$\geq \frac{1}{n} \|Z_{B'B}\|_{\text{LOCC}_{\rightarrow}} - \frac{n-1}{n} \|\tau(Z_{AB} \otimes \sigma_{A'B'})\|_{\text{LOCC}_{\rightarrow}} \tag{3.20}$$

$$\geq \frac{1}{n} \|Z_{AB}\|_1 - \frac{n-1}{n} \|Z_{AB}\|_{\text{LOCC}_{\rightarrow}}. \tag{3.21}$$

We conclude that

$$\|Z_{AB}\|_{\text{LOCC}_{\rightarrow}} \geq \frac{1}{2n-1} \|Z_{AB}\|_1,$$

enforcing $R_{QM}(\text{LOCC}_{\rightarrow}) \leq 2n - 1$ in view of Proposition 3.6.

In order to deduce the lower bound $R_{QM}(\text{SEP})$, we appeal to Werner states [91, 130]. These can be thought of as convex combinations of the normalised projectors onto the symmetric and antisymmetric subspace of $\mathbb{C}^n \otimes \mathbb{C}^n$, denoted by ρ_S and ρ_A , respectively. In terms of the ‘flip operator’ F defined by $F|\alpha\beta\rangle = |\beta\alpha\rangle$ for all $|\alpha\rangle, |\beta\rangle \in \mathbb{C}^n$, we have

$$\rho_S = \frac{\mathbb{1} + F}{n(n+1)}, \quad \rho_A = \frac{\mathbb{1} - F}{n(n-1)}. \quad (3.22)$$

Since $n = \min\{n_A, n_B\}$, we can safely imagine to give one share of this bipartite system to A and the other to B . We already saw how the preparation with equal a priori probabilities of the two extremal states is well-known to produce data hiding, as shown by the fact that $\|\rho_S - \rho_A\|_1 = 2$ but $\|\rho_S - \rho_A\|_{\text{SEP}} = \|\rho_S - \rho_A\|_{\text{LOCC}} = 2/(n+1)$ [151, 156]. Curiously, there is an optimised version of this construction with different weights that does not seem to have been considered before. Namely, via the same techniques it can be shown that

$$\begin{aligned} \left\| \frac{n+1}{n} \rho_S - \frac{n-1}{n} \rho_A \right\|_1 &= 2, \\ \left\| \frac{n+1}{n} \rho_S - \frac{n-1}{n} \rho_A \right\|_{\text{SEP}} &= \frac{2}{n}. \end{aligned} \quad (3.23)$$

Observe that $\frac{n+1}{n} \rho_S - \frac{n-1}{n} \rho_A = \frac{2}{n^2} F$. Since the proof of (3.23) is just a variation of a standard calculation, we relegate it to Appendix E. Thanks to Proposition 3.6, this yields the lower bound in the claim. Finally, combining $R_{QM}(\text{SEP}) \geq n$ with the upper bound in (3.12), we see that when $n_A = n_B = n$ we must have $R_{QM}(\text{SEP}) = n$. \square

Remark. The fact that the upper bound for $R_{QM}(\text{LOCC})$ in Theorem 3.11 only depends on $\min\{n_A, n_B\}$ is crucial in solving the data hiding problem in quantum mechanics (up to constants) for all pairs (n_A, n_B) . To our knowledge, this complete solution was not known before. We find the simplicity of the above proof quite instructive on its own, but it does not seem like the teleportation argument can encompass the case of purely local measurements, by its very nature. Therefore, we must leave open the problem of finding the optimal scaling of $R_{QM}(\text{LO})$ in the general case.

Remark. We do not want to give the reader the wrong impression that the inequality (3.10) found in [151] is now useless, since what you deduce from it concerning data hiding ratios, i.e. (3.11), is a looser bound than our (3.14). On the contrary, (3.10) has to be thought as giving some kind of complementary information. First, it tells us that in order to achieve high data hiding efficiency we need to engineer high-rank states, something not apparent from (3.13). Second, it involves the norm $\|\cdot\|_{\text{LO}}$ instead of the larger $\|\cdot\|_{\text{LOCC}_{\rightarrow}}$ that appears in (3.13). Finally, (3.10) can be generalised to the multipartite case. While the same is true for our (3.13), in this more general case these two

approaches can be shown to lead to incomparable results concerning the ratio $\|\cdot\|_1/\|\cdot\|_{\text{LOCC}}$.

Remark. In terms of the real dimensions of the local spaces, which are given by $d_A = n_A^2$, $d_B = n_B^2$ according to (2.32), Theorem 3.11 shows that the data hiding ratio against separable protocols scales as $\min\{\sqrt{d_A}, \sqrt{d_B}\}$. Thus, we deduce a first estimate $R_{\text{SEP}}(d_A, d_B) \geq \min\{\sqrt{d_A}, \sqrt{d_B}\}$ (valid when $\sqrt{d_A}, \sqrt{d_B}$ are integers).

3.3.2 Data hiding in W -theory

In view of Proposition 3.10, the reader might wonder, whether considering a modified version of quantum mechanics in which composite systems are obtained via the minimal tensor product (2.34) exhibits better data hiding properties. We already encountered such a modified version of standard quantum mechanics, and dubbed it W -theory (Definition 2.10). Remember that the base norm of such theory takes the form (2.37). Perhaps surprisingly, it turns out that the answer to the aforementioned question is substantially negative, meaning that the improvement in the highest data hiding efficiency is bounded by a constant.

In order to arrive at this conclusion, we need to investigate a bit more the concept of LOCC protocols in W -theory. Here, by protocol we mean a state transformation that is achieved by combining elementary operations, possibly subjected to operational constraints, and possibly probabilistic. In the language of Subsection 2.2.4, we could also call these ‘instruments’. As remarked there, this has not to be confused with the concept of measurement. The outcome of a measurement is a classical label, while the outcome of a protocol or instrument is a state (possibly together with a classical label).

Let us then consider the class of protocols that can be realised in the W -theory framework when two agents A and B are allowed to use just local operations and $A \rightarrow B$ classical communication. We denote this class of protocols by $\mathbf{LWCC}_{\rightarrow}$. Despite the fact that $\mathbf{LWCC}_{\rightarrow}$ constitutes a more general class of *protocols* than $\text{LOCC}_{\rightarrow}$ in standard quantum mechanics, it is not difficult to see that they are not more powerful than the latter within the context of state discrimination. This is a consequence of the fact that the *measurements* one can implement with $\mathbf{LWCC}_{\rightarrow}$ operations are necessarily of the form $(E_i \otimes F_j^{(i)})_{(i,j) \in I \times J}$ for some local measurements $(E_i)_{i \in I}$ on A and $(F_j^{(i)})_{j \in J}$ on B . Since local measurements in W -theory are the same as in standard quantum mechanics, the same measurement is also obtainable via $\text{LOCC}_{\rightarrow}$ operations. More generally, we saw already that the locally constrained sets of measurements defined in (2.21), (2.22), (2.23), (2.25) do not depend on the composition rule we choose for assembling multipartite systems.

The above discussion allows us to write the identity

$$\|\cdot\|_{\text{LOCC}_{\rightarrow}} = \|\cdot\|_{\mathbf{LWCC}_{\rightarrow}}. \quad (3.24)$$

Here, the right-hand side can be thought of as defined via the minimal error probability according to (3.3), where said error probability has to be achieved by A and B running a $\mathbf{LWCC}_{\rightarrow}$ protocol followed by a classical decision. Of course, one can also look at it as defined by the usual formula (3.2), where it is understood that the corresponding set \mathcal{M} includes in this case all those measurements that are implementable through an $\mathbf{LWCC}_{\rightarrow}$ protocol.

As it turns out, data hiding in W -theory is not much better than in quantum theory, in the sense that the scaling with the local dimensions is exactly the same. The proof of this latter result constitutes another example of how the techniques used in [156, 151] seem not to be applicable in a more general scenario. In fact, the approach taken there, based on (3.10), would start with the inequality $\|\cdot\|_{\text{LOCC}_{\rightarrow}} \geq \kappa \|\cdot\|_2$, for κ universal constant and $\|\cdot\|_2$ the Hilbert-Schmidt norm. As we saw, when the system under examination is $\mathbb{C}^n \otimes \mathbb{C}^n$ the elementary relation $\|\cdot\|_2 \leq \frac{1}{n} \|\cdot\|_1$ yields $\|\cdot\|_{\text{LOCC}_{\rightarrow}} \geq \frac{\kappa}{n} \|\cdot\|_1$, which is optimal up to a constant factor. However, it is not difficult to prove that the ratio between the norm $\|\cdot\|_W$ defined in (2.37) and the Hilbert-Schmidt norm $\|\cdot\|_2$ can be asymptotically as large as $n^{3/2}$ (see Appendix F). Therefore, the tighter inequality we can deduce by making use of the Hilbert-Schmidt norm in an intermediate step is $\|\cdot\|_{\text{LOCC}_{\rightarrow}} \geq \kappa \|\cdot\|_2 \geq \frac{\kappa}{n^{3/2}} \|\cdot\|_W$. As we will see in a moment, the scaling of the lower bound is not tight.

While a direct approach via the other techniques previously exploited in the literature does not lead to a satisfactory answer to the problem, the teleportation argument can be quickly adapted to compute *exactly* the data hiding ratios against SEP or $\text{LOCC}_{\rightarrow}$ in W -theory.

Proposition 3.12. *For a bipartite W -theory with local Hilbert spaces of dimensions n_A, n_B , define $n := \min\{n_A, n_B\}$. Then the data hiding ratios against separable and $\text{LOCC}_{\rightarrow}$ measurements are given by*

$$R_W(\text{SEP}) = R_W(\text{LOCC}_{\rightarrow}) = 2n - 1, \quad (3.25)$$

where the communication direction in $\text{LOCC}_{\rightarrow}$ is from the smaller to the larger subsystem.

Proof. We assume without loss of generality that $n_B \geq n_A = n$. Let us start by showing that the above data hiding ratios can be upper bounded by $2n - 1$. The idea is that the argument in (3.17)-(3.21) can be adapted to encompass also the case of W -theory, by replacing everywhere $\|\cdot\|_{\text{LOCC}_{\rightarrow}}$ with $\|\cdot\|_{\mathbf{LWCC}_{\rightarrow}}$ and the trace norm $\|\cdot\|_1$ with the minimal tensor product base norm (2.37). The first step consists in acknowledging the fact that we can compute the $\text{LOCC}_{\rightarrow}$ distinguishability norm by making use of more general $\mathbf{LWCC}_{\rightarrow}$ protocols that are available in W -theory, as expressed in (3.24).

Now, we choose a particular $\mathbf{LWCC}_{\rightarrow}$ protocol in order to lower bound the norm $\|Z_{AB}\|_{\mathbf{LWCC}_{\rightarrow}}$. Such a protocol resembles the one we devised for the proof of Theorem 3.11, with one important difference.

Since separable states can be created with local operations and shared randomness, we can safely start by supplying A and B with a separable isotropic

state $\left(\frac{1}{n}\Phi + \frac{n-1}{n}\sigma\right)_{A'B'}$, defined on an ancillary system $A'B'$ with local dimension $n_{A'} = n_{B'} = n$. As usual, σ is an appropriate normalised, separable state. Then, we perform the teleportation τ defined in (3.16), which is an $\mathbf{LWCC}_{\rightarrow}$ (even $\mathbf{LOCC}_{\rightarrow}$) operation with respect to the splitting $AA'|BB'$, where classical communication goes from AA' to BB' . After applying the triangular inequality, we are left with two terms, i.e. $\|Z_{B'B}\|_{\mathbf{LWCC}_{\rightarrow}}$ and $\|\tau(Z_{AB} \otimes \sigma_{A'B'})\|_{\mathbf{LWCC}_{\rightarrow}}$. The first one can be computed exactly, since the operator $Z_{B'B}$ obtained after teleportation belongs to the local subsystem BB' , and therefore measuring any witness $H_{B'B}$ satisfying the constraints in (2.37) is an allowed $\mathbf{LWCC}_{\rightarrow}$ operation, leading to the equality $\|Z_{B'B}\|_{\mathbf{LWCC}_{\rightarrow}} = \|Z\|_W$. As for the second term, we observe that $\|\tau(X_{AB} \otimes \sigma_{A'B'})\|_{\mathbf{LWCC}_{\rightarrow}} \leq \|X\|_{\mathbf{LWCC}_{\rightarrow}}$, since adding the ancillary system $A'B'$ in a separable state $\sigma_{A'B'}$ and subsequently applying τ is clearly an $\mathbf{LWCC}_{\rightarrow}$ protocol (which is why this latter inequality is in fact an equality). The above reasoning can be summarised in the following chain of inequalities, totally analogous to (3.17)-(3.21):

$$\begin{aligned}
\|Z_{AB}\|_{\mathbf{LOCC}_{\rightarrow}} &= \|Z_{AB}\|_{\mathbf{LWCC}_{\rightarrow}} \\
&= \left\| Z_{AB} \otimes \left(\frac{1}{n}\Phi + \frac{n-1}{n}\sigma \right)_{A'B'} \right\|_{\mathbf{LWCC}_{\rightarrow}} \\
&\geq \left\| \tau \left(Z_{AB} \otimes \left(\frac{1}{n}\Phi + \frac{n-1}{n}\sigma \right)_{A'B'} \right) \right\|_{\mathbf{LWCC}_{\rightarrow}} \\
&= \left\| \frac{1}{n} Z_{B'B} + \frac{n-1}{n} \tau(Z_{AB} \otimes \sigma_{A'B'}) \right\|_{\mathbf{LWCC}_{\rightarrow}} \\
&\geq \frac{1}{n} \|Z_{B'B}\|_{\mathbf{LWCC}_{\rightarrow}} - \frac{n-1}{n} \|\tau(Z_{AB} \otimes \sigma_{A'B'})\|_{\mathbf{LWCC}_{\rightarrow}} \\
&\geq \frac{1}{n} \|Z_{AB}\|_W - \frac{n-1}{n} \|Z_{AB}\|_{\mathbf{LWCC}_{\rightarrow}} \\
&= \frac{1}{n} \|Z_{AB}\|_W - \frac{n-1}{n} \|Z_{AB}\|_{\mathbf{LOCC}_{\rightarrow}}.
\end{aligned}$$

In conclusion, we find

$$\|Z_{AB}\|_{\mathbf{LOCC}_{\rightarrow}} = \|Z_{AB}\|_{\mathbf{LWCC}_{\rightarrow}} \geq \frac{1}{2n-1} \|Z_{AB}\|_W, \quad (3.26)$$

which implies $R_W(\mathbf{LOCC}_{\rightarrow}) \leq 2n-1$. Once more, to derive a lower bound on $R_W(\mathbf{SEP})$ we use Werner states [91, 130]. With the same notation as in the proof of Theorem 3.11, it can be shown that

$$\begin{aligned}
\left\| \frac{n+1}{2n-1} \rho_S - \frac{n-1}{2n-1} \rho_A \right\|_W &= 2, \\
\left\| \frac{n+1}{2n-1} \rho_S - \frac{n-1}{2n-1} \rho_A \right\|_{\mathbf{SEP}} &= \frac{2}{2n-1},
\end{aligned} \quad (3.27)$$

enforcing the complementary bound $R_W(\mathbf{SEP}) \geq 2n-1$. The proof of (3.27) is provided in Appendix E. \square

3.3.3 Further implications of the teleportation argument

This subsection is devoted to investigating some implications of the bound (3.13) we proved thanks to the teleportation argument (Theorem 3.11). As it turns out, the fact that (3.13) depends only on the minimal of the local dimensions has all sort of implications in improving previously known results. It is our purpose here to examine some of these improvements in detail, although a more thorough analysis will be conducted elsewhere.¹

Lower bounds on squashed entanglement

Among all the many entanglement measures that have been explored in the vast literature on the subject, one of the most mathematically appealing is perhaps **squashed entanglement** [158, 11]. For a bipartite quantum state ρ_{AB} , it is defined via the equation

$$E_{\text{sq}}(A : B)_\rho := \inf_{\rho_{ABC}} \frac{1}{2} I(A : B|C)_\rho, \quad (3.28)$$

where: (i) the infimum runs over all possible *extensions* ρ_{ABC} of ρ_{AB} , i.e. over all possible ancillary quantum systems C and over all possible density matrices ρ_{ABC} such that $\text{Tr}_C \rho_{ABC} = \rho_{AB}$; (ii) the **quantum conditional mutual information** $I(A : B|C)_\rho$ is defined via

$$I(A : B|C)_\rho := S(\rho_{AC}) + S(\rho_{BC}) - S(\rho_{ABC}) - S(\rho_C), \quad (3.29)$$

and

$$S(\rho) := -\text{Tr}[\rho \log \rho] \quad (3.30)$$

is the celebrated **von Neumann entropy**. Here, \log stands for the logarithms to base 2. It is a classic result of quantum information theory that the conditional mutual information (3.29) is always positive also in the quantum case, the corresponding inequality being known as **strong subadditivity** [159, 160, 161]. Consequently, the squashed entanglement of any state is necessarily non-negative. The fact that it is a legitimate entanglement measure, in the sense that it can not be increased by LOCC and it vanishes on separable states, was proven already in [11]. However, it took few years to show that the squashed entanglement is also *faithful*, in the sense that it vanishes *only* on separable states. This fact was finally shown in [162] by means of the inequality [162, Eq. (9) and (12)]

$$I(A : B|C)_\rho \geq 2E_{\text{sq}}(A : B)_\rho \geq \frac{1}{8 \ln 2} \|\rho_{AB} - \mathcal{S}_{A:B}\|_{\text{LOCC} \rightarrow}^2, \quad (3.31)$$

where

$$\|\rho_{AB} - \mathcal{S}_{A:B}\| := \inf_{\sigma \in \mathcal{S}_{A:B}} \|\rho_{AB} - \sigma_{AB}\|$$

¹Most of what I will say here came out of a series of very instructive conversations with Matthias Christandl, whom I thank wholeheartedly.

is the ‘distance’ of ρ_{AB} from the set $\mathcal{S}_{A:B}$ of normalised separable states of the bipartite system AB , as measured in an arbitrary norm $\|\cdot\|$. Observe that since the squashed entanglement is invariant under the exchange of A and B , there is no need to specify a specific direction of one-way classical communication (as long as it is fixed from the start). Clearly, since $\|\cdot\|_{\text{LOCC}\rightarrow}$ is a norm and not only a seminorm, (3.31) shows that a vanishing squashed entanglement ensures that the state under examination is separable.

Later on, in [163, p.4] the bound (3.31) was improved by a factor 4 to

$$I(A : B|C)_\rho \geq 2E_{\text{sq}}(A : B)_\rho \geq \frac{1}{2 \ln 2} \|\rho_{AB} - \mathcal{S}_{A:B}\|_{\text{LOCC}\rightarrow}^2 \quad (3.32)$$

as a by-product of the approach put forward in [163].

In [162], the lower bound (3.31) is employed in conjunction with (3.10) to obtain relations like

$$I(A : B|C)_\rho \geq 2E_{\text{sq}}(A : B)_\rho \quad (3.33)$$

$$\geq \frac{1}{306 \ln 2} \|\rho_{AB} - \mathcal{S}_{A:B}\|_2^2 \quad (3.34)$$

$$\geq \frac{1}{306 \ln 2 n_A n_B} \|\rho_{AB} - \mathcal{S}_{A:B}\|_1^2, \quad (3.35)$$

where we made use of the better constant in (3.32). A different strategy to the problem of finding lower bounds on the squashed entanglement was pursued in [164], where results of [165] are exploited in order to obtain a different lower bound

However, now that we have (3.13), we can employ it to obtain the best lower bound on the squashed entanglement in terms of the trace norm that is currently available.

Corollary 3.13. *For all states ρ_{AB} of a bipartite $n_A \times n_B$ quantum system, one has*

$$I(A : B|C)_\rho \geq 2E_{\text{sq}}(A : B)_\rho \geq \frac{1}{2 \ln 2 (2n - 1)^2} \|\rho_{AB} - \mathcal{S}_{A:B}\|_1^2, \quad (3.36)$$

where $n := \min\{n_A, n_B\}$, or equivalently

$$\begin{aligned} \|\rho_{AB} - \mathcal{S}_{A:B}\|_1 &\leq 4\sqrt{\ln 2} (n - 1/2) \sqrt{E_{\text{sq}}(A:B)_\rho} \\ &\leq 3.34 (n - 1/2) \sqrt{E_{\text{sq}}(A:B)_\rho}. \end{aligned} \quad (3.37)$$

Proof. The bound (3.36) is found by chaining (3.32) with (3.13), while (3.37) follows after straightforward algebraic manipulations. \square

The lower bound on the squashed entanglement in Corollary 3.13 combines the advantages of: (i) what you get from [162, Eq. (12)] by using the results of [151], here reported as (3.35); and (ii) [164, Corollary 2]. Namely, the trace-norm dependence is quadratic as in the former (and not quartic as in the latter), but the constant depends only on the minimal dimension as in the latter (and not on both as in the former).

A modified quantum de Finetti theorem

We already discussed briefly the complete extendibility criterion at the end of Subsection 2.4.1. In brief, it states that any separable state ρ_{AB} admits a *symmetric k -extension* for all positive integers k . This is a state $\rho_{AB_1 \dots B_k}$ of a quantum system $AB_1 \dots B_k$ (here the B_i are all isomorphic to B) which is: (i) symmetric under the exchange of any two B systems; and (ii) an extension of ρ_{AB} , in the sense that $\text{Tr}_{B_2 \dots B_k} \rho_{AB_1 \dots B_k}$ is nothing but the original state ρ (written in the registers AB_1). We also saw that a bipartite state that is k -extendible for all k , also called completely extendible, is necessarily separable [126]. Of course, checking infinitely many k s is not practically feasible, hence an approximate version of this latter statement is needed. Greatly simplifying the matter, we could say that a *quantum de Finetti theorem* serves this purpose, for it provides an estimate of how *close* a k -extendible state must be from the separable set. Delving into the details of the rich theory of quantum de Finetti theorems is beyond the scope of this thesis. We therefore limit ourselves to pointing the interested reader to some of many papers dedicated to the subject [166, 167, 168, 128, 162, 169].² A good introduction can also be found in [129, §9]. The theorem is named after Bruno de Finetti, who proved an earlier (classical) version of the statement [170].

Our goal here is way less ambitious. We follow [162] closely and rely heavily on the results presented there to show the following.

Corollary 3.14. *Let ρ_{AB} be a quantum state of a bipartite $n_A \times n_B$ quantum system that is k -extendible on the B side. Calling $n := \min\{n_A, n_B\}$, we have*

$$\|\rho_{AB} - \mathcal{S}_{A:B}\|_1 \leq 4\sqrt{\ln 2} (n - 1/2) \sqrt{\frac{\log n}{k}}. \quad (3.38)$$

Proof. It suffices to combine [162, Corollary 2] (with the better constant found in (3.32)) with our estimate (3.13). Alternatively, one can resort directly to (3.37) and to the observation that the squashed entanglement of any k -extendible state is no larger than $\frac{\log n}{k}$ [162]. \square

An interesting feature of the right hand side of (3.38) is that it stays finite when the dimension of one of the two subsystems diverges.

Algorithms for the separability problem

Following [162, 171], we know that any quantum de Finetti theorem can be applied to devise an algorithm for the weak membership problem for separability [171, Problem 1]. In [162, Corollary 3] (or in [171, Theorem 1]), the existence of an algorithm for deciding $W_{\text{SEP}}(\epsilon, \|\cdot\|_{\text{LOCC}}$) that takes an amount of time $\exp(O(\epsilon^{-2} \log n_A \log n_B))$ is shown. Plugging in (3.13), we can turn this into an algorithm for the same membership problem in terms the trace norm.

²The list is by no means complete nor representative.

Corollary 3.15. *Consider a bipartite $n_A \times n_B$ quantum system. There is an algorithm for deciding $W_{\text{SEP}}(\epsilon, \|\cdot\|_1)$ in $\exp(O(\epsilon^{-2}n^2 \log n_A \log n_B))$ time, where $n := \min\{n_A, n_B\}$.*

Proof. Because of the inequality (3.13), once we have an algorithm that can decide $W_{\text{SEP}}\left(\frac{\epsilon}{2^{n-1}}, \|\cdot\|_{\text{LOCC}}\right)$ in time T , we can also decide $W_{\text{SEP}}(\epsilon, \|\cdot\|_1)$ in time T . Thus, using [162, Corollary 3] (equivalently, [171, Theorem 1]), we get an algorithm for $W_{\text{SEP}}(\epsilon, \|\cdot\|_1)$ running in $\exp(O(\epsilon^{-2}n^2 \log n_A \log n_B))$ time. \square

The above result shows that when one of the two dimensions is fixed, the dependence on the other one is only *polynomial*, and not exponential as suggested by the best known algorithms, which have worst case complexity $\exp(\text{poly}(n_A, n_B) \log(1/\epsilon))$. You get such an improvement at the price of a worse error dependence, which becomes in turn exponential instead of logarithmic. However, Corollary 3.15 is pretty remarkable.

3.4 Ultimate bound on data hiding effectiveness

Throughout this section, we will determine the ultimate data hiding ratios of Definition 3.9 up to an additive constant, thus proving the main result of this chapter, Theorem 3.21. In order to do this, we need to: (i) find an explicit example of two GPTs that combined achieve a high data hiding ratio (Subsection 3.4.1); and (ii) show that this is the largest ratio that is compatible with our assumptions on the composition of local theories (Subsection 3.4.3). As discussed at the end of Section 3.1.2, this answers the central question of our investigation, that is, the determination of the general constraints that data hiding is forced to obey in any GPTs.

3.4.1 Data hiding in spherical models

As it turns out, quantum mechanical data hiding, whose maximum efficiency scales as $\min\{\sqrt{d_A}, \sqrt{d_B}\}$, is not the strongest possible within the realm of GPTs. Instead, throughout this subsection we will show that the spherical model, as introduced at the end of Subsection 2.3.3 (see for instance (2.44)), exhibits a data hiding ratio against separable operations that is as large as $\min\{d_A, d_B\}$, thus quadratically larger than that in quantum mechanics.

Let us start with a quick recap of the notation we adopt for spherical models. For more details, we refer the reader to Subsection 2.3.3. The arena is as usual the vector space \mathbb{R}^d . Vectors are denoted by $x = (x_0, x_1, \dots, x_{d-1})^T$, and since $u := (1, 0, \dots, 0)^T$, the 0-th component represents the normalisation of x . The space can be thought of as $\mathbb{R}^d = \mathbb{R}u_* \oplus \mathbb{R}^{d-1}$, where $u_* = (1, 0, \dots, 0)^T$ is a distinguished normalised state, and the vector component \mathbb{R}^{d-1} is equipped with the Euclidean norm $|\cdot|_2$. A vector $v \in \mathbb{R}^{d-1}$ can be turned into an element of the global space by simply adding a zero as the first component, which we denote by $\hat{v} = (0, v) \in \mathbb{R}^d$. Conversely, we can extract from any

$x = (x_0, \bar{x}) \in \mathbb{R}^d$ its vector component $\bar{x} \in \mathbb{R}^{d-1}$. With this notation, the cone of unnormalised states takes the form $C_d = \{(x_0, \bar{x}) : x_0 \geq |\bar{x}|_2\}$. Remarkably, with the canonical identification $\mathbb{R}^d \simeq (\mathbb{R}^d)^*$, we have $C_d \simeq C_d^*$, i.e. the spherical model is (strongly) self-dual.

Consider now two spherical models of dimensions d_A and d_B . States of the bipartite system AB are naturally identifiable with matrices Z of size $d_A \times d_B$. The tensor product of the two vector components of the local spaces, that is, $\mathbb{R}^{d_B-1} \otimes \mathbb{R}^{d_B-1}$, is represented by the $(d_A - 1) \times (d_B - 1)$ bottom right corner of Z . Retaining only this corner of Z yields another matrix that we denote by $\bar{Z} \in \mathbb{R}^{(d_A-1) \times (d_B-1)}$. As usual, one can do the opposite as well, and build a $d_A \times d_B$ matrix out of a $(d_A - 1) \times (d_B - 1)$ matrix M by adding zeros on the first row and column. Explicitly,

$$\widehat{M}_{ij} = \begin{cases} 0 & \text{if } i = 0 \text{ or } j = 0, \\ M_{ij} & \text{if } i, j \geq 1. \end{cases}$$

In what follows, we will often use the notation $U = u_A \otimes u_B$ and $U_* = u_{A*} \otimes u_{B*}$.

Since our primary interest is in the exploration of the data hiding properties, according to Proposition 3.10 we construct a bipartite system AB by joining two spherical models $A = \text{Sph}_{d_A}$ and $B = \text{Sph}_{d_B}$ via the minimal tensor product, i.e. taking $AB = \text{Sph}_{d_A \text{ min}} \otimes \text{Sph}_{d_B}$. This latter assumption will be made throughout the rest of this subsection.

According to Proposition 3.6, the data hiding ratio against separable measurements can be computed once we know the expressions for both the separability norm and the base norm induced by the minimal tensor product. Instead of treating the general case, we show how to compute these norms for a restricted yet large class of matrices, that is, those having zero entries in the first row and column.

Lemma 3.16. *Consider a bipartite system $AB = \text{Sph}_{d_A \text{ min}} \otimes \text{Sph}_{d_B}$. Then for all $M \in \mathbb{R}^{(d_A-1) \times (d_B-1)}$ we have*

$$\|\widehat{M}\| = \|M\|_1, \quad \|\widehat{M}\|_{\text{SEP}} = \|M\|_\infty, \quad (3.39)$$

where it is understood that the base norm $\|\cdot\|$ is induced by the minimal tensor product, and $\|M\|_1, \|M\|_\infty$ denote the trace and operator norm of M , i.e. the sum and the largest of its singular values, respectively.

Proof. Consider an arbitrary dual tensor $W \in \mathbb{R}^{d_A \times d_B}$. Since the spherical model is self-dual, we can safely apply Corollary 2.26 and conclude that: (a) if W is separable then necessarily $\|\overline{W}\|_1 \leq W_{00}$; and (b) this condition is also sufficient when all cross terms W_{i0}, W_{0j} ($i, j \geq 1$) vanish.

Now, let us compute the distinguishability norm against separable operations, denoted by $\|\cdot\|_{\text{SEP}}$. On the one hand, the above necessary condition for

separability of effects shows that

$$\begin{aligned}
\|\widehat{M}\|_{\text{SEP}} &= \max \left\{ \langle W, \widehat{M} \rangle : U \pm W \in C_{d_A}^* \otimes_{\min} C_{d_B}^* \right\} \\
&\leq \max \left\{ \langle W, \widehat{M} \rangle : \|\overline{W}\|_1 \leq 1 \pm W_{00} \right\} \\
&= \max \left\{ \text{Tr} [\overline{W}^T M] : \|\overline{W}\|_1 + |W_{00}| \leq 1 \right\} \\
&= \|M\|_{\infty},
\end{aligned} \tag{3.40}$$

where we employed (3.5) to find a more compact expression for the separability norm, and we exploited the fact that trace norm and operator norm are dual to each other. On the other hand, the fact that $\|N\|_1 \leq 1$ is sufficient to guarantee the separability of $U \pm \widehat{N}$ leads us, again via (3.5), to the complementary bound

$$\begin{aligned}
\|\widehat{M}\|_{\text{SEP}} &\geq \max \left\{ \langle \widehat{N}, \widehat{M} \rangle : \|N\|_1 \leq 1 \right\} \\
&= \max \left\{ \text{Tr} [N^T M] : \|N\|_1 \leq 1 \right\} \\
&= \|M\|_{\infty}.
\end{aligned} \tag{3.41}$$

Our final task is the calculation of the base norm induced by the minimal tensor product. Thanks to the formula (2.6), we can write

$$\|\widehat{M}\| = \min \{ \langle U, Z_+ + Z_- \rangle : Z_{\pm} \in C_{d_A} \otimes_{\min} C_{d_B}, \widehat{M} = Z_+ - Z_- \}. \tag{3.42}$$

Since $Z_{\pm} \in C_{d_A} \otimes_{\min} C_{d_B}$ implies $\langle U, Z_{\pm} \rangle \geq \|\overline{Z}_{\pm}\|_1$ and $\widehat{M} = Z_+ - Z_-$ implies $M = \overline{Z}_+ - \overline{Z}_-$, we see that

$$\begin{aligned}
\|\widehat{M}\| &\geq \min \{ \|\overline{Z}_+\|_1 + \|\overline{Z}_-\|_1 : M = \overline{Z}_+ - \overline{Z}_-, Z_{\pm} \in C_{d_A} \otimes_{\min} C_{d_B} \} \\
&\geq \|M\|_1,
\end{aligned} \tag{3.43}$$

where the last lower bound follows from the triangular inequality. On the other hand, we construct an ansatz for Z_{\pm} achieving the above lower bound. From the singular value decomposition theorem, it is immediately seen that for all real matrices M there exists a decomposition $M = M_+ - M_-$ such that $\|M_+\|_1 = \|M_-\|_1 = \frac{\|M\|_1}{2}$, and consequently $\|M\|_1 = \|M_+\|_1 + \|M_-\|_1$. Then, consider $Z_{\pm} = \frac{\|M\|_1}{2} U + \widehat{M}_{\pm}$, so that $Z_+ - Z_- = \widehat{M}$. Since $(Z_{\pm})_{i0} = 0 = (Z_{\pm})_{0j}$ when $i, j \geq 1$, Corollary 2.26 tells us that the condition $(Z_{\pm})_{00} \geq \|\overline{Z}_{\pm}\|_1$ (satisfied by construction) is sufficient to ensure the separability of Z_{\pm} , hence this is a valid ansatz. We find

$$\|\widehat{M}\| \leq \|M_+\|_1 + \|M_-\|_1 = \|M\|_1, \tag{3.44}$$

thus showing that $\|\widehat{M}\| = \|M\|_1$. \square

Corollary 3.17. *In the bipartite GPT $AB = \text{Sph}_{d_A} \otimes_{\min} \text{Sph}_{d_B}$, the data hiding ratio against separable measurements can be lower bounded as*

$$R_{\text{Sph}}(\text{SEP}) \geq \min\{d_A, d_B\} - 1. \quad (3.45)$$

Consequently, the ultimate data hiding ratios against locally constrained sets of measurements obey

$$\begin{aligned} R_{\text{LO}}(d_A, d_B) &\geq R_{\text{LOSR}}(d_A, d_B) \\ &\geq R_{\text{LOCC}\rightarrow}(d_A, d_B) \\ &\geq R_{\text{LOCC}}(d_A, d_B) \\ &\geq R_{\text{SEP}}(d_A, d_B) \\ &\geq \min\{d_A, d_B\} - 1. \end{aligned} \quad (3.46)$$

Proof. From Lemma 3.16 we know that $R_{\text{Sph}}(\text{SEP})$ can be lower bounded by the maximal ratio between trace norm and operator norm of matrices in $\mathbb{R}^{(d_A-1) \times (d_B-1)}$, which is well known to be $\min\{d_A-1, d_B-1\} = \min\{d_A, d_B\} - 1$. Since we can provide an example of bipartite GPT with local dimensions d_A, d_B for which the data hiding ratio against separable measurements is no smaller than $\min\{d_A, d_B\} - 1$, this constitutes a lower bound on the ultimate data hiding ratio $R_{\text{SEP}}(d_A, d_B)$. \square

The problem of finding a complementary upper bound for the data hiding ratios in the spherical model is solved by the general result expressed in the forthcoming Theorem 3.21, which implies that $R_{\text{Sph}}(\text{LO}) \leq \min\{d_A, d_B\}$. This yields the almost tight, two-sided bound

$$\begin{aligned} \min\{d_A, d_B\} - 1 &\leq R_{\text{Sph}}(\text{SEP}) \\ &\leq R_{\text{Sph}}(\text{LOCC}) \\ &\leq R_{\text{Sph}}(\text{LOCC}\rightarrow) \\ &\leq R_{\text{Sph}}(\text{LOSR}) \\ &\leq R_{\text{Sph}}(\text{LO}) \\ &\leq \min\{d_A, d_B\}, \end{aligned} \quad (3.47)$$

which fully determines the scaling of all the data hiding ratios against locally constrained measurements up to an additive constant.

Remark. Corollary 3.17 shows that quantum mechanics, even when it is modified to encompass the minimal tensor product composition rule according to Proposition 3.10, is not optimal from the point of view of data hiding, in the sense that there exist GPTs with the same local dimensions but with a (quadratically) higher data hiding ratio against all locally constrained sets of measurements.

3.4.2 A result on tensor norms

In this subsection, we will lay the foundations for the proof of the main result of the present chapter, Theorem 3.21 in the forthcoming Subsection 3.4.3. A decisive tool in our analysis will be the theory of tensor norms as discussed in Subsection 2.2.1. The reason why these objects play an important role here may not be a priori clear. However, consider that we are mainly interested in understanding the asymptotic behaviour of certain quantities constructed out of local Banach spaces as the dimension of those spaces goes to infinity. This study fits in the so called local theory of Banach spaces, which is concerned with the quantitative analysis of d -dimensional normed spaces (as $d \rightarrow \infty$). The investigation of tensor norms and their relations with operator ideals and factorizing operators is a crucial part of those studies [97]. Thus, in that respect it is not surprising that tensor norms play an important role in our approach.

More in detail, a crucial problem for us will be the comparison of injective and projective norm constructed out of general local Banach spaces of fixed (finite) dimension. Let us mention in passing that the same type of problem in the setting of infinite dimensional Banach leads to very deep results in functional analysis [172, 173]. In our context, the problem we want to address asks for the smallest constant $\kappa = \kappa(d_A, d_B)$ such that $\|\cdot\|_\pi \leq \kappa \|\cdot\|_\varepsilon$ holds true for all Banach spaces V_A, V_B of dimensions d_A, d_B , respectively. At this stage it is not even obvious that such a quantity will be finite. The answer to this question is provided by the following result, which could be known to experts in the topic, although we did not find any explicit reference.

Proposition 3.18. *For all pairs of finite dimensional Banach spaces V_A, V_B with dimensions $d_A = \dim V_A, d_B = \dim V_B$, we have*

$$\|\cdot\|_\pi \leq \min\{d_A, d_B\} \|\cdot\|_\varepsilon. \quad (3.48)$$

Furthermore, the constant on the right-hand side of the above inequality is the best possible for all pairs of positive integers d_A, d_B .

Proof. We start by recalling Auerbach's lemma, which we reported here as Lemma 2.1. This states that any finite-dimensional Banach space admits a basis $\{v_i\}_i$, whose associated dual basis we denote by $\{v_j^*\}_j$, such that (2.98) holds. Suppose without loss of generality that $d_A \leq d_B$. Expand any tensor $Z \in V_A \otimes V_B$ in the local Auerbach basis for V_A , that is, $Z = \sum_{j=1}^{d_A} v_j \otimes y_j$.³ Choose d_A functionals $\lambda_i \in V_B^*$ such that $\|\lambda_i\|_* \leq 1$ and $\langle \lambda_i, y_i \rangle = \|y_i\|$. Since

³We thank Guillaume Aubrun for having brought Auerbach's lemma and its usefulness in the context of this problem to our attention.

also $\|v_i^*\|_* \leq 1$, using (2.1) we can lower bound the injective norm as follows:

$$\begin{aligned} \|Z\|_\varepsilon &\geq \langle v_i^* \otimes \lambda_i, Z \rangle \\ &= \sum_{j=1}^{d_A} \langle v_i^*, v_j \rangle \langle \lambda_i, y_j \rangle \\ &= \sum_{j=1}^{d_A} \delta_{ij} \langle \lambda_i, y_j \rangle \\ &= \langle \lambda_i, y_i \rangle \\ &= \|y_i\|. \end{aligned}$$

Then, the definition of projective norm as given in (2.2) tells us that

$$\begin{aligned} \|Z\|_\pi &\leq \sum_{i=1}^{d_A} \|v_i\| \|y_i\| \\ &\leq \sum_{i=1}^{d_A} \|v_i\| \|Z\|_\varepsilon \\ &= d_A \|Z\|_\varepsilon, \end{aligned}$$

where we employed also the other defining property of the Auerbach basis, i.e. $\|v_i\| \leq 1$ for all i .

To see that the constant $\min\{d_A, d_B\}$ is optimal for all d_A, d_B , we resort to the results of Example 2.2. There, we saw that when V_A, V_B are both Euclidean spaces, the injective norm $\|\cdot\|_\varepsilon$ coincides with the matrix operator norm $\|\cdot\|_\infty$, while the corresponding projective norm is nothing but the matrix trace norm $\|\cdot\|_1$. Since $\|Z\|_1 = \min\{d_A, d_B\} \|Z\|_\infty$ whenever all the singular values of Z coincide, we see that the constant in (3.48) is the best possible. \square

3.4.3 Ultimate data hiding ratios

This subsection is devoted to the proof of the main result of the present chapter, Theorem 3.21. Namely, we want to show that the lower bound (3.46) on the ultimate data hiding ratios is substantially tight. In order to be able to do this, we have to address all GPTs at once, demonstrating that none of them can exhibit a larger ratio.

As we anticipated, the theory of tensor norms and in particular Proposition 3.48 are going to play a major role in our approach. In order to apply Proposition 3.18 to the problem, we need to relate the distinguishability norms on a composite system to the injective and projective norm constructed out of the local base norms. This translation between purely mathematical and physically motivated quantities is the subject of our next result.

Proposition 3.19. *Let $A = (V_A, C_A, u_A)$ and $B = (V_B, C_B, u_B)$ be two GPTs. The local base norms turn V_A, V_B into Banach spaces, and we can construct*

injective and projective tensor norms on $V_A \otimes V_B$, denoted simply by $\|\cdot\|_\varepsilon$ and $\|\cdot\|_\pi$. We can also define on $V_A \otimes V_B$: (i) the base norm $\|\cdot\|_{A \otimes_{\min} B}$ associated to the minimal tensor product $A \otimes_{\min} B$ taken as in (2.19); and (ii) the local distinguishability norm $\|\cdot\|_{\text{LO}}$. Then the following holds true:

$$\|\cdot\|_\pi = \|\cdot\|_{A \otimes_{\min} B}, \quad \|\cdot\|_\varepsilon \leq \|\cdot\|_{\text{LO}}. \quad (3.49)$$

Proof. We start by showing the first relation in (3.49). Consider $Z \in V_A \otimes V_B$, and decompose it as $Z = \sum_{i=1}^n x^i \otimes y^i$ in such a way that $\|Z\|_\pi = \sum_{i=1}^n \|x^i\| \|y^i\|$ (see the definition (2.2)). According to 2.6, we can construct $x_\pm^i \in C_A$ and $y_\pm^i \in C_B$ such that $x^{(i)} = x_+^i - x_-^i$ and $\|x^i\| = \langle u_A, x_+^i + x_-^i \rangle$ and analogously for y_\pm^i . Then

$$X = \sum_{i=1}^n (x_+^i \otimes y_+^i + x_-^i \otimes y_-^i) - \sum_{i=1}^n (x_+^i \otimes y_-^i + x_-^i \otimes y_+^i),$$

and since

$$\sum_{i=1}^n (x_+^i \otimes y_+^i + x_-^i \otimes y_-^i), \quad \sum_{i=1}^n (x_+^i \otimes y_-^i + x_-^i \otimes y_+^i) \in C_A \otimes_{\min} C_B,$$

the dual formula (2.6) yields

$$\begin{aligned} \|X\|_{A \otimes_{\min} B} &\leq \left\langle u_A \otimes u_B, \sum_{i=1}^n (x_+^i \otimes y_+^i + x_-^i \otimes y_-^i) \right\rangle \\ &\quad + \left\langle u_A \otimes u_B, \sum_{i=1}^n (x_+^i \otimes y_-^i + x_-^i \otimes y_+^i) \right\rangle \\ &= \sum_{i=1}^n \langle u_A, x_+^i + x_-^i \rangle \langle u_B, y_+^i + y_-^i \rangle \\ &= \sum_{i=1}^n \|x^i\| \|y^i\| \\ &= \|Z\|_\pi. \end{aligned}$$

To show the converse, notice first that $\|Z\|_\pi = \langle u_A \otimes u_B, Z \rangle$ holds true for all separable $Z \in C_A \otimes_{\min} C_B$. In fact, writing $Z = \sum_i x_i \otimes y_i$ for $x_i, y_i \geq 0$ positive, we see that on the one hand $\|Z\|_\pi \leq \sum_i \|x_i\| \|y_i\| = \sum_i \langle u_A, x_i \rangle \langle u_B, y_i \rangle = \langle u_A \otimes u_B, Z \rangle$, while on the other hand $\langle u_A \otimes u_B, Z \rangle \leq \|Z\|_\varepsilon \leq \|X\|_\pi$ by inequality (2.4). Now, for a generic $Z \in V_A \otimes V_B$ apply once again (2.6) to construct $Z_\pm \in C_A \otimes_{\min} C_B$ such that $Z = Z_+ - Z_-$ and

$$\|X\|_{A \otimes_{\min} B} = \langle u_A \otimes u_B, Z_+ + Z_- \rangle.$$

By using the above observation and the triangular inequality, we find

$$\begin{aligned} \|Z\|_\pi - \langle u_A \otimes u_B, Z_- \rangle &= \|Z\|_\pi - \|Z_-\|_\pi \\ &\leq \|Z + Z_-\|_\pi \\ &= \|Z_+\|_\pi \\ &= \langle u_A \otimes u_B, Z_+ \rangle, \end{aligned}$$

from which

$$\|Z\|_\pi \leq \langle u_A \otimes u_B, Z_- + Z_+ \rangle = \|Z\|_{A \otimes_{\min} B}.$$

This completes the proof of the equality $\|\cdot\|_\pi = \|\cdot\|_{A \otimes_{\min} B}$.

Let us now show the second relation in (3.49). To find a lower bound on the separability norm, we will employ the expression (3.5) with $\mathcal{M} = \text{LO}$. For arbitrary $Z \in V_A \otimes V_B$ and $\varphi \in V_A^*$, $\lambda \in V_B^*$ such that $\|\varphi\|_*, \|\lambda\|_* \leq 1$, we show that $\varphi \otimes \lambda$ is a valid test functional to be plugged into (3.5) since $\mu := (\frac{1}{2}(u_A \otimes u_B + \varphi \otimes \lambda), \frac{1}{2}(u_A \otimes u_B - \varphi \otimes \lambda)) \in \langle \text{LO} \rangle$. Before we prove this latter claim, let us note in passing that μ is easily seen to be at least a separable measurement, as follows from Theorem 2.23. Now, we show that $\mu \in \langle \text{LO} \rangle$. In fact, notice that

$$\frac{1}{2}(u_A \otimes u_B + \varphi \otimes \lambda) = \frac{u_A + \varphi}{2} \otimes \frac{u_B + \lambda}{2} + \frac{u_A - \varphi}{2} \otimes \frac{u_B - \lambda}{2}, \quad (3.50)$$

$$\frac{1}{2}(u_A \otimes u_B - \varphi \otimes \lambda) = \frac{u_A + \varphi}{2} \otimes \frac{u_B - \lambda}{2} + \frac{u_A - \varphi}{2} \otimes \frac{u_B + \lambda}{2}. \quad (3.51)$$

Thanks to the fact that the unit balls of the dual to the local base norms have the form $B_{\|\cdot\|_*} = [-u, u]$ (Definition 1.32), we know that

$$\left(\frac{1}{2}(u_A + \varphi), \frac{1}{2}(u_A - \varphi) \right), \quad \left(\frac{1}{2}(u_B + \lambda), \frac{1}{2}(u_B - \lambda) \right)$$

are valid measurements on A and B , respectively. Using (2.21), it is easy to see that μ is indeed obtainable from a product measurement via a coarse graining procedure as defined in (2.8). Then, equation (3.5) yields $\|Z\|_{\text{LO}} \geq \langle \varphi \otimes \lambda, Z \rangle$, which becomes in turn $\|Z\|_{\text{LO}} \geq \|Z\|_\varepsilon$ once we maximise over the functionals φ, λ satisfying $\|\varphi\|_*, \|\lambda\|_* \leq 1$. \square

Corollary 3.20. *Any global base norm $\|\cdot\|$ in a bipartite GPT constructed according to (2.18) must obey $\|x \otimes y\| = \|x \otimes y\|_{\text{LO}} = \|x\| \|y\|$ for all $x \in V_A$ and $y \in V_B$, where $\|x\|, \|y\|$ stand for the local base norms.*

Proof. On the one hand, since (2.18) holds, the global base norms must be upper bounded by the one associated to the minimal tensor product, which coincides with $\|\cdot\|_\pi$ by Proposition 3.19. On the other hand, the second identity in (3.49) ensures that $\|\cdot\|_{\text{LO}} \geq \|\cdot\|_\varepsilon$. We already saw how injective

and projective norm coincide on simple tensors. Then, putting all together we obtain

$$\begin{aligned}
\|x\| \|y\| &= \|x \otimes y\|_\varepsilon \\
&\leq \|x \otimes y\|_{\text{LO}} \\
&\leq \|x \otimes y\| \\
&\leq \|x \otimes y\|_{A \otimes_{\min} B} \\
&= \|x \otimes y\|_\pi \\
&= \|x\|, \|y\|,
\end{aligned}$$

concluding the proof. \square

We are finally ready to prove one of our main results, that is, the optimality of the lower bound (3.46) (up to an additive constant).

Theorem 3.21 (Upper bound on ultimate data hiding ratios). *Let A, B be two GPTs, and let their composite AB obey (2.18). Then the data hiding ratios against locally constrained sets of measurements satisfy the bound*

$$\begin{aligned}
R(\text{SEP}) &\leq R(\text{LOCC}) \\
&\leq R(\text{LOCC}_{\rightarrow}) \\
&\leq R(\text{LOSR}) \\
&\leq R(\text{LO}) \\
&\leq \max_{0 \neq Z \in V_A \otimes V_B} \frac{\|Z\|_\pi}{\|Z\|_\varepsilon}.
\end{aligned} \tag{3.52}$$

In particular, the corresponding ultimate data hiding ratios, as given in Definition 3.9, are bounded as follows:

$$\begin{aligned}
\min\{d_A, d_B\} - 1 &\leq R_{\text{SEP}}(d_A, d_B) \\
&\leq R_{\text{LOCC}}(d_A, d_B) \\
&\leq R_{\text{LOCC}_{\rightarrow}}(d_A, d_B) \\
&\leq R_{\text{LOSR}}(d_A, d_B) \\
&\leq R_{\text{LO}}(d_A, d_B) \\
&\leq \min\{d_A, d_B\}.
\end{aligned} \tag{3.53}$$

Proof. Since the inequalities (3.8) hold, we have to upper bound only the data hiding ratio against local operations. Using (3.6) and Proposition 3.10, we know that for a fixed pair of GPTs $A = (V_A, C_A, u_A)$ and $B = (V_B, C_B, u_B)$, this latter ratio satisfies

$$R(\text{LO}) \leq \max_{Z \neq 0} \left\{ \|Z\|_{A \otimes_{\min} B} \|Z\|_{\text{LO}}^{-1} \right\}.$$

Now, Proposition 3.19 states that $\|Z\|_{A \otimes_{\min} B} = \|Z\|_\pi$ and $\|Z\|_{\text{LO}}^{-1} \leq \|Z\|_\varepsilon^{-1}$, from which we deduce (3.52). Using also Proposition 3.18, we see that the

right-hand side of (3.52) is upper bounded by $\min\{d_A, d_B\}$ for all GPTs of fixed local dimensions d_A, d_B , proving also (3.53). \square

The above Theorem 3.21 is remarkable for several reasons. First, and most obviously, it solves the main problem that guided our investigation in the present chapter, by determining the exact scaling of the ultimate data hiding ratios against locally constrained sets of measurements in the pair of local dimensions. Second, it gives us a way to explicitly upper bound those ratios in any given model, via (3.52). Third, it shows a surprising similarity among all the classes of locally constrained measurements from the point of view of ultimate data hiding efficiency, which is not to be expected a priori.

Finally, by comparing Theorem 3.11 and 3.21 we can now tell how strong quantum mechanical data hiding is as compared to the strongest that is conceivable given the non-signalling constraint. When the strength is measured by the data hiding ratios, it turns out that quantum mechanics sits right in the middle between classical theories and the other extremal example provided by the spherical model. That is, the quantum mechanical data hiding ratio scales as the geometric mean between those pertaining to the two extremes.⁴

Remark. Another example of a nontrivial consequence of Theorem 3.21 is as follows. Applying (3.53) to a bipartite quantum mechanical system $\mathbb{C}^{n_A} \otimes \mathbb{C}^{n_B}$, we obtain the estimate $R(\text{LO}) \leq \min\{d_A, d_B\} = \min\{n_A^2, n_B^2\}$. This must be compared with the best known upper bound $R(\text{LO}) \leq \sqrt{153 n_A n_B}$ [151]. The comparison shows how neither of the two results is tight. If n_A and n_B are very different from each other the former bound will be more effective, while if they are of the same order the latter will be preferable. In conclusion, as we already stressed when discussing data hiding in quantum mechanics, determining the optimal scaling of $R(\text{LO})$ remains an interesting open problem. At the same time, this example shows how consequences drawn from general results like Theorem 3.21 can shed some light even on well-studied problems in quantum information.

3.5 Data hiding in special classes of GPTs

Until now, we have been mainly interested in investigating the strongest cases of data hiding, in order to study the ultimate, intrinsic bounds characterising this non-classical phenomenon. For how well-motivated this inclination to universality can be, throughout this section we want to take a different approach and look into particular classes of models.

Subsection 3.5.1 is concerned with all the GPTs whose state space, just like in the spherical model, is centrally symmetric. It turns out that in this case all the data hiding ratios against locally restricted sets of measurements are equal up to an additive constant, and moreover can be computed as a maximal ratio

⁴If this rings a Bell, it should. Indeed, the same happens with the CHSH inequality, namely, the maximal quantum mechanical violation (‘Tsirelson bound’) is $2\sqrt{2}$, i.e. the geometric mean between the classical value, i.e. 2, and the PR-box violation, which is 4.

between a projective and an injective norm. This has the merit of explaining the results of Subsection 3.4.1 in a more general context.

This is reminiscent of the proof of Theorem 3.21, where we upper bounded $R(\text{LO})$ with the maximal projective-injective ratio induced by the local base norms. However, while in that case the only piece of information we could extract was the existence of the upper bound, here computing such a quantity yields the exact data hiding ratio up to additive constants (Theorem 3.24). That this is not just a alternative way of rephrasing the problem, but that on the contrary it can be instrumental in solving it, is apparent from our work in Subsection 3.4.1, where we solved the spherical model thanks to this trick. We show how this approach can be pushed further, by using the machinery we develop to solve another ‘natural’ model (Example 3.25).

Throughout Subsection 3.5.2, we look into those GPTs whose vector space is endowed with a representation of a compact group that maps states to states but is otherwise irreducible on the section $\{x \in V : \langle u, x \rangle = 0\}$. Although this assumption is quite strong, we show how it encompasses the main physically relevant examples of GPTs, like classical probability theory and quantum theory. With the group integral at hand, we are able to generalise the construction of Werner states and to use them for estimating data hiding in terms of simple geometrical parameters of the model (Theorem 3.33).

3.5.1 Centrally symmetric models

The solution of the spherical model we gave in Subsection 3.4.1 through Lemma 3.16 and Corollary 3.17 was based on two remarkable facts: on the one hand, we could achieve the data hiding ratio (up to an additive constant) by employing only tensors of the form \widehat{M} for $M \in \mathbb{R}^{(d_A-1) \times (d_B-1)}$, and on the other hand we were able to find simple expressions for base and separability norm of tensors of this simplified form. Here, we want to take the chance to generalise these intuitions a bit further, to encompass any centrally symmetric model (Definition 2.11).

For a thorough introduction to these models we refer the reader to Subsection 2.3.3. Here, let us just recall that a centrally symmetric GPT is defined on a vector space $V = \mathbb{R}^d = \mathbb{R}u_* \oplus \mathbb{R}^{d-1}$ (2.41), where $u_* = (1, 0, \dots, 0)^T$ is a distinguished state. For $x \in \mathbb{R}^d$, we denote the second component with respect to the above decomposition with \bar{x} . Conversely, for $v \in \mathbb{R}^{d-1}$ one can construct $\widehat{v} := (0, v) \in \mathbb{R}^d$. The positive cone C is defined by $C = \{(x_0, \bar{x}) \in \mathbb{R} \oplus \mathbb{R}^{d-1} : x_0 \geq |\bar{x}|\}$, with $|\cdot|$ being an arbitrary norm on \mathbb{R}^{d-1} . As can be easily verified, the dual cone to C is given by $C^* = \{(y_0, \bar{y}) \in \mathbb{R} \oplus \mathbb{R}^{d-1} : y_0 \geq |\bar{y}|_*\}$ (2.39), where $|\cdot|_*$ is the dual to the norm $|\cdot|$. In particular, C^* is centrally symmetric if so is C . We remind the reader that the base norm of a centrally symmetric model is given by $\|(x_0, \bar{x})\| = \max\{|x_0|, |\bar{x}|\}$ (2.40). For centrally symmetric models there exists a simple positive linear map, given by $T := 1 \oplus (-I_{d-1})$ (2.42).

When one combines two centrally symmetric models of dimensions d_A, d_B , the resulting vector space contains a ‘double’ vector component $\mathbb{R}^{(d_A-1) \times (d_B-1)}$

that is obtained by tensoring the vector components of the two spaces (2.43). Analogously to the case of a single system, we can ‘lower’ matrices $Z \in \mathbb{R}^{d_A \times d_B}$ to $\bar{Z} \in \mathbb{R}^{(d_A-1) \times (d_B-1)}$ by retaining only this component, and conversely ‘lift’ $M \in \mathbb{R}^{(d_A-1) \times (d_B-1)}$ to $\widehat{M} \in \mathbb{R}^{d_A \times d_B}$ by adding a row and a column of zeros.

As we proved in Proposition 2.25, a bipartite tensor of the form $U_* + \widehat{M}$, where $U_* = u_{A*} \otimes u_{B*}$, belongs to the maximal tensor product iff $|M|_\pi \leq 1$, and to the minimal tensor product iff $|M|_\varepsilon \leq 1$. Here, $|\cdot|_\varepsilon$ and $|\cdot|_\pi$ denote the injective and projective tensor norms constructed out of the local norms $|\cdot|$. As a side remark, observe that the maps $T \otimes I$, $I \otimes T$, $T \otimes T$ preserve separability of states and of effects. Therefore, the norm $\|\cdot\|_{\text{SEP}}$ is left invariant by any of these maps. As is easy to verify, the same is true for all the locally constrained distinguishability norms, with the possible exception of $\|\cdot\|_{\text{LOCC}}$.

Concerning the analysis of the data hiding properties, what we did in the case of the spherical model was to consider only tensors of the form \widehat{M} , and to a posteriori justify this restriction by employing Theorem 3.21 to show that the obtained result is tight up to an additive constant. As it turns out, this procedure can be always followed for centrally symmetric GPTs without loss of generality. This is the content of our first result.

Proposition 3.22. *Let AB be a bipartite system formed by two centrally symmetric GPTs joined with any rule that respects (2.18). Then for all $Z \in \mathbb{R}^{d_A \times d_B}$ we have*

$$\frac{\|Z\|}{\|Z\|_{\text{LO}}} \leq \frac{\|\bar{Z}\|}{\|\bar{Z}\|_{\text{LO}}} + 2. \quad (3.54)$$

Proof. Consider $Z = s u_* \otimes u_* + x \otimes u_* + u_* \otimes y + \bar{Z} \in \mathbb{R}^{d_A \times d_B}$, where we omitted the subscripts A, B for the sake of simplicity. Applying the triangular inequality to the global base norm, we see that $\|Z\| \leq \|(s u_* + x) \otimes u_*\| + \|u_* \otimes y\| + \|\bar{Z}\|$. Now, Corollary 3.20 guarantees that $\|(s u_* + x) \otimes u_*\| = \|(s u_* + x) \otimes u_*\|_{\text{LO}} = \|s u_* + x\|$ and $\|u_* \otimes y\| = \|u_* \otimes y\|_{\text{LO}} = \|y\|$. Moreover, it is also clear by discarding the system B and performing an arbitrary operation on A , that $\|Z\|_{\text{LO}} \geq \|s u_* + x\|$. Proceeding in an analogous fashion with exchanged subsystems, remembering that Λ defined in (2.42) leaves the base norm invariant, and exploiting the triangular inequality, yields

$$\begin{aligned} \|Z\|_{\text{LO}} &\geq \|s u_* + y\| \\ &= \|T(s u_* + y)\| \\ &= \|s u_* - y\| \\ &\geq \left\| \frac{1}{2}(s u_* + y) - \frac{1}{2}(s u_* - y) \right\| \\ &= \|y\|. \end{aligned}$$

Finally, using the readily verified invariance of $\|\cdot\|_{\text{LO}}$ under any of the maps

$T \otimes I, I \otimes T, T \otimes T$, we obtain

$$\begin{aligned} \|X\|_{\text{LO}} &\geq \left\| \frac{1}{4}Z - \frac{1}{4}(T \otimes I)(Z) - \frac{1}{4}(I \otimes T)(Z) + \frac{1}{4}(T \otimes T)(Z) \right\|_{\text{LO}} \\ &= \|\bar{Z}\|_{\text{LO}}. \end{aligned}$$

Putting all together, we find

$$\begin{aligned} \frac{\|Z\|}{\|Z\|_{\text{LO}}} &\leq \frac{\|su_* + x\| + \|y\| + \|\bar{Z}\|}{\|Z\|_{\text{LO}}} \\ &= \frac{\|su_* + x\|}{\|X\|_{\text{LO}}} + \frac{\|y\|}{\|Z\|_{\text{LO}}} + \frac{\|\bar{Z}\|}{\|Z\|_{\text{LO}}} \\ &\leq 2 + \frac{\|\bar{Z}\|}{\|Z\|_{\text{LO}}} \end{aligned}$$

□

What Proposition 3.22 is telling us is that up to an additive constant we can restrict the search for data hiding against local operations to projected tensors of the form \bar{Z} . From now on, we denote by $\bar{R}(\mathcal{M})$ the *restricted data hiding ratio* against a set of measurements \mathcal{M} that is obtained by considering only those tensors. With this notation, Proposition 3.22 can be cast into the form of the inequality $R(\text{LO}) \leq \bar{R}(\text{LO}) + 2$. In order to carry out the analysis of restricted ratios, we need an analogue of Lemma 3.16.

Proposition 3.23. *Let $|\cdot|_\varepsilon, |\cdot|_\pi$ be the injective and projective norm constructed on $\mathbb{R}^{(d_A-1) \times (d_B-1)} \simeq \mathbb{R}^{d_A-1} \otimes \mathbb{R}^{d_B-1}$ out of the local norms $|\cdot|$ on $\mathbb{R}^{d_A-1}, \mathbb{R}^{d_B-1}$. Then for $M \in \mathbb{R}^{(d_A-1) \times (d_B-1)}$ the locally constrained distinguishability norms and the global base norm induced by the minimal tensor product are respectively given by*

$$\|\widehat{M}\|_{\text{LO}} = \|\widehat{M}\|_{\text{LOSR}} = \|\widehat{M}\|_{\text{LOCC} \rightarrow} = \|\widehat{M}\|_{\text{LOCC}} = \|\widehat{M}\|_{\text{SEP}} = |M|_\varepsilon, \quad (3.55)$$

$$\|\widehat{M}\|_{A \otimes_{\min} B} = |M|_\pi. \quad (3.56)$$

Proof. The argument follows the guidelines of the proof of Proposition 3.16, so we omit some of the details. First of all, thanks to Proposition 2.25 tensor $W \in \mathbb{R}^{d_A \times d_B}$ is separable then necessarily $|\overline{W}|_{*\pi} \leq W_{00}$, and this condition is also sufficient when all cross terms W_{i0}, W_{0j} ($i, j \geq 1$) vanish.

To compute the separability norm, on the one hand as in (3.40) we find

$$\begin{aligned} \|\widehat{M}\|_{\text{SEP}} &\leq \max \{ \text{Tr} [\overline{W}^T M] : |\overline{W}|_{*\pi} + |W_{00}| \leq 1 \} \\ &= |M|_{*\pi*} \\ &= |M|_\varepsilon, \end{aligned}$$

where in the last step we used (2.3). On the other hand, the complementary inequality $\|\widehat{M}\|_{\text{LO}} \geq |M|_\varepsilon$ is a simple consequence of the second relation

in (3.49). In fact, since for all $v \in \mathbb{R}^{d_A-1}$ the identity $|v|_* = \|\widehat{v}\|$ holds true, we obtain

$$\begin{aligned} \|\widehat{M}\|_{\text{LO}} &\geq \|\widehat{M}\|_\varepsilon \\ &= \max \left\{ \langle \varphi \otimes \lambda, \widehat{M} \rangle : \|\varphi\|_*, \|\lambda\|_* \leq 1 \right\} \\ &\geq \max \left\{ v^T M w : |v|_*, |w|_* \leq 1 \right\} \\ &= |M|_\varepsilon. \end{aligned}$$

Putting all together, we obtain

$$|M|_\varepsilon \leq \|\widehat{M}\|_{\text{LO}} \leq \|\widehat{M}\|_{\text{LOSR}} \leq \|\widehat{M}\|_{\text{LOCC}\rightarrow} \leq \|\widehat{M}\|_{\text{LOCC}} \leq \|\widehat{M}\|_{\text{SEP}} \leq |M|_\varepsilon,$$

from which (3.55) follows. \square

Theorem 3.24. *For a fixed pair of centrally symmetric GPTs whose composite is formed with the minimal tensor product rule according to Proposition 3.10, all the four restricted data hiding ratios against locally constrained measurements coincide, and can be computed as*

$$\bar{R} := \bar{R}(\text{LO}) = \bar{R}(\text{LOSR}) = \bar{R}(\text{LOCC}\rightarrow) = \bar{R}(\text{LOCC}) = \bar{R}(\text{SEP}) \quad (3.57)$$

$$= \max_{0 \neq M \in \mathbb{R}^{(d_A-1) \times (d_B-1)}} \frac{|M|_\pi}{|M|_\varepsilon}. \quad (3.58)$$

Moreover, the ‘true’ data hiding ratios are equal to the restricted ones up to an additive constant. Namely,

$$\bar{R} \leq R(\text{SEP}) \leq R(\text{LOCC}) \leq R(\text{LOCC}\rightarrow) \leq R(\text{LOSR}) \leq R(\text{LO}) \leq \bar{R} + 2. \quad (3.59)$$

Proof. First of all, (3.58) follows directly from Proposition 3.23. To show (3.59), note that, by definition, restricted data hiding ratios are smaller than the true ones. In particular, $R(\text{SEP}) \geq \bar{R}(\text{SEP}) = \bar{R}$. Using (3.8) and the result of Proposition 3.22 in the form of the inequality $R(\text{LO}) \leq \bar{R}(\text{LO}) + 2$, we obtain immediately (3.59). \square

Remark. Sometimes the quantity \bar{R} appearing in (3.58) can be better computed with the help of a simple trick. Namely, remember that for all norms $|\cdot|^{(1)}, |\cdot|^{(2)}$ on any space V we have $\max_{0 \neq x \in V} \frac{|x|^{(1)}}{|x|^{(2)}} = \max_{0 \neq \varphi \in V^*} \frac{|\varphi|^{(2)}}{|\varphi|^{(1)}}$. Thanks to this identity and to the duality relation (2.3), we see that \bar{R} can alternatively be expressed as

$$\bar{R} = \max_{0 \neq M \in \mathbb{R}^{(d_A-1) \times (d_B-1)}} \frac{|M|_{*\varepsilon}}{|M|_{*\pi}}. \quad (3.60)$$

Theorem 3.24 suggests a precise recipe for computing data hiding ratios of centrally symmetric models. It is in fact enough to analyse the behaviours of injective and projective norms, compute \bar{R} defined in (3.58), and finally use (3.59) to determine all the ratios against locally constrained measurements up to a universal additive constant. In fact, this is basically what we did in Subsection 3.4.1, since trace norm and operator norm are exactly the injective and projective norms constructed out of local Euclidean spaces, as we saw in the proof of Proposition 3.18. In the remaining part of this subsection, we are going to show how to apply all this machinery to solve another explicit example.

Example 3.25 (Cubic model). Let us consider the class of cubic models G_n defined in Subsection 2.3.4. We remind the reader that the state space of G_n is an n -dimensional hypercube. Here, we are mainly interested in discussing the data hiding properties of GPTs of the form $G_n \otimes_{\min} G_m$, for n, m positive integers. Since the n -dimensional hypercube is centrally symmetric, this is a perfect playground for testing the machinery we developed throughout this subsection. Notice that in the case of G_n the norm $|\cdot|$ on \mathbb{R}^n appearing in Definition 2.11 is given by the ∞ -norm $|v|_\infty := \max_{1 \leq i \leq n} |v_i|$, with dual $|w|_{\infty*} = |w|_1 = \sum_{i=1}^n |w_i|$. It is worth noticing that the extreme points of the unit ball $B_{|\cdot|_1}$ coincide with the elements of the standard basis of \mathbb{R}^n up to a sign.

According to Theorem 3.24, the first step in solving the GPT $G_n \otimes_{\min} G_m$ is the determination of the restricted data hiding ratio \bar{R} in (3.58), where in our case $|\cdot|_\varepsilon, |\cdot|_\pi$ are the injective and projective norm induced on $\mathbb{R}^{n \times m}$ by the local ∞ -norms. We already observed how this can be dually rephrased as (3.60). In our case, this is helpful because the injective and projective dual norms $|\cdot|_{*\varepsilon}, |\cdot|_{*\pi}$ can be easily computed as follows. The former is given by

$$\begin{aligned} |M|_{*\varepsilon} &= \max_{|v|_\infty, |w|_\infty \leq 1} v^T M w \\ &= \max_{s \in \{\pm 1\}^n, t \in \{\pm 1\}^m} s^T M t \\ &= \|M\|_{\infty \rightarrow 1}, \end{aligned} \tag{3.61}$$

while the latter can be found thanks to the chain of inequalities

$$|M|_{*\pi} = |M|_{\varepsilon*} = |M|_{\infty*} = |M|_1, \tag{3.62}$$

where we used the fact that since

$$|N|_\varepsilon = \max_{|v|_1, |w|_1 \leq 1} v^T N w = \max_{i,j} |N_{ij}| = |N|_\infty,$$

one has $|M|_{\varepsilon*} = |M|_1 = \sum_{i,j} |M_{ij}|$, where we extended the definitions of ∞ -norm and 1-norm to matrices in an obvious way.

Now, we have to find the largest ratio $|M|_1 / \|M\|_{\infty \rightarrow 1}$ that is achievable by $M \in \mathbb{R}^{n \times m}$. The argument comprises two parts: first, we have to exhibit an

explicit example M_0 displaying a high ratio, and secondly, we have to show that this is optimal up to a constant factor.

Let us start with the first task. From now on, we will assume without loss of generality $n \leq m$. Suppose that there exists a matrix $H \in \{\pm 1\}^{n \times m}$ made of signs such that $HH^T = m\mathbf{1}_n$ (such matrices are often called *partial Hadamard matrices*). Then simple considerations very close in spirit to Lindsey's lemma [174, 175] (see also [176, 177]) show that for all $s \in \{\pm 1\}^n$ and $t \in \{\pm 1\}^m$ we have

$$\begin{aligned} s^T H t &\leq \sum_j \left| \sum_i s_i H_{ij} \right| \\ &\leq \sqrt{m} \sqrt{\sum_j \left| \sum_i s_i H_{ij} \right|^2} \\ &= \sqrt{m} \sqrt{\sum_{ijk} s_i s_k H_{ij} H_{kj}} \\ &= \sqrt{m} \sqrt{\sum_{ik} s_i s_k n \delta_{ik}} \\ &= n\sqrt{m}. \end{aligned}$$

Maximising over sign vectors s, t yields $\|H\|_{\infty \rightarrow 1} \leq n\sqrt{m}$. Since $|H|_1 = nm$, we obtain the lower bound $|H|_1 / \|H\|_{\infty \rightarrow 1} \geq \sqrt{n}$. Now, the existence of a matrix H with the properties required for the above construction is not guaranteed for arbitrary integers n, m . In fact, this is never the case when $n = m > 2$ and $4 \nmid n$, while when $n = m = 4k$ it is the content of the so-called *Hadamard conjecture*. However, observe that Hadamard matrices are elementarily guaranteed to exist for $n = m = 2^k$, as the explicit example $\left(\begin{smallmatrix} 1 & 1 \\ -1 & 1 \end{smallmatrix}\right)^{\otimes k}$ shows. Since for all n there is n' power of 2 such that $\frac{n}{2} \leq n' \leq n$, we get $|H|_1 / \|H\|_{\infty \rightarrow 1} \geq \sqrt{n'} \geq \sqrt{\frac{n}{2}}$. This shows that

$$\max_{0 \neq M \in \mathbb{R}^{n \times m}} \frac{|M|_{*\pi}}{|M|_{*\varepsilon}} = \max_{0 \neq M \in \mathbb{R}^{n \times m}} \frac{|M|_1}{\|M\|_{\infty \rightarrow 1}} \geq \sqrt{\frac{n}{2}}. \quad (3.63)$$

We note in passing that the crude estimate (3.63) can be improved to \sqrt{n} when $n \leq \frac{m}{2}$ by the same tricks. Further asymptotic refinements can be obtained with the help of sophisticated results such as [178], but this is beyond the scope of the present paper.

Now, we have to show that the above lower bound is optimal, i.e. that the scaling of the two sides of (3.63) are the same. In order to do so, we resort to a celebrated inequality known as Khintchine inequality [179, 180, 181], which states that for all $c \in \mathbb{R}^n$, if s_1, \dots, s_n is an i.i.d. sequence with Rademacher distribution, then

$$\mathbb{E}_s \left| \sum_i s_i c_i \right| \geq \frac{|c|_2}{\sqrt{2}}, \quad (3.64)$$

where $|c|_2^2 = \sum_i c_i^2$. Thanks to this result, we know that whenever $M \in \mathbb{R}^{n \times m}$ we have

$$\begin{aligned}
\|M\|_{\infty \rightarrow 1} &= \max_{s \in \{\pm 1\}^n, t \in \{\pm 1\}^m} s^T M t \\
&= \max_{s \in \{\pm 1\}^n} \sum_j \left| \sum_i s_i M_{ij} \right| \\
&\geq \mathbb{E}_s \sum_j \left| \sum_i s_i M_{ij} \right| \\
&= \sum_j \mathbb{E}_s \left| \sum_i s_i M_{ij} \right| \\
&\geq \frac{1}{\sqrt{2}} \sum_j \sqrt{\sum_i M_{ij}^2} \\
&\geq \frac{1}{\sqrt{2n}} \sum_{ij} |M_{ij}| \\
&= \frac{1}{\sqrt{2n}} |M|_1.
\end{aligned}$$

This is nothing but the complementary upper bound to (3.63), and reads

$$\max_{0 \neq M \in \mathbb{R}^{n \times m}} \frac{|M|_{*\pi}}{|M|_{*\varepsilon}} = \max_{0 \neq M \in \mathbb{R}^{n \times m}} \frac{|M|_1}{\|M\|_{\infty \rightarrow 1}} \leq \sqrt{2n}. \quad (3.65)$$

The two inequalities (3.63) and (3.65) together solve the problem of data hiding in the cubic model. We summarise this solution as follows.

Proposition 3.26. *Let the composite of two cubic models $A = G_n$, $B = G_m$ be given by the minimal tensor product, $AB = G_n \otimes_{\min} G_m$. Then the restricted data hiding ratio $\bar{R}_{G_n \otimes_{\min} G_m}$ defined via Proposition 3.23 satisfies*

$$\sqrt{\frac{\min\{n, m\}}{2}} \leq \bar{R}_{G_n \otimes_{\min} G_m} \leq \sqrt{2 \min\{n, m\}}. \quad (3.66)$$

Consequently, all the data hiding ratios against locally constrained measurements scale as

$$R_{G_n \otimes_{\min} G_m}(\mathcal{M}) = \Theta\left(\sqrt{\min\{n, m\}}\right) = \Theta\left(\sqrt{\min\{d_A, d_B\}}\right), \quad (3.67)$$

for $\mathcal{M} = \text{LO}, \text{LOSR}, \text{LOCC}_{\rightarrow}, \text{LOCC}, \text{SEP}$.

3.5.2 Werner construction for symmetric models

A remarkable feature possessed by all the examples of GPTs we have seen so far, including the two that are undoubtedly the most physically relevant, i.e. classical probability theory (Subsection 2.3.1) and quantum theory (Subsection 2.3.2), is the existence of a wide group of symmetries, i.e. linear transformations sending states to states. In the classical case, these are just permutations of the entries of the probability vector, while in the quantum case a symmetry is any conjugation by a unitary matrix. As for the spherical and cubic models (Subsection 2.3.3 and Example 3.25), there are the natural actions of $SO(d-1)$ and of signed permutations on the last $d-1$ components of the state vector. We will show how these symmetries can be exploited in order to define a relevant class of bipartite states in a theory made of two copies of the same symmetric GPT. Let us start with the following definition.

Definition 3.27 (Completely symmetric models). *A GPT (V, C, u) is said to be completely symmetric if there is a compact group G and a representation $\zeta : G \rightarrow \mathcal{L}(V)$ that: (i) is such that ζ_g sends normalised states to normalised states, for all $g \in G$; and (ii) is irreducible and nontrivial on the preserved subspace $V_0 := \{x \in V : u(x) = 0\}$.*

Note. Throughout this subsection, we will denote the action of a functional φ on a vector x with $\varphi(x)$ rather than $\langle \varphi, x \rangle$. This is to avoid confusion with other scalar products we will encounter along the way.

We will see at the end of this Subsection that the main examples of GPTs we have considered so far (namely, classical and quantum theory, and spherical and cubic model) are in fact completely symmetric. However, for now we are interested in keeping the reasoning as abstract and general as possible. Thus, consider a completely symmetric GPT as in Definition 3.27. We start by listing some elementary consequences of the existence of a group symmetry. Let us remind the reader that if $\zeta : G \rightarrow \mathcal{L}(V)$ is a representation, then the dual space V^* is naturally endowed with the dual representation $\zeta^* : G \rightarrow \mathcal{L}(V^*)$ given by $\zeta_g^* = \zeta_{g^{-1}}^T$. With this notation, the fact that V_0 is preserved by the action of G can be written as $\zeta_g^* u = u$ for all $g \in G$. In other words, $\mathbb{R}u$ is a trivial component of V^* under the action of G through ζ^* . Since it is well-known that if V is real and G is compact then ζ and ζ^* are isomorphic, there must necessarily exist also a G -invariant vector $u_* \in V$. We deduce from the fact that V_0 is irreducible and nontrivial that $u_* \notin V_0$, so that we are free to rescale it in such a way that $u, (u_*) = 1$. Note that this fixes the decomposition of V into G -irreducible representations as

$$V = \mathbb{R}u_* \oplus V_0, \tag{3.68}$$

where the two pieces are non-isomorphic. Now, we proceed to show that u_* is in fact a state, as implied by property (i) in Definition 3.27. In order to do so, we use the Haar integral on G , denoted by $\int_G dg$ and whose existence is guaranteed by the compactness of G . For every state ω , we have that $u_* = \int_G dg \zeta_g \omega$,

because the right-hand side is G -invariant and normalised and therefore must coincide with u_* . This expression for u_* as a positive combination of states $\zeta_g \omega$ reveals that u_* is itself a state.

As is easy to see by referring to (3.68), the dual vector space V^* decomposes as

$$V^* = \mathbb{R}u \oplus V_0^* \quad (3.69)$$

under the action of G , where the first addend is a trivial representation and the second one is G -isomorphic to V_0 . Therefore, any G -isomorphism $\chi : V^* \rightarrow V$ must map u into a multiple of u_* and V_0^* into V_0 . If we think of χ as a tensor in $V \otimes V$, we can write $\chi = \alpha u_* \otimes u_* + \beta \mathcal{E} = \alpha U_* + \beta \mathcal{E}$, where we used the shorthand $U_* := u_* \otimes u_*$, and $\mathcal{E} \in V_0 \otimes V_0$ is (canonically identified with) a fixed G -isomorphism $V_0^* \rightarrow V_0$. Depending on the representation we choose, we can alternatively write $(\zeta_g \otimes \zeta_g)(\mathcal{E}) = \mathcal{E}$ for all $g \in G$ or $\mathcal{E} \zeta_g^* = \zeta_g \mathcal{E}$ for all $g \in G$. From this latter expression we see that $\mathcal{E}_* := (\mathcal{E}^{-1})^T : V \rightarrow V^*$ is also a G -isomorphism. As for the tensor $\mathcal{E}_* \in V_0^* \otimes V_0^*$, we will rephrase this as $(\zeta_g^* \otimes \zeta_g^*)(\mathcal{E}_*) = \mathcal{E}_*$ for all $g \in G$.

It is perhaps convenient to think of \mathcal{E} and \mathcal{E}_* also as scalar products $\langle \cdot, \cdot \rangle_{\mathcal{E}}$, $\langle \cdot, \cdot \rangle_{\mathcal{E}_*}$ on V_0^* and V_0 , respectively. This can be done via the definitions $\langle f, g \rangle_{\mathcal{E}} := (f \otimes g)(\mathcal{E})$ and $\langle v, w \rangle_{\mathcal{E}_*} := \mathcal{E}_*(v \otimes w)$. It is well known that it is possible to choose \mathcal{E} such that both these scalar products are positive definite. We will always make this assumption throughout the rest of this Section. If $\{f_i\}_{i=1}^{d-1}$ is an orthonormal basis for $\langle \cdot, \cdot \rangle_{\mathcal{E}}$, we will have

$$\mathcal{E} = \sum_{i=1}^{d-1} f_i^* \otimes f_i^* \in V_0^* \otimes V_0^*, \quad \mathcal{E}_* = \sum_{i=1}^{d-1} f_i \otimes f_i \in V_0 \otimes V_0, \quad (3.70)$$

with $\{f_i^*\}_i$ being the dual basis to $\{f_i\}_i$. As a simple consequence, we see that $\mathcal{E}_*(\mathcal{E}) = d - 1$, and the two norms induced by the above scalar products are dual to each other. Moreover, observe that the completion $\{u, f_1, \dots, f_d\}$ is an orthonormal basis for a global G -invariant scalar product. We are now ready to give the following definition.

Definition 3.28 (Werner states). *For a composite AB made of two copies A, B of the same completely symmetric GPT (V, C, u) and such that (2.18) is obeyed, a Werner state is a normalised state in $V \otimes V$ that corresponds to a G -isomorphism $V^* \rightarrow V$.*

With the language developed throughout the above discussion, we can express a generic Werner state as

$$\chi_s = U_* + s \mathcal{E}, \quad (3.71)$$

where s is a real parameter whose range depends on specific features of the model. For any bipartite cone C_{AB} satisfying (2.18), let us define

$$k_{\pm} := \max \{s : \chi_{\pm s} \in C_{AB}\}, \quad (3.72)$$

so that the allowed range of s in (3.71) will be $[-k_-, k_+]$. Besides the obvious observation that $k_{\pm} \geq 0$, there seems to be nothing we can say a priori about these parameters. In order to proceed further, we need a little lemma.

Lemma 3.29. *Let (V, C, u) be a completely symmetric GPT with Werner states given by (3.71). If $\mathcal{E}, \mathcal{E}_*$ are given by (3.70), the identity*

$$\int_G dg \zeta_g \otimes \zeta_g = U_* U + \frac{1}{d-1} \mathcal{E} \mathcal{E}_* \quad (3.73)$$

between linear operators on $V \otimes V$ holds true. Here, $U := u \otimes u$, and $u_* u : V \rightarrow V$ acts as $(u_* u)(x) = \langle u, x \rangle u_*$ for all $x \in V$.

Proof. For an arbitrary $Z \in V \otimes V$, the properties of the Haar measure guarantee that $\int_G dg \zeta_g \otimes \zeta_g(Z)$ is a G -invariant tensor, and thus can be expressed as $\alpha U_* + \beta \mathcal{E}$ for some $\alpha, \beta \in \mathbb{R}$. Applying U on both sides we see that

$$\begin{aligned} \alpha &= \int_G dg U (\zeta_g \otimes \zeta_g Z) \\ &= \int_G dg (\zeta_g^* \otimes \zeta_g^* U) (Z) \\ &= \int_G dg U(Z) \\ &= U(Z). \end{aligned}$$

Moreover, it is easy to verify that since $\mathcal{E}_*(\mathcal{E}) = d-1$ and $\zeta_g^* \otimes \zeta_g^* \mathcal{E}_* = \mathcal{E}_*$ for all $g \in G$, the analogous equality $(d-1)\beta = \mathcal{E}_*(Z)$ holds, thus completing the proof of (3.73). \square

With Lemma 3.29 in our hands, we can say a bit more about the constants k_{\pm} introduced in (3.72). Namely, we can determine lower bounds that correspond to the *separability region* for the family of Werner states (3.71). As expected, these lower bounds will depend only on the local structure of the model, not on the particular choice of the composite cone, which will instead affect the values of k_{\pm} .

Proposition 3.30. *For a completely symmetric GPT (V, C, u) of dimension d and a scalar product $\langle \cdot, \cdot \rangle_{\mathcal{E}_*}$ on the corresponding V_0 , introduce the quantities*

$$m_+ := \max \{ \langle v, v \rangle_{\mathcal{E}_*} : v \in V_0, u_* + v \in C \}, \quad (3.74)$$

$$m_- := \max \{ \langle v, w \rangle_{\mathcal{E}_*} : v, w \in V_0, u_* + v, u_* - w \in C \}. \quad (3.75)$$

Then the constants k_{\pm} defined through (3.72) satisfy $k_{\pm} \geq \frac{m_{\pm}}{d-1}$, and a Werner state χ_s as given in (3.71) is separable if and only if $-\frac{m_-}{d-1} \leq s \leq \frac{m_+}{d-1}$.

Proof. Take $v \in V_0$ such that $u_* + v \in C$ and $\langle v, v \rangle_{\mathcal{E}_*} = m_+$. Compute

$$\begin{aligned} & \int_G dg (\zeta_g \otimes \zeta_g) ((u_* + v) \otimes (u_* + v)) \\ &= \left(U_* U + \frac{1}{d-1} \mathcal{E} \mathcal{E}_* \right) ((u_* + v) \otimes (u_* + v)) \\ &= U_* + \frac{\mathcal{E}_*(v \otimes v)}{d-1} \mathcal{E} \\ &= U_* + \frac{\langle v, v \rangle_{\mathcal{E}_*}}{d-1} \mathcal{E} \\ &= U_* + \frac{m_+}{d-1} \mathcal{E}. \end{aligned}$$

From the leftmost side we see that this is an allowed separable state of the bipartite system. Looking at the rightmost side and comparing it to (3.71), we see that $k_+ \geq \frac{m_+}{d-1}$ and that $\chi_{m_+/(d-1)} \in C_A \otimes_{\min} C_B$. An analogous construction shows that we can find $v, w \in V_0$ such that $u_* + v, u_* - w \in C$ and

$$\int_G dg \zeta_g \otimes \zeta_g ((u_* + v) \otimes (u_* - w)) = U_* - \frac{m_-}{d-1} \mathcal{E} \in C_A \otimes_{\min} C_B,$$

so that $k_- \geq \frac{m_-}{d-1}$ and $\chi_{-k_-} \in C_A \otimes_{\min} C_B$.

In order to show that $\chi_s \in C_A \otimes_{\min} C_B$ if and only if $s \in \left[-\frac{m_-}{d-1}, \frac{m_+}{d-1} \right]$, we start by proving that $\omega \in C_A \otimes_{\min} C_B$ implies $-m_- \leq \mathcal{E}_*(\omega) \leq m_+$. In fact, if $\omega = \sum_i p_i (u_* + v_i) \otimes (u_* + w_i)$ we obtain $\mathcal{E}_*(\omega) = \sum_i p_i \langle v_i, w_i \rangle_{\mathcal{E}_*}$, and the claim follows from the inequalities

$$\langle v_i, w_i \rangle_{\mathcal{E}_*} = -\langle v_i, -w_i \rangle_{\mathcal{E}_*} \geq -m_-,$$

valid because $u_* + v_i, u_* - (-w_i) \in C$ and thus $\langle v_i, -w_i \rangle_{\mathcal{E}_*} \leq k_-$, and

$$\langle v_i, w_i \rangle_{\mathcal{E}_*} \leq \sqrt{\langle v_i, v_i \rangle_{\mathcal{E}_*} \langle w_i, w_i \rangle_{\mathcal{E}_*}} \leq m_+.$$

Applying this to a Werner state χ_s we obtain that a necessary condition for separability is $-m_- \leq \mathcal{E}_*(\chi_s) = (d-1)s \leq m_+$, concluding the proof. \square

In order to study the data hiding properties of the family of Werner states, we need to understand the separability conditions at the dual level. Luckily enough, there is no need to repeat any calculation. This is because if (V, C, u) is completely symmetric then (V^*, C^*, u_*) is itself a completely symmetric GPT. Therefore, we start by giving the following ‘dual’ definitions:

$$k_{\pm}^* := \max \{ s : U \pm s \mathcal{E}_* \in C_{AB}^* \}, \quad (3.76)$$

$$m_+^* := \max \{ \langle f, f \rangle_{\mathcal{E}} : f \in V_0^*, u + f \in C^* \}, \quad (3.77)$$

$$m_-^* := \max \{ \langle f, g \rangle_{\mathcal{E}} : f, g \in V_0^*, u + f, u - g \in C^* \}. \quad (3.78)$$

Exactly as in Proposition 3.30, we obtain

$$k_{\pm}^* \geq \frac{m_{\pm}^*}{d-1}. \quad (3.79)$$

Moreover, one can prove the following.

Lemma 3.31. *In a completely symmetric GPT (V, C, u) , the constants k_{\pm} , k_{\pm}^* , m_{\pm} , m_{\pm}^* defined by (3.72), (3.76), (3.74), (3.75), (3.77), (3.78), satisfy $m_- \leq m_+$, $m_-^* \leq m_+^*$, and moreover*

$$k_{\pm}^* k_{\mp} = \frac{1}{d-1}, \quad 1 \leq m_{\pm} m_{\mp}^* \leq d-1. \quad (3.80)$$

Proof. The inequalities $m_- \leq m_+$ and $m_-^* \leq m_+^*$ follow trivially from the definition by applying once Cauchy-Schwartz inequality. Thus, let us proceed to show the relation between k_{\pm} and k_{\pm}^* . By definition of dual cone, $U \pm s\mathcal{E}_* \in C_{AB}^*$ if and only if $(U + s\mathcal{E}_*)(\omega) \geq 0$ for all (normalised) $\omega \in C_{AB}$. However, because of the group symmetry it is enough to test Werner states. In fact, since $(\zeta_g^* \otimes \zeta_g^*)(U + s\mathcal{E}_*) = U + s\mathcal{E}_*$ for all $g \in G$, we obtain

$$\begin{aligned} (U + s\mathcal{E}_*)(\omega) &= \int_G dg ((\zeta_g^* \otimes \zeta_g^*)(U + s\mathcal{E}_*))(\omega) \\ &= \int_G dg (U + s\mathcal{E}_*)(\langle \zeta_g \otimes \zeta_g, \omega \rangle) \\ &= (U + s\mathcal{E}_*)(\chi_t) \\ &= 1 + (d-1)st, \end{aligned}$$

where $(d-1)t = \mathcal{E}_*(\omega)$. The rightmost side of the above equation is positive for all $\varphi \in [-k_-, k_+]$ if and only if $-\frac{1}{(d-1)k_+} \leq k \leq \frac{1}{(d-1)k_-}$.

Now, let us devote our attention to the bounds on the m quantities in (3.80). Clearly, it is enough to show that $1 \leq m_+ m_-^* \leq d-1$, up to exchanging primal and dual GPT. Consider $f \in V_0^*$ such that $\langle f, f \rangle_{\mathcal{E}} = 1$. Then it is easy to see that $u - \frac{f}{\sqrt{m_+}} \in C^*$, since for all normalised $\omega \in C$, that can be equivalently parametrised as $\omega = u_* + v$ with $v \in V_0$ satisfying $\langle v, v \rangle_{\mathcal{E}_*} = 1$, one has

$$\begin{aligned} \left(u \pm \frac{f}{\sqrt{m_+}} \right)(\omega) &= 1 \pm \frac{1}{\sqrt{m_+}} f(v) \\ &\geq 1 - \frac{1}{\sqrt{m_+}} \sqrt{\langle f, f \rangle_{\mathcal{E}} \langle v, v \rangle_{\mathcal{E}_*}} \\ &\geq 0. \end{aligned}$$

Since $u \pm \frac{f}{\sqrt{m_+}} \in C^*$, from the definition (3.78) we infer that

$$m_-^* \geq \left\langle \frac{f}{\sqrt{m_+}}, \frac{f}{\sqrt{m_+}} \right\rangle_{\mathcal{E}} = \frac{1}{m_+}.$$

In order to prove the upper bound $m_+m_-^* \leq d-1$, let us write

$$m_+ \leq (d-1)k_+ = \frac{1}{k_-^*} \leq \frac{d-1}{m_-^*},$$

where we employed in order: the results of Proposition 3.30, the first relation in (3.80) and finally (3.79). \square

With the tools we have developed so far, we are ready to discuss quantitatively the existence of allowed and separable measurements displaying Werner symmetry.

Proposition 3.32. *In a completely symmetric GPT (V, C, u) , for $\alpha, \beta \in \mathbb{R}$ the functionals $(\alpha U + \beta \mathcal{E}_*, (1-\alpha)U - \beta \mathcal{E}_*)$ form an allowed measurement if and only if*

$$-k_-^* \alpha \leq \beta \leq k_+ \alpha, \quad -k_+^* (1-\alpha) \leq \beta \leq k_-^* (1-\alpha). \quad (3.81)$$

In the (α, β) -plane, these two conditions identify a parallelogram with vertices

$$(0, 0), (1, 0), \left(\frac{k_-^*}{k_+^* + k_-^*}, \frac{k_+^* k_-^*}{k_+^* + k_-^*} \right), \left(\frac{k_+^*}{k_+^* + k_-^*}, -\frac{k_+^* k_-^*}{k_+^* + k_-^*} \right). \quad (3.82)$$

The separability conditions for the measurement under examination can be deduced from (3.81) by making the substitutions $k_\pm^* \mapsto \frac{m_\pm^*}{d-1}$.

Proof. We omit the details, since the proof consists in a systematic application of the definitions (3.76), together with the dual conditions to those already given in Proposition 3.30. \square

Equipped with Proposition 3.32, we are ready to explore the data hiding properties of Werner states. Since we did not make any general claim about anything but separability, we will compute the highest data hiding efficiency against separable measurements that is obtainable by using Werner states in Definition 3.5.

Theorem 3.33 (Data hiding with Werner states). *For a composite system made of two copies of the same completely symmetric GPT, the highest data hiding efficiency against separable measurements that is achievable with only Werner states is given by*

$$\begin{aligned} R_{\text{Werner}}(\text{SEP}) &= \max_{0 \neq (a,b) \in \mathbb{R}^2} \frac{\|aU_* + b\mathcal{E}\|}{\|aU_* + b\mathcal{E}\|_{\text{SEP}}} \\ &= 1 + \frac{2}{k_+ + k_-} \max \left\{ \frac{1}{m_+^*} - k_-, \frac{1}{m_-^*} - k_+ \right\}. \end{aligned} \quad (3.83)$$

Proof. Let us start by computing the base norm $\|aU_* + b\mathcal{E}\|$, for $a, b \in \mathbb{R}$. Thanks to Proposition 3.32, we have to test just one nontrivial measurement, namely

$$\left(\frac{k_-^*}{k_+^* + k_-^*} U + \frac{k_+^* k_-^*}{k_+^* + k_-^*} \mathcal{E}, \frac{k_+^*}{k_+^* + k_-^*} U - \frac{k_+^* k_-^*}{k_+^* + k_-^*} \mathcal{E} \right).$$

Thus, using (3.2) we find

$$\begin{aligned} \|aU_* + b\mathcal{E}\| &= \left| \left(\frac{k_-^*}{k_+^* + k_-^*} U + \frac{k_+^* k_-^*}{k_+^* + k_-^*} \mathcal{E} \right) (aU_* + b\mathcal{E}) \right| \\ &\quad + \left| \left(\frac{k_+^*}{k_+^* + k_-^*} U - \frac{k_+^* k_-^*}{k_+^* + k_-^*} \mathcal{E} \right) (aU_* + b\mathcal{E}) \right| \end{aligned} \quad (3.84)$$

$$\begin{aligned} &= \left| \frac{k_-^*}{k_+^* + k_-^*} a + \frac{k_+^* k_-^*}{k_+^* + k_-^*} (d-1)b \right| \\ &\quad + \left| \frac{k_+^*}{k_+^* + k_-^*} a - \frac{k_+^* k_-^*}{k_+^* + k_-^*} (d-1)b \right| \end{aligned} \quad (3.85)$$

$$= \max \left\{ |a|, \frac{|2(d-1)b k_+^* k_-^* + a(k_-^* - k_+^*)|}{k_+^* + k_-^*} \right\}, \quad (3.86)$$

where we used the elementary formula $|\alpha + \beta| + |\gamma - \beta| = \max\{|\alpha + \gamma|, |2\beta + \alpha - \gamma|\}$ in the last step. As suggested by Proposition 3.32, we can obtain the separability norm by replacing everywhere k_\pm^* with $\frac{m_\pm^*}{d-1}$:

$$\|aU_* + b\mathcal{E}\|_{\text{SEP}} = \max \left\{ |a|, \frac{|2b m_+^* m_-^* + a(m_-^* - m_+^*)|}{m_+^* + m_-^*} \right\}. \quad (3.87)$$

Observe that plugging (3.80) into (3.86) yields the somehow more handy expression

$$\|aU_* + b\mathcal{E}\| = \max \left\{ |a|, \frac{|2b + a(k_- - k_+)|}{k_+ + k_-} \right\}. \quad (3.88)$$

Now, computing the ratio $\max_{0 \neq (a,b) \in \mathbb{R}^2} \frac{\|aU_* + b\mathcal{E}\|}{\|aU_* + b\mathcal{E}\|_{\text{SEP}}}$ amounts to minimising (3.87) for a fixed value of (3.88). It is very easy to see that such a minimum is always achieved for pairs (a, b) such that the two expressions in the maximum appearing in (3.87) coincide. Up to scalar multiples, there are exactly two such pairs, namely those satisfying $b = \pm \frac{a}{m_\mp^*}$. By substituting these values into (3.88) and dividing by (3.87), we find

$$R_{\text{Werner}}(\text{SEP}) = \max_{0 \neq (a,b) \in \mathbb{R}^2} \frac{\|aU_* + b\mathcal{E}\|}{\|aU_* + b\mathcal{E}\|_{\text{SEP}}} \quad (3.89)$$

$$= \max \left\{ 1, \frac{\left| -\frac{2}{m_-^*} + k_- - k_+ \right|}{k_+ + k_-}, \frac{\left| -\frac{2}{m_+^*} + k_- - k_+ \right|}{k_+ + k_-} \right\}. \quad (3.90)$$

Now, using (3.80) it is easy to see that

$$\begin{aligned} \frac{2}{m_{\pm}^*} \pm (k_+ - k_-) &\geq \frac{2}{(d-1)k_{\pm}^*} \pm (k_+ - k_-) \\ &= 2k_{\mp} \pm (k_+ - k_-) \\ &= k_+ + k_- . \end{aligned}$$

Thanks to the above inequalities, we can further simplify (3.90) to

$$\begin{aligned} R_{\text{Werner}}(\text{SEP}) &= \max \left\{ \left| \frac{\frac{2}{m_-^*} + k_- - k_+}{k_+ + k_-} \right|, \left| \frac{-\frac{2}{m_+^*} + k_- - k_+}{k_+ + k_-} \right| \right\} = \\ &= \frac{1}{k_+ + k_-} \max \left\{ \frac{2}{m_+^*} + (k_+ - k_-), \frac{2}{m_-^*} - (k_+ - k_-) \right\} = \\ &= 1 + \frac{2}{k_+ + k_-} \max \left\{ \frac{1}{m_+^*} - k_-, \frac{1}{m_-^*} - k_+ \right\}, \end{aligned}$$

finally proving (3.83). \square

The above result shows how the maximal data hiding ratio obtainable by employing only Werner states depends on just few geometric parameters characterising the model under examination. The usefulness of this theorem rests on its immediate applicability to several natural classes of highly symmetric GPTs. Since computing the relevant parameters is often a simple and intuitive task, as we shall see in a moment, (3.83) gives a quick lower bound on all the data hiding ratios against locally constrained sets of measurements. To demonstrate the power of Theorem 3.33, we apply it to the symmetric models that we have examined so far: classical probability theory, quantum mechanics, spherical model, and cubic model.

- *Classical probability theory* (Subsection 2.3.1). By looking at the definition (2.27), it is easy to see that the relevant group here is the symmetric group S_d . We have $u = (1, \dots, 1)^T = d u_*$, and the action on the subspace $V_0 = \{x \in \mathbb{R}^d : \sum_i x_i = 0\}$ is well-known to be irreducible. This can be either proven with elementary tools or shown in one line via character theory and Burnside's formula. In fact, denote with $\eta : S_d \rightarrow \mathbb{R}^{d \times d}$ the standard representation of S_d on \mathbb{R}^d . Since we can exhibit an explicit one-dimensional irrep (vectors of constant entries), it is enough to show that $\frac{1}{d!} \sum_{\pi \in S_d} (\text{Tr } \eta(\pi))^2 = \langle \chi_\eta, \chi_\eta \rangle = 2$. Observing that $\text{Tr } \eta(\pi)$ is the number of elements fixed by π and applying Burnside's formula to the natural action of S_d on $\{1, \dots, d\}^{\times 2}$, which has exactly 2 orbits, we find $\frac{1}{d!} \sum_{\pi \in S_d} (\text{Tr } \eta(\pi))^2 = 2$.

Going back to the analysis of classical probability theory as a completely symmetric model, and exploiting the identification between elements in the tensor product and $d \times d$ real matrices we can also write $U = uu^T =$

d^2U_* and $\mathcal{E} = \mathbb{1} - \frac{1}{d}uu^T = \mathcal{E}_*$. Furthermore, observe that since the upper and lower bound in (2.18) coincide, the bipartite cone is composed of all the entrywise positive matrices, collectively denoted by $\mathbb{R}_+^{d \times d}$. Computing the relevant quantities is now elementary:

$$\begin{aligned} m_+^* &= \max \left\{ v^T v : u + v \in \mathbb{R}_+^d, \sum_i v_i = 0 \right\} = d(d-1) = d^2 m_+, \\ m_-^* &= \max \left\{ v^T w : u+v, u-w \in \mathbb{R}_+^d, \sum_i v_i = \sum_j w_j = 0 \right\} = d = d^2 m_-, \\ k_+ &= \max \left\{ s : u_* u_*^T + s(\mathbb{1} - du_* u_*^T) \in \mathbb{R}_+^{d \times d} \right\} = \frac{1}{d}, \\ k_- &= \max \left\{ s : u_* u_*^T - s(\mathbb{1} - du_* u_*^T) \in \mathbb{R}_+^{d \times d} \right\} = \frac{1}{d(d-1)}, \end{aligned}$$

Applying (3.83), we find $R_{\text{Werner}}^{\text{Cl}}(\text{SEP}) = 1$, which is expected since the collapse of the hierarchy (2.18) when either of the two cones is simplicial forbids the existence of data hiding altogether.

- *Quantum mechanics* (Subsection 2.3.2). We refer to (2.31) for the notation. The symmetry group in quantum mechanics is $U(n)$, the group of unitary $n \times n$ matrices. It is easy to see that we can choose $u = \mathbb{1} = n u_*$ and that the orthogonal complement to this trivial representation (that is, the space of traceless hermitian matrices) is irreducible. In fact, this follows already from the same result for the classical case, since permutation matrices are also unitaries.

For a bipartite system we see that $U = \mathbb{1} = n^2 U_*$ and $\mathcal{E} = F - \frac{\mathbb{1}}{n} = \mathcal{E}_*$, where $F|\alpha\beta\rangle = |\beta\alpha\rangle$ denotes the flip operator. If the composite is assembled according to the standard quantum mechanical rule (2.36), we find

$$\begin{aligned} m_+^* &= \max \left\{ \text{Tr } X^2 : \mathbb{1} + X \geq 0, \text{Tr } X = 0 \right\} = n(n-1) = n^2 m_+, \\ m_-^* &= \max \left\{ \text{Tr } XY : \mathbb{1} + X, \mathbb{1} - Y \geq 0, \text{Tr } X = \text{Tr } Y = 0 \right\} = n = n^2 m_-, \\ k_+ &= \max \left\{ s : \mathbb{1}/n^2 + s(F - \mathbb{1}/n) \geq 0 \right\} = \frac{1}{n(n+1)}, \\ k_- &= \max \left\{ s : \mathbb{1}/n^2 - s(F - \mathbb{1}/n) \geq 0 \right\} = \frac{1}{n(n-1)}, \end{aligned}$$

from which it follows easily $R_{\text{Werner}}^{\text{QM}}(\text{SEP}) = n$. According to Theorem 3.11, this is even the optimal data hiding ratio against all separable measurements.

If the composite is formed with the minimal tensor product rule (what we called ‘ W -theory’ in Definition 2.10, we write instead

$$\begin{aligned} k_+ &= \frac{m_+}{n^2 - 1} = \frac{1}{n(n+1)}, \\ k_- &= \frac{m_-}{n^2 - 1} = \frac{1}{n(n^2 - 1)}, \end{aligned}$$

and thus $R_{\text{Werner}}^{\text{W}}(\text{SEP}) = 2n - 1$, in accordance with Proposition 3.12.

- *Spherical model* (Subsection 2.3.3). Using the same notation as in (2.44), we see that in this case the relevant group is $O(d-1)$, the set of $(d-1) \times (d-1)$ orthogonal matrices acting on the last $d-1$ components of \mathbb{R}^d . We can choose $u = (1, 0, \dots, 0)^T = u_*$, and again the orthogonal complement is irreducible as follows from the same result for classical probability theory. For a bipartite system $U = uu^T = U_*$ and $\mathcal{E} = \widehat{\mathbb{1}}_{d-1} = \mathcal{E}_*$. If the composite is formed with the minimal tensor product rule, then it is very easy to verify that

$$\begin{aligned} m_{\pm}^* &= \max\{v^T v : u + v \in C_d\} = 1 = m_{\pm}, \\ k_{\pm} &= \frac{m_{\pm}}{d-1} = \frac{1}{d-1}, \end{aligned}$$

so that $R_{\text{Werner}}^{\text{Sph}}(\text{SEP}) = d - 1$, which coincides with the lower bound for the data hiding ratio against separable measurements given in Corollary 3.17.

- *Cubic model* (Example 3.25). We follow the notation previously established. The symmetry group of the cubic model in d dimensions is simply the symmetry group of the $(d-1)$ -dimensional hypercube. As for the spherical model, we have $u = (1, 0, \dots, 0)^T = u_*$, $U = uu^T = U_*$ and u^{\perp} is irreducible. Furthermore, we can choose $\mathcal{E} = \widehat{\mathbb{1}}_{d-1} = \mathcal{E}_*$. However, even for a minimal tensor product composite we have

$$\begin{aligned} m_{\pm}^* &= \max\{v^T v : v \in \mathbb{R}^{d-1}, |v|_1 \leq 1\} = 1, \\ m_{\pm} &= \max\{v^T v : v \in \mathbb{R}^{d-1}, |v|_{\infty} \leq 1\} = d - 1, \\ k_{\pm} &= \frac{m_{\pm}}{d-1} = 1, \end{aligned}$$

and therefore $R_{\text{Werner}}^{\text{G}}(\text{SEP}) = 1$. This example shows how it might be the case that Werner states do not display any data hiding property, despite the fact that there is global data hiding in the cubic model, as shown by Proposition 3.26.

3.6 Conclusions

Throughout this chapter, we have presented a general theory of data hiding in bipartite systems composed of general probabilistic theories. It significantly generalises ideas and results from quantum mechanics; in particular, we were able to determine the maximum so-called data hiding ratio in terms of the state space dimension, by making a connection between this problem and Grothendieck's tensor norms. This maximum is essentially attained for GPTs over spherical cones.

Inspired by the prominent role played by Werner states in quantum mechanical data hiding, we investigated Werner-like states in classes of theories

with symmetric state spaces. By using these states as ansatzes, we could find a general lower bound on the data hiding ratio that depends on few geometrically meaningful parameters.

For quantum mechanics on finite-dimensional spaces we proved a new upper bound on the data hiding ratio against LOCC operations by exploiting the celebrated quantum teleportation protocol. We saw how this improves previous results and definitively settles the problem of determining the optimal data hiding ratio against LOCC for fixed local dimensions. However, the same problem for the smaller set of local operations remains open, as we showed that none of the bounds that we provide or that are available in the literature is generally optimal.

Perhaps surprisingly, data hiding ratios in quantum mechanics are not as large as the maximal ones, being of the order of the square root of the latter, which are exhibited by spherical cones. This answers the general question we posed in Subsection 2.1.1, point (iii), for the particular example of non-classical phenomenon known as data hiding.

If one were to summarise the results of our research in one single sentence, one could say that the laws of Nature, which move the sun and the other stars, are intrinsically non-classical, but not as non-classical as they could have been.

Part II

Some aspects of quantum entanglement

Chapter 4

Bipartite depolarising maps

4.1 Introduction

We now leave behind the technically fierce world of general probabilistic theories, and let our modest investigation set course for the more kindly and familiar waters of quantum information theory. Namely, while most of Chapter 2 has been devoted to discussing the concept of entanglement in the GPT framework, in this second part of the thesis we focus instead on some aspects of quantum entanglement. In this chapter, in particular, we address the problem of characterising the entanglement transformation properties of a class of quantum channels acting on a bipartite system. The maps we chose to analyse, which generalise the well-known depolarising channel [118], are good candidates to model simple forms of noise acting on quantum devices.

The present chapter is organised as follows. The rest of this section is devoted to discussing the main problem and the motivation behind it (Subsection 4.1.1) and to highlighting our original contributions (Subsection 4.1.2). After reviewing in Section 4.2 some basic concepts about notable classes of maps in quantum information, with special attention devoted to the depolarising channel and to its generalisation, we move on to the characterisation of the properties of the maps we introduce. Section 4.3 is dedicated to positivity, the most basic requirement for maps that are relevant in quantum information. Section 4.4 is then devoted to complete positivity, while the main goal of Section 4.5 is to find necessary and sufficient conditions under which the maps under examination are entanglement-breaking. The same problem but with entanglement annihilation is the subject of Section 4.6, where we present what is probably the main result of the chapter, Theorem 4.10, and discuss its applications to the solution of some open problems recently raised in [12, 13]. Finally, in Section 4.7 we draw the conclusions and discuss some future perspectives.

4.1.1 The problem

In quantum information science, the notion of entanglement plays a major role from both the foundational and the operational point of view [92, 109], and

thus its study constitutes one of the cornerstones of the field. Entanglement is however a very fragile resource and once distributed, the local participants in any kind of communication protocol can never increase it via local operations, even if assisted by unlimited classical communication (i.e. with so-called LOCC protocols, see Subsection 2.2.4). Much effort has thus been invested into characterising the resource of entanglement [92, 182].

This programme can be declined in several ways. For example, the problems have been considered, of experimentally detecting [182] or quantifying entanglement from an operationally meaningful point of view [182]. Another related question concerns the transformation of entanglement. Naturally, the importance of this problem stems primarily from physics. In fact, on the one hand we want to manipulate our resource by implementing any given transformation that is physically accessible, in order to make it usable for running quantum information protocols. On the other hand, once we have some amount of entanglement stored in our quantum device, we should also be able to counter the deleterious effects of any spurious interaction with the external world. The very basic question that we tackle in this chapter is the following: *given a model of the noise acting on our entangled state, how long do we have to wait for the entanglement to be completely lost?* More loosely speaking, *how strong does the noise have to be in order to destroy the initial entanglement?* Here I am taking a passive viewpoint, but of course equivalent questions can be asked in a dynamical setting, where one sends states through physical channels (such as optic fibres) instead of being subjected to a harmful external noise.

Of course, we are going to answer this question in a very special case, i.e. for a very special model of noise. The reason behind this restriction will become clearer soon. However, the reason why we should impose *some kind* of restriction, that is, why we can not solve the problem in a closed form in full generality, can already be appreciated. In fact, as we mentioned in Section 2.4, the problem of deciding whether a given quantum state exhibits entanglement (called the *separability problem*) is in general computationally NP-hard (under certain assumptions on the approximation error) [114]. This precludes a closed solution of the above problem in full generality, but nonetheless some paradigmatic cases can be addressed in a practically useful way. Complete solutions for the separability problem indeed exist for various special classes of states [91, 117, 118, 130, 183, 184, 185, 186, 187, 188]. A central tool for addressing these problems are positive maps, the canonical example being transposition, from which the probably most widely used entanglement test derives [116].

Let us now delve a bit more into the mathematical formulation of the aforementioned questions. In general, under certain physically reasonable assumptions [109] any physical transformation of a quantum system can be represented by a completely positive (CP) map. Studying which CP maps preserve and more importantly which maps adversely affect entanglement is therefore our main goal here. Two important examples of deleterious action are constituted by: (i) maps that act on a bipartite system as a whole and break any entanglement between the two parties, called *entanglement-annihilating* [189]; and

(ii) maps that act on a single quantum system and the rest of the world, called *entanglement-breaking* [190].

In this chapter, we tackle the above problem for a special class of maps acting on a bipartite system that are suitable generalisations of *depolarising channels* [191, 192, 193, 194]. The one-parameter class of depolarising channels provides undoubtedly some of the most physically significant examples of quantum channels, its importance ultimately stemming from the fact that it models the process of addition of white noise to the input system. At the same time, its mathematically simple structure allows to express in closed form the solution to many problems [118, 194]. Indeed, the depolarising channel Δ_λ acting on an n -level quantum system can be expressed as

$$\Delta_\lambda := \lambda I + (1 - \lambda) \frac{\mathbb{1}}{n} \text{Tr}, \quad (4.1)$$

meaning that its action on an input density matrix ρ is given by $\Delta_\lambda(\rho) := \lambda\rho + (1 - \lambda) \frac{\mathbb{1}}{n}$. More details will be provided in Subsection 4.2.2.

4.1.2 Our contributions

The material covered in this chapter is taken from the homonymous paper [195]:

- L. Lami and M. Huber. Bipartite depolarizing maps. *J. Math. Phys.*, 57(9):092201, 2016.

In this chapter, we address the problems discussed in the preceding Subsection 4.1.1 for a specific class of maps, that are chosen as suitable generalisations of those in (4.1). Since we want them to act on a bipartite system AB and to contain the same terms as in (4.1), they will be most naturally expressed as a linear combination of four elementary maps of the form $\mathcal{N}_A^i \otimes \mathcal{N}_B^j$ ($i, j = 0, 1$), where \mathcal{N}^i can be either the identity channel I ($i = 0$) or the completely depolarising map $\mathbb{1} \text{Tr}$ ($i = 1$). Therefore, we will consider the family of maps

$$\Gamma[\alpha, \beta, \gamma] := \mathbb{1}_{AB} \text{Tr}_{AB} + \alpha \mathbb{1}_A \text{Tr}_A \otimes I_B + \beta I_A \otimes \mathbb{1}_B \text{Tr}_B + \gamma I_{AB}. \quad (4.2)$$

Note. In the original paper [195], we use the letter Φ instead of Γ for the maps in (4.2). This naturally creates a potential conflict with the symbol for a maximally entangled state, which there is taken to be ε . Here, we chose to stick to the more customary convention of reserving the letter Φ for the maximally entangled state, and thus chose Γ for the bipartite depolarising maps in (4.2).

It is immediately obvious that maps of the form (4.2) generalise those in (4.1) in the case of a bipartite system. However, the main motivation for studying this class comes from the fact that it can accommodate a variety of models for the noise acting on a system. As a first and most obvious example, consider the case of a the two shares of an entangled state being stored in a single device, which is then subjected to white noise of the form (4.1). In our case, such noise takes the form $(\Delta_\lambda)_{AB} \propto \Gamma \left[0, 0, \frac{n_A n_B \lambda}{1 - \lambda} \right]$. The second

extreme example we present corresponds to the case of the two subsystems being sent through different depolarising channels or subjected to different local white noises in spatially separated storing devices. The relevant channel becomes $(\Delta_\lambda)_A \otimes (\Delta_\mu)_B \propto \Gamma \left[\frac{n_B \mu}{1-\mu}, \frac{n_A \lambda}{1-\lambda}, \frac{n_A n_B \lambda \mu}{(1-\lambda)(1-\mu)} \right]$, thus falls again within the class (4.2).

Despite the richer structure of the maps Γ in (4.2), in this chapter we determine exactly the parameter regions for which these maps are: *positive* (Theorem 4.4); *completely positive* (Theorem 4.7); *entanglement-breaking* (Theorem 4.9) and *entanglement-annihilating* (Theorem 4.10). The results on complete positivity and entanglement-breaking region are extensions of those in [196, 197]. The full characterisation we give here was not available before, but a modified version of the argument in [196] can cover the case $n_A = n_B$. In [197], a class of states is considered, that intersects the Choi states of the maps (4.2) in a one-dimensional subspace, for all n_A, n_B .

On the contrary, the characterisation of the entanglement-annihilating region for the maps Γ (Theorem 4.10) is fully original. Besides its physical significance, our initial motivation for tackling this question was to answer the main open problems in [12, 13]. There, the problem of determining all pairs λ, μ for which the local depolarising noise $\Delta_\lambda \otimes \Delta_\mu$ is entanglement-annihilating for the system AB was considered. However, the bounds established in [12, 13] are not strong enough to solve the problem completely. Instead, an application of Theorem 4.10 yields the desired characterisation (Corollary 4.11), showing in particular that a product of depolarising maps is entanglement-annihilating iff it is (positive and) PPT-inducing.

As by-products of the proofs, some minor results are found, and some interesting techniques are devised and applied to the quantum separability problem. An important tool in our analysis is in fact the observation that separable matrices are closed with respect to the so-called *Hadamard* (or entry-wise) *product* [198, 199], essentially because the latter can be stochastically implemented via local operations and classical communication (LOCC). We believe that this can be a useful technique in entanglement theory, so in Subsection 4.2.4 we take a small detour to explore some of its properties. Along the way to the main results of the chapter, we stumble upon some small results previously found by others. For instance, we retrieve a fairly simple example of a PPT entangled state (4.24) living in a system $AB : A'B'$ that was originally discovered in [200], and make use of a positive indecomposable map (4.25) acting on a bipartite system AB , which is an instance of a more general construction presented in [197, §IX].

4.2 Preliminaries

4.2.1 Classes of maps in quantum information

As discussed in Subsection 2.3.2, states of an n -level quantum system are represented by $n \times n$ positive semidefinite matrices ρ with normalised trace (i.e.

subjected to the constraint $\text{Tr } \rho = 1$), generically called **density matrices**. The real span of the set of density matrices is the set of $n \times n$ Hermitian matrices \mathcal{H}_n . The extremal points of the set of quantum states are rank-one orthogonal projectors $|\psi\rangle\langle\psi|$, called **pure states**. Joining two quantum systems A and B of n_A and n_B levels, respectively, yields another quantum system of $n_A n_B$ levels. Therefore, bipartite density matrices are naturally thought of as living in $\mathcal{H}_{n_A} \otimes \mathcal{H}_{n_B} \simeq \mathcal{H}_{n_A n_B}$. Often, we will say for short that AB is an $n_A \times n_B$ quantum system.

Bipartite pure states $|\psi\rangle_{AB}$ admit a remarkable decomposition in terms of local bases. Namely, using the notation $n := \min\{n_A, n_B\}$, one can find orthonormal vectors $|e_1\rangle_A, \dots, |e_n\rangle_A$ and $|f_1\rangle_B, \dots, |f_n\rangle_B$ on the first and second subsystem, respectively, and non-negative scalars $\lambda_1, \dots, \lambda_n$, such that

$$|\psi\rangle_{AB} = \sum_{i=1}^n \sqrt{\lambda_i} |e_i\rangle_A |f_i\rangle_B, \quad (4.3)$$

a representation known as **Schmidt decomposition**. Observe that if $|\psi\rangle$ is appropriately normalised then $\sum_{i=1}^n \lambda_i = 1$.

In this chapter, we are mainly interested in linear maps acting on density matrices, or equivalently on \mathcal{H}_n . The reason why this is so, remember, is that we care about transformations undergone by quantum states when some external noise is acting on them, either because they are stored in a non-isolated environment, or because they are sent through an imperfect channel to a distant agent. For a complete review of the topic, we refer the reader to the classic textbook [109, VIII]. A very natural requirement to impose on a map that aims to represent a physical transformation is positivity. A map $\Lambda : \mathcal{H}_n \rightarrow \mathcal{H}_m$ is called **positive** if $\Lambda(\rho) \geq 0$ whenever $\rho \geq 0$ (where $X \geq 0$ stands for ‘ X is positive semidefinite’). This is a special instance of the definition we gave in Subsection 2.4.1 for general maps between ordered vector spaces.

As it turns out, this requirement is not strong enough to ensure that Λ is physically implementable. In fact, we need to make sure that also the partial application of Λ on one share of a bipartite system in a state ρ , that we represent as $(\Lambda \otimes I)(\rho)$, yields a physical state. This is the case iff $\Lambda \otimes I_k$ is a positive map for all integers k , where I_k stands for the identity map acting on \mathcal{H}_k . A map satisfying the above requirement is usually called **completely positive** (CP), a concept we already defined in more general terms in Subsection 2.2.4. This is a strictly stronger requirement than positivity, a classic example showing this being the transposition map $T : \mathcal{H}_n \rightarrow \mathcal{H}_n$ that acts as $T(X) := X^T$. The transposition operation plays a relevant role in the theory, so relevant that it deserves some dedicated nomenclature: a map Λ is called **completely copositive** (coCP) if $T\Lambda$ (equivalently, ΛT) is completely positive.

Despite sounding a bit less natural at first glance, complete positivity is actually much more easily checkable than positivity, since one can show that a map $\Lambda : \mathcal{H}_n \rightarrow \mathcal{H}_m$ is completely positive iff its **Choi state**

$$R_\Lambda := (\Lambda \otimes I)(\Phi) \quad (4.4)$$

is positive semidefinite, a statement known as **Choi's theorem** [201]. Here, Φ denotes the rank-one projector onto the **maximally entangled state**

$$|\Phi\rangle := \frac{1}{\sqrt{n}} \sum_{i=1}^n |ii\rangle. \quad (4.5)$$

As an easy corollary, one sees that $\Lambda \otimes I_k$ is positive for all integers k iff it is positive for $k = n$ (the dimension of the input Hilbert space of Λ). The correspondence between maps and states expressed by (4.4) is usually called **Choi-Jamiolkowski isomorphism**.

Usually, a completely positive map Λ that is also **trace-preserving**, meaning that $\text{Tr } \Lambda X = \text{Tr } X$ for all Hermitian X , is called a **quantum channel**. Physically, quantum channels are exactly the state transformations that one obtains by letting the input system interact with an ancillary system prepared in an fixed state for some time (so that system and ancilla together undergo a unitary evolution), and by discarding the ancilla afterwards. This is known as **Stinespring's dilation theorem** [202].

Despite not representing directly any physical transformation, positive maps constitute a valuable tool for entanglement detection. We remind the reader the definition given in Subsection 2.2.3 and (for the quantum case) in Subsection 2.3.2: a state ρ_{AB} is called **separable** [91] if there are collections of local states σ_A^i and τ_B^i and probabilities p_i such that

$$\rho_{AB} = \sum_i p_i \sigma_A^i \otimes \tau_B^i, \quad (4.6)$$

and **entangled** otherwise. We already defined the cone of unnormalised quantum separable states in (2.34). As we saw in Subsection 2.4.1, already in [117] it was proven that a bipartite quantum state ρ is separable iff $(\Lambda \otimes I)(\rho) \geq 0$ for all positive maps Λ , a statement that in Subsection 2.4.1 we called – following the terminology of [124] – **Woronowicz criterion**. As the reader may remember, an equivalent formulation of the criterion involves observables instead of maps: ρ is separable iff $\text{Tr } \rho W \geq 0$ for all $W \in \mathcal{H}_{n_A n_B}$ that are **entanglement witnesses**, i.e. are such that $\langle \psi\varphi | W | \psi\varphi \rangle \geq 0$ for all pure states $|\psi\rangle, |\varphi\rangle$ of the local subsystems (Subsection 2.3.2). This reformulation is made possible by the remarkable fact that *entanglement witnesses are precisely the Choi states of positive maps*.

A prominent example of a positive map used as a test for entanglement is the transposition. The resulting criterion, called **PPT criterion** [116], states that a separable state ρ_{AB} is such that $\rho_{AB}^{T_A} = (T_A \otimes I_B)(\rho_{AB}) \geq 0$. This test is so strong that it turns out to be even *sufficient* to ensure separability in low-dimensional systems, namely when $n_{AB} \leq 6$ [117]. In higher dimensions this is no longer the case, and one can find examples of PPT entangled states.

Of particular relevance here are the properties of maps with respect to the entanglement of the input states. A map $\Lambda : \mathcal{H}_n \rightarrow \mathcal{H}_m$ is called **entanglement-breaking** [190] if it outputs a separable state whenever it acts on a share of a bipartite system prepared in *any* initial state, i.e. if $(\Lambda \otimes I_k)(\rho)$ is separable for

all states ρ of all bipartite $n \times k$ quantum systems. Clearly, an entanglement-breaking map Λ is in particular completely positive, and even more, has the property that $\Lambda' \Lambda$ is completely positive for all positive maps Λ' . As shown in [190], a map Λ is entanglement-breaking iff its Choi state (4.4) is separable. This allows for an operational characterisation of entanglement-breaking maps as those obtained by measuring and re-preparing the input state.

Since we learnt that PPT-ness is a mathematically convenient relaxation of the more elusive concept of separability, we can accordingly relax the notion of entanglement-breaking to that of **PPT-binding** map. This is defined as a map Λ such that $(\Lambda \otimes I_k)(\rho)$ is PPT (instead of separable) for all bipartite input states ρ of all bipartite $n \times k$ quantum systems. As is easy to check thanks to Choi's theorem, this happens iff the Choi state R_Λ is PPT.

If the input system is already bipartite, one could look at the entanglement between the two subsystems that compose it, instead of focusing on that between the system and the rest of the world. This attitude leads to the definition of **entanglement-annihilating** maps [189], which are maps acting on some bipartite system AB and characterised by the property that $\Lambda_{AB}(\rho_{AB})$ is separable for all density matrices ρ_{AB} . Although it is sufficient to check pure input states $\rho_{AB} = |\psi\rangle\langle\psi|_{AB}$, a more compact criterion on the same lines as that for entanglement-breaking maps is in general not available. By virtue of the above definitions, one sees that a map Λ is entanglement-breaking iff $\Lambda \otimes I_k$ is entanglement-annihilating for all k .

We can use again the PPT test to obtain a mathematically convenient relaxation of the entanglement-annihilation condition. Namely, if Λ_{AB} is entanglement-annihilating, then $(T_A \otimes I_B) \Lambda_{AB}$ is a positive map, with T_A denoting partial transposition on the first subsystem. Maps for which the latter condition holds are called **PPT-inducing** in [13].

An elementary observation is that positive, completely positive, entanglement-breaking, PPT-binding, entanglement-annihilating and PPT-inducing maps form convex sets.

4.2.2 Depolarising maps

The **depolarising channel** constitutes one of the simplest yet most physically significant examples of quantum channel, and has been the subject of extensive study [191, 192, 193, 194]. For a real parameter λ , the channel Δ_λ acting on an n -dimensional system is defined by the formula (4.1). We remind the reader [118] that the map Δ_λ is:

- positive iff $-\frac{1}{n-1} \leq \lambda \leq 1$;
- completely positive iff $-\frac{1}{n^2-1} \leq \lambda \leq 1$;
- completely copositive iff $-\frac{1}{n-1} \leq \lambda \leq \frac{1}{n+1}$;
- entanglement-breaking iff it is at the same time CP and coCP, iff $-\frac{1}{n^2-1} \leq \lambda \leq \frac{1}{n+1}$.

In particular, it is useful to remember that whenever Δ_λ is positive it is also either completely positive or completely copositive (or both). Another notable consequence is that $\mathbb{1}\text{Tr} - I \propto \Delta_{-1/(n-1)}$ is a completely copositive map, i.e. that $\mathbb{1}\text{Tr} - T$ (where T denotes transposition) is completely positive. Indeed, its Choi state is proportional to the projector onto the antisymmetric subspace of $\mathbb{C}^n \otimes \mathbb{C}^n$. These facts are used several times in the rest of the chapter.

4.2.3 Bipartite depolarising maps

The main subject of study in the present chapter are the ‘bipartite depolarising maps’ defined in (4.2). As a first remark, one could wish to consider a more general linear combination than (4.2), with another parameter for the coefficient of $\mathbb{1}_{AB}\text{Tr}_{AB}$. However, already the positivity condition easily implies that such a parameter must be non-negative. Up to a normalisation constant, one can take it to be either 1 or 0. The latter case can be deduced from the former, because $\alpha \mathbb{1}_A \text{Tr}_A \otimes I_B + \beta I_A \otimes \mathbb{1}_B \text{Tr}_B + \gamma I_{AB}$ is positive iff $\frac{1}{M} \mathbb{1}_{AB} \text{Tr}_{AB} + \alpha \mathbb{1}_A \text{Tr}_A \otimes I_B + \beta I_A \otimes \mathbb{1}_B \text{Tr}_B + \gamma I_{AB}$ is positive for all $M > 0$, which is in turn equivalent to $\Gamma[M\alpha, M\beta, M\gamma]$ being positive for all $M > 0$.

Observe that the maps Γ we defined commute with all the local unitary conjugations, i.e. with all maps of the form $\rho_{AB} \mapsto U_A \otimes V_B \rho_{AB} U_A^\dagger \otimes V_B^\dagger$, for all local unitaries U_A, V_B . This is the same as to say that the corresponding Choi states are left invariant by conjugations by $U_A \otimes V_B \otimes U_{A'}^* \otimes V_{B'}^*$. But there is more: up to including a coefficient for the first addend (which is an irrelevant modification in the sense clarified above), these maps are *all* the maps on AB displaying this feature, as the following reasoning shows. Before going into the details, we remind the reader that the states of a bipartite system AA' commuting with all the conjugations by local unitaries of the form $U_A \otimes U_{A'}^*$ are exactly the linear combinations of the identity $\mathbb{1}_{AA'}$ and the maximally entangled state $\Phi_{AA'}$ [118]. Now, up the application of the Choi-Jamiolkowski isomorphism, we have to prove that the set of $ABA'B'$ states that are left invariant by conjugations by $U_A \otimes V_B \otimes U_{A'}^* \otimes V_{B'}^*$ coincides with the set of linear combinations of $\mathbb{1}_{ABA'B'}$, $\mathbb{1}_{AA'} \otimes \Phi_{BB'}$, $\Phi_{AA'} \otimes \mathbb{1}_{BB'}$, and $\Phi_{AA'} \otimes \Phi_{BB'}$, i.e. with the tensor product of the two locally invariant subspaces of AA' and BB . This is nothing but a particular instance of a more general phenomenon: given two representations of compact groups $\zeta_i : G_i \rightarrow GL(V_i)$ (with $i = 1, 2$), the invariant subspace of the *external* tensor product $\zeta_1 \boxtimes \zeta_2$ is the tensor product of the two ζ_i -invariant subspaces; in fact, the latter is trivially contained in the former, while at the same time the dimensions are equal, thanks to character theory:

$$\begin{aligned} \langle \chi_1, \chi_{\zeta_1 \boxtimes \zeta_2} \rangle &= \int dg_1 dg_2 \text{Tr} [(\zeta_1 \otimes \zeta_2)(g_1, g_2)] \\ &= \int dg_1 dg_2 \text{Tr} \zeta_1(g_1) \text{Tr} \zeta_2(g_2) \\ &= \langle \chi_1, \chi_{\zeta_1} \rangle \langle \chi_1, \chi_{\zeta_2} \rangle. \end{aligned}$$

In our case, the two groups are $G_1 = SU(n_A)$ and $G_2 = SU(n_B)$, the two spaces are the set of hermitian operators on AA' and BB' , while the representations are defined through $\zeta_1(U)(X_{AA'}) = U_A \otimes U_{A'}^* X_{AA'} U_A^\dagger \otimes U_{A'}^T$ and analogously for BB' .

Note that for $n_A = n_B$ the separability properties of the Choi states corresponding to (4.2) have been already considered in [203]. In that paper, the entanglement-breaking conditions for (4.2) are found under the above simplifying assumption $n_A = n_B$. However, we will see that the general case $n_A \neq n_B$ is in a sense more interesting, because new phenomena such as the existence of PPT entangled states appear, as already observed in [197]. A family similar to the Choi states of (4.2) but possessing $U_A \otimes V_B \otimes U_{A'} \otimes V_{B'}^*$ instead of $U_A \otimes V_B \otimes U_{A'}^* \otimes V_{B'}^*$ symmetry was an important tool in [196]. Translating the results of that paper on the channel level would produce straightforwardly another 3-parameter class of maps for which necessary and sufficient entanglement-breaking conditions are available. As usual, we will denote by n the minimum between the two local dimensions n_A, n_B , i.e. the maximum Schmidt rank of a global pure state. A trick that turns out to be useful in analysing the above $\Gamma[\alpha, \beta, \gamma]$ requires the construction of the following family of maps acting on an n -dimensional system:

$$\chi[a, c] := \mathbb{1}_n \text{Tr} + aD + cI, \quad (4.7)$$

where D is the projection onto the diagonal, i.e. $D(X) = \mathbb{1} \circ X$ and the symbol \circ denotes the Hadamard or entry-wise product, defined by $(X \circ Y)_{ij} := X_{ij}Y_{ij}$.

4.2.4 Entanglement properties of the Hadamard product

Now that we mentioned it, let us take a small detour to explore some of the properties exhibited by the Hadamard product in relation to the separability problem. For an introduction to other basic properties, we refer the reader to [199, Chapter 5]. The main feature of this seemingly bizarre operation is that the cone of positive matrices is closed with respect to it; in other words, if $M \geq 0$ and $N \geq 0$ then $M \circ N \geq 0$ (this is usually called *Schur product theorem*). Actually, it is easily verified by means of Choi's theorem that for a given $M \geq 0$ the corresponding *Hadamard channel* $\phi_M(\cdot) := M \circ (\cdot)$ is not only positive but even completely positive.

Thanks to the Schur product theorem, given two quantum states ρ, σ , it makes sense to consider the (unnormalised) state $\rho \circ \sigma$. We stress that the definition of Hadamard product is explicitly dependent on the basis we choose to represent the operators ρ, σ as matrices. As it turns out, if ρ_{AB}, σ_{AB} are bipartite states and we fix a local basis to represent them (as we will always do from now on), then the Hadamard product behaves well with respect to the entanglement properties of the input states. Namely, the following simple yet very useful lemma holds.

Lemma 4.1. *Let ρ_{AB}, σ_{AB} be two bipartite density matrices represented with respect to local bases. Then:*

(a) $\rho_{AB} \circ \sigma_{AB}$ can be obtained stochastically from $\rho_{AB} \otimes \sigma_{A'B'}$ through a local measurement on the bipartite system $AA' : BB'$;

(b) in particular, if ρ_{AB}, σ_{AB} are separable, the same is true for $\rho_{AB} \circ \sigma_{AB}$.

Proof. Clearly, claim (b) follows straightforwardly once (a) has been established. In order to prove (a), it suffices to write

$$\rho_{AB} \circ \sigma_{AB} = \Pi_{AA'}^0 \otimes \Pi_{BB'}^0 \rho_{AB} \otimes \sigma_{A'B'} \Pi_{AA'}^0 \otimes \Pi_{BB'}^0, \quad (4.8)$$

here $\Pi_{AA'}^0 = \sum_{i=1}^{n_A} |i_A\rangle\langle i_{A'}|$ and analogously for $\Pi_{BB'}^0$. \square

Lemma 4.1(b) is a useful trick up our sleeve. This is all we will use in the rest of the chapter, particularly in the proof of Theorem 4.10. However, let us take some time to proceed a bit further.

As it turns out, there are examples of weakly entangled states that can be (stochastically) efficiently distilled by taking Hadamard powers. This holds also in the multipartite setting. For instance, the computation described in [204] (together with some later results from [187]) shows the existence of $(n-2)$ -separable n -qubit states having a genuinely multipartite entangled Hadamard square.

As a generalisation of the separability-preserving property of the Hadamard product, let us analyse the behaviour of the Schmidt rank function $SR(\cdot)$ defined in [205]. Let us remind the reader that the Schmidt rank of a pure state $|\psi\rangle_{AB}$ of a bipartite $n_A \times n_B$ system AB is the minimum k such that one can write

$$|\psi\rangle_{AB} = \sum_{i=1}^k |v_i\rangle |w_i\rangle$$

for appropriate local vectors $|v_i\rangle, |w_i\rangle$. The Schmidt rank (or Schmidt number) of a bipartite state ρ_{AB} is then defined as the minimum k such that ρ_{AB} can be written as a convex combination of bipartite pure states of Schmidt rank no larger than k .

Clearly, (4.8) implies that

$$SR(\rho_{AB} \circ \sigma_{AB}) \leq SR_{AA':BB'}(\rho_{AB} \otimes \sigma_{A'B'}) = SR(\rho_{AB})SR(\sigma_{AB}). \quad (4.9)$$

Now, we proceed to show that this bound can always be saturated, as long as it does not conflict with the trivial requirements $SR(\omega_{AB}) \leq \min\{n_A, n_B\} =: n$.

Before coming to the details, we remind the reader of some well-known facts.

- (i) The rank of the $n \times m$ Vandermonde matrix $M_{ij} = \lambda_j^i$ is always equal to $\min\{n, k\}$, where k is the number of distinct elements among the complex numbers $\lambda_1, \dots, \lambda_m$.
- (ii) The Schmidt rank of a vector $|Z\rangle = \sum_i |v_i\rangle |w_i\rangle$ coincides with the rank of $v^T w$, where v is a matrix defined via $|v_i\rangle = \sum_j v_{ij} |j\rangle$ for some fixed

computational basis, and w is constructed analogously. As a first consequence, we deduce that $SR(Z) = r$ whenever $|Z\rangle = \sum_{i=1}^r |v_i\rangle |w_i\rangle$ with $\{|v_i\rangle\}_{i=1}^r, \{|w_i\rangle\}_{i=1}^r$ linearly independent families in their own local spaces. As a second corollary, observe that $SR(\sum_i |v_i\rangle |v_i^*\rangle) = \text{rk}(v)$.

We are now in position to prove the following.

Proposition 4.2. *Let r, s, n be positive integers such that $r, s \leq n$. Then there are (pure) states ρ_{AB}, σ_{AB} of an $n \times n$ bipartite quantum system such that $SR(\rho_{AB} \circ \sigma_{AB}) = \min\{n, rs\}$.*

Proof. Consider a bipartite $n \times n$ quantum system. Choose a large $N \geq \max\{n, rs\}$, and define $\omega := e^{2\pi i/N}$. Then, for all $0 \leq i \leq r-1$ and $0 \leq j \leq s-1$, introduce the local (unnormalised) states

$$|\alpha_i\rangle := \sum_{l=0}^{n-1} \omega^{il} |l\rangle, \quad |\beta_j\rangle := \sum_{l=0}^{n-1} \omega^{jrl} |l\rangle.$$

Using these states as building blocks, we construct

$$|\psi\rangle := \sum_{i=0}^{r-1} |\alpha_i \alpha_i^*\rangle, \quad |\varphi\rangle := \sum_{j=0}^{s-1} |\beta_j \beta_j^*\rangle.$$

It is easy to see that $\{|\alpha_i\rangle\}_{0 \leq i \leq r-1}$ and $\{|\beta_j\rangle\}_{0 \leq j \leq s-1}$ are two linearly independent families, since all the $\{\omega^i\}_{i=0}^{r-1}$ are distinct, as well as all the $\{\omega^{jr}\}_{j=0}^{s-1}$. Consequently, we see that $SR(|\psi\rangle\langle\psi|) = r$ and $SR(|\varphi\rangle\langle\varphi|) = s$. Taking the Hadamard product we obtain

$$|\psi\rangle\langle\psi| \circ |\varphi\rangle\langle\varphi| = |\psi \circ \varphi\rangle\langle\psi \circ \varphi|,$$

with

$$|\psi \circ \varphi\rangle = \sum_{i=0}^{r-1} \sum_{j=0}^{s-1} |\alpha_i \circ \beta_j\rangle |\alpha_i^* \circ \beta_j^*\rangle.$$

Observe that $|\alpha_i \circ \beta_j\rangle = \sum_l \omega^{(i+rj)l} |l\rangle$; therefore, the Schmidt rank of the above state is equal to the rank of the $n \times rs$ matrix M whose entries are given by $M_{l, i+rj} := \omega^{(i+rj)l}$. Since this is a Vandermonde matrix with rs distinct generating numbers, we conclude that $\text{rk } M = \min\{rs, n\}$. \square

4.3 Positivity region

This section is devoted to answering the most basic question concerning the maps defined in (4.2): *for what range of the parameters α, β, γ is $\Gamma[\alpha, \beta, \gamma]$ positive?* The most direct way of solving the problem is reported in the following Subsection 4.3.1. However, as in many cases, a direct proof is not necessarily the most illuminating. Thus, in Subsection 4.3.2 we choose to follow a longer

path to prove the main result of this section (Theorem 4.4), passing through Proposition 4.5 and Theorem 4.6. This is instructive because it shows the link between the map Γ of (4.2), acting on a bipartite system, and the map χ of (4.7), whose input is instead a single system.

4.3.1 Direct approach

Let us start with a preliminary result of elementary linear algebra.

Lemma 4.3. *Let $M > 0$ be a strictly positive definite matrix, and let $\psi = |\psi\rangle\langle\psi|$ be a (non necessarily normalised) pure state. For κ a real coefficient, $M + \kappa\psi \geq 0$ iff $1 + \kappa\langle\psi|M^{-1}|\psi\rangle \geq 0$.*

Proof. Observe that

$$M + \kappa|\varphi\rangle\langle\varphi| = M^{1/2}(\mathbb{1} + \kappa M^{-1/2}|\varphi\rangle\langle\varphi|M^{-1/2})M^{1/2}$$

is positive semidefinite iff so is $\mathbb{1} + \kappa M^{-1/2}|\varphi\rangle\langle\varphi|M^{-1/2}$, which happens iff $1 + \kappa\langle\varphi|M^{-1}|\varphi\rangle \geq 0$. \square

Now, the main result of the present section is as follows.

Theorem 4.4. *The map $\Gamma[\alpha, \beta, \gamma]$ defined by (4.2) is positive iff:*

$$\begin{aligned} \alpha + 1, \beta + 1 &\geq 0, \\ \frac{\alpha + \beta}{n} + \gamma + 1 &\geq 0, \\ \alpha + \beta + \gamma + 1 &\geq 0. \end{aligned} \tag{4.10}$$

Here, we used the notation $n := \min\{n_A, n_B\}$.

Before delving into the proof of the above result, let us try to represent the region determined by the inequalities (4.10) on a 3-dimensional plot. This will also help us to get some geometrical understanding of the various steps of the argument.

Proof of Theorem 4.4. First of all, conditions (4.10) are necessary, as can be seen by choosing suitable input states. In the following, we often write Γ instead of $\Gamma[\alpha, \beta, \gamma]$.

- A maximally entangled state $|\Phi\rangle$ gives $\Gamma(\Phi) = \left(1 + \frac{\alpha + \beta}{n}\right)\mathbb{1} + \gamma\Phi \geq 0$, from which $1 + \frac{\alpha + \beta}{n} + \gamma \geq 0$ follows.
- A product state $|11\rangle$ gives $\Phi(|11\rangle\langle 11|) = \mathbb{1} + \alpha\mathbb{1} \otimes |1\rangle\langle 1| + \beta|1\rangle\langle 1| \otimes \mathbb{1} + \gamma|11\rangle\langle 11| \geq 0$. Taking the matrix element on the same state produces $1 + \alpha + \beta + \gamma \geq 0$.
- As above, input $|11\rangle$ but take the matrix element on pure states $|21\rangle$ and $|12\rangle$, producing the condition $1 + \alpha \geq 0$ and $1 + \beta \geq 0$.

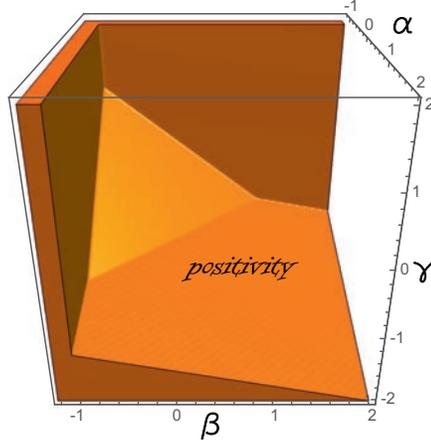


Figure 4.1: The solid region represents the parameter region for which the map $\Gamma[\alpha, \beta, \gamma]$ of (4.2) is not positive. Here the the $n = 4$ case is considered.

Observe that if $\alpha + \beta \geq 0$ the third condition in (4.10) is subsumed under the second one, while on the contrary if $\alpha + \beta < 0$ the second condition is subsumed under the third one. This completes the proof of the necessity of the inequalities (4.10).

In order to prove their sufficiency, observe the region that they determine in Figure 4.1. It is apparent that every point belonging to that region is a convex combination of three points on the three half-lines that constitute the ‘edges’ of the region. We proceed to list them.

- (1) The vertical line that lies at the intersection between the $\alpha\gamma$ and $\beta\gamma$ coordinate planes:

$$\{(-1, -1, \gamma) : \gamma \geq 1\}. \quad (4.11)$$

- (2) The lower line which runs parallel to the $\beta\gamma$ plane:

$$\left\{ \left(-1, \beta, -1 - \frac{\beta - 1}{n} \right) : \beta \geq 1 \right\}. \quad (4.12)$$

- (3) The lower line which runs parallel to the $\alpha\gamma$ plane:

$$\left\{ \left(\alpha, -1, -1 - \frac{\alpha - 1}{n} \right) : \alpha \geq 1 \right\}. \quad (4.13)$$

If we prove that all of the above half-lines are entirely composed of positive maps, we are done. Let us proceed in the same order.

- (1) $\Gamma[-1, -1, \gamma] = (\mathbf{1}\text{Tr} - I) \otimes (\mathbf{1}\text{Tr} - I) + (\gamma - 1)I$ is positive if $\gamma \geq 1$, because the first addend is the tensor product of two completely copositive

maps (Subsection 4.2.2). In other words, up to composing it with a global transposition (that is positive and invertible), it is nothing but the product of two completely positive maps.

(2) $\Gamma[-1, \beta, -1 - \frac{\beta-1}{n}] = (\mathbb{1}\text{Tr} + I) \otimes (\mathbb{1}\text{Tr} - I) + (\beta - 1) I \otimes (\mathbb{1}\text{Tr} - \frac{1}{n}I)$ is positive if $\beta \geq 1$, as we are set to show.

- The first addend is positive since it is a tensor product of an entanglement-breaking channel on the first subsystem and a positive map on the second subsystem (Subsection 4.2.2).
- The second addend is positive because by evaluating it directly on a pure state $\psi = |\psi\rangle\langle\psi|$ one gets a positive result. In fact, writing $|\psi\rangle$ as in (4.3), one has

$$\begin{aligned} \left(I \otimes \left(\mathbb{1}\text{Tr} - \frac{I}{n} \right) \right) (\psi_{AB}) &= (\text{Tr}_B \psi_{AB}) \otimes \mathbb{1}_B - \frac{1}{n} \psi_{AB} \\ &= \left(\sum_{i=1}^n \lambda_i |e_i\rangle\langle e_i| \right) \otimes \mathbb{1} - \frac{1}{n} |\psi\rangle\langle\psi| \\ &\geq 0, \end{aligned}$$

where the last step follows easily from Lemma 4.3 with $\kappa = -1$, considering that

$$\frac{1}{n} \langle \psi | \left(\sum_{i=1}^n \lambda_i |e_i\rangle\langle e_i| \right)^{-1} \otimes \mathbb{1} | \psi \rangle = \sum_{i=1}^n \lambda_i = 1.$$

The above argument works essentially because $\mathbb{1}_B \text{Tr}_B - \frac{1}{n} I_B$ is completely positive if $n = n_B$ and at least n -positive if $n = n_A$.

(3) The third case is completely analogous to the second one and can be treated in a symmetrically identical way.

□

4.3.2 An alternative proof of Theorem 4.4

As we mentioned, there is a less direct pathway to arrive at Theorem 4.4 that is worth exploring. In fact, it involves a characterisation theorem for the maps χ presented in (4.7), which might be of independent interest. The first important observation we present is then the reduction of the problem of positivity for $\Gamma[\alpha, \beta, \gamma]$ to the analogous problem for $\chi[a, c]$.

Proposition 4.5. *The map $\Gamma[\alpha, \beta, \gamma]$ is positive iff $\alpha, \beta \geq -1$ and $\chi[\alpha + \beta, \gamma]$ is positive.*

Proof. Clearly, in order to ensure the positivity of Γ we have to require that $\Gamma(\psi) \geq 0$ for all global pure states $\psi = |\psi\rangle\langle\psi|$. Since Γ commutes with the conjugation by local unitaries, we can suppose that the input state is in Schmidt normal form as in (4.3), with $|e_i\rangle = |i\rangle$ and $|f_j\rangle = |j\rangle$. Then,

$$\begin{aligned}\Gamma(\psi) &= \mathbb{1} + \sum_{i,j} (\alpha\lambda_j + \beta\lambda_i) |ij\rangle\langle ij| + \gamma\psi \\ &= \sum_{i \neq j} (1 + \alpha\lambda_j + \beta\lambda_i) |ij\rangle\langle ij| \\ &\quad + \sum_{ij} \left((1 + (\alpha + \beta)\lambda_i) \delta_{ij} + \gamma \sqrt{\lambda_i\lambda_j} \right) |ii\rangle\langle jj| .\end{aligned}$$

From the above block decomposition it is apparent that $\Gamma(\psi) \geq 0$ iff $1 + \alpha\lambda_j + \beta\lambda_i \geq 0$ and

$$\sum_{ij} \left((1 + (\alpha + \beta)\lambda_i) \delta_{ij} + \gamma \sqrt{\lambda_i\lambda_j} \right) |i\rangle\langle j| \geq 0 .$$

Once we require the validity of these conditions for all the probability distributions λ , the first one reads $\alpha, \beta \geq -1$. At the same time, the above matrix is exactly $\chi[\alpha + \beta, \gamma] \left(\tilde{\psi} \right)$, with $|\tilde{\psi}\rangle := \sum_i \sqrt{\lambda_i} |i\rangle$, and therefore we are led to require the positivity of the map $\chi[\alpha + \beta, \gamma]$ as the second and last condition. \square

The above proposition shows how the map $\chi[\alpha + \beta, \gamma]$ is in fact the restriction of $\Gamma[\alpha, \beta, \gamma]$ to the maximally coherent subspace. This, in turn, reduces our main problem to the question of determining what is the region of positivity of the map (4.7). Now, we proceed to show how to solve this latter problem.

Theorem 4.6. *The map $\chi[a, c]$ defined by (4.7) is positive iff*

$$\begin{aligned}a + 2 &\geq 0, \\ \frac{a}{n} + c + 1 &\geq 0, \\ a + c + 1 &\geq 0.\end{aligned}\tag{4.14}$$

Proof. Suppose first that $a \geq 0$. Denoting with D_λ the diagonal matrix with diagonal $\lambda \in \mathbb{R}^n$, the positivity of $\chi[a, c]$ amounts to require that

$$\mathbb{1} + a D_\lambda + c |\psi\rangle\langle\psi| \geq 0\tag{4.15}$$

for all pure states $|\psi\rangle = \sum_{i=1}^n \sqrt{\lambda_i} |i\rangle$, where λ is an arbitrary probability distribution over the alphabet $\{1, \dots, n\}$ (phases on $|\psi\rangle$ are irrelevant, as the application of a diagonal unitary immediately reveals). Since $\mathbb{1} + a D_\lambda > 0$, one can apply Lemma 4.3 with $\kappa = c$ and rewrite the above condition as

$$1 + c \langle\psi| (\mathbb{1} + a D_\lambda)^{-1} |\psi\rangle \geq 0,$$

i.e.

$$c \geq -\frac{1}{\langle \psi | (\mathbb{1} + aD_\lambda)^{-1} | \psi \rangle} = -\frac{1}{\sum_i \frac{\lambda_i}{1+a\lambda_i}}.$$

We have to impose the above constraint for all probability vectors $\lambda \in \mathbb{R}^n$. The strictest condition is that corresponding to the vector λ that achieves the maximum of $f(\lambda) := \sum_i \frac{\lambda_i}{1+a\lambda_i}$. From the strict concavity of $\frac{x}{1+ax}$ on $x \geq 0$ it follows that $f(\lambda)$ is strictly concave on the simplex of the allowed vectors λ . Therefore, any internal stationary point is automatically a global maximum. Applying the method of Lagrange multipliers gives us $\lambda_i := \text{const} = \frac{1}{n}$ as the unique stationary point of f , and thus $f_{\max} = \frac{1}{1+\frac{a}{n}}$. The final condition on c becomes exactly

$$c \geq -1 - \frac{a}{n},$$

which is what (4.14) dictates for the present case $a \geq 0$.

Now, let us consider the opposite possibility, $a < 0$. Since the diagonal entries of the operator in (4.15) are $1 + (a+c)\lambda_i$, imposing their positivity for all the allowed λ amounts to requiring that $a+c+1 \geq 0$. Moreover, observe that if $a < -2$ it is possible to force $\mathbb{1} + aD_\lambda$ to have a 2-dimensional negative eigenspace. In this case, an addition of pure state could not make the whole operator positive. Therefore, we have to demand $a \geq -2$. In order to show that these two conditions are sufficient, consider an arbitrary pure state $|v\rangle$ and take the matrix element of (4.15) on it.

$$\langle v | (\mathbb{1} + aD_\lambda + c|\psi\rangle\langle\psi|) | v \rangle = 1 + a \sum_i \lambda_i |v_i|^2 + c \left| \sum_i \sqrt{\lambda_i} v_i \right|^2 \quad (4.16)$$

In order to prove that the above quantity is always positive if $a \geq -2$ and $a+c+1 \geq 0$, we want to formalise the following intuition. On the one hand, it is not possible to make the coefficient of c small without choosing at least two non-zero (and comparable) v_i , thus reducing the negative impact of the coefficient of a . On the other hand, if one wants to concentrate the weights $\sqrt{\lambda_i} v_i$ on a single element, then a appears always summed with c . We can make the above reasoning rigorous by considering the following inequality, valid for arbitrary $z_1, \dots, z_n \in \mathbb{C}$:

$$\sum_i |z_i|^2 \leq \frac{1}{2} \left(\left(\sum_i |z_i| \right)^2 + \left| \sum_i z_i \right|^2 \right). \quad (4.17)$$

Its proof follows easily once one rewrites the difference of its two sides as

$$\sum_{ij} (|z_i||z_j| + \Re(z_i^* z_j) - 2|z_i|^2 \delta_{ij}) \geq 0,$$

where the latter inequality holds trivially because the left-hand side is a sum of positive terms.¹ Applying 4.17 with $z_i = \sqrt{\lambda_i}v_i$ yields

$$\sum_i \lambda_i |v_i|^2 \leq \frac{1+k}{2},$$

where we noticed that $|\sum_i \sqrt{\lambda_i}v_i| \leq (\sum_i \lambda_i)^{1/2} (\sum_i |v_i|^2)^{1/2} \leq 1$ and wrote for brevity $k := |\langle \psi | v \rangle|^2$. Then, taking into account that $a < 0$, (4.16) can be lower bounded by

$$\begin{aligned} \langle v | (\mathbb{1} + aD_\lambda + c|\psi\rangle\langle\psi|) | v \rangle &\geq 1 + a\frac{1+k}{2} + ck \\ &= (1+a+c)k + \left(1 + \frac{a}{2}\right)(1-k) \\ &\geq 0, \end{aligned}$$

i.e. it is non-negative, as we claimed. \square

Remark. One could also give a direct proof of the positivity of $\chi[a, c]$ when $-2 \leq a \leq 0$, $1 + a + c \geq 0$, based on Hadamard product arguments. Clearly, up to taking convex combinations with known positive maps, one has to prove only the positivity of the extreme point $\chi[-2, 1] = \mathbb{1}\text{Tr} - 2D + I$. Then, the crucial observation consists in writing the above map as a Hadamard square of a completely copositive map, i.e. $\chi[-2, 1] = (\mathbb{1}\text{Tr} - I)^{\circ 2}$, where the basis we choose from now on to take Hadamard products is the canonical basis of operators $\{|i\rangle\langle j|\}_{i,j}$ (or its Choi–Jamiolkowski equivalent $\{|ij\rangle\}_{i,j}$). The following reasoning then shows that Hadamard multiplying two completely copositive maps yields another completely copositive map. Using the notation (4.4), one has

$$R_{\Lambda_1}^{T_A}, R_{\Lambda_2}^{T_A} \geq 0 \implies R_{\Lambda_1 \circ \Lambda_2}^{T_A} = (R_{\Lambda_1} \circ R_{\Lambda_2})^{T_A} = R_{\Lambda_1}^{T_A} \circ R_{\Lambda_2}^{T_A} \geq 0.$$

As an easy corollary of the above discussion, we deduce the following result.

Alternative proof of Theorem 4.4. Follows by combining Proposition 4.5 and Theorem 4.6. \square

4.4 Complete Positivity region

Determining the range of α, β, γ for which the map $\Gamma[\alpha, \beta, \gamma]$ given by (4.2) is completely positive requires the construction of the Choi state associated to Φ .

¹A nice geometrical interpretation arises when one considers a closed polygon, that is, numbers z_1, \dots, z_n such that $\sum_i z_i = 0$. In this case, (4.17) states that the sum of the square sides is at most *half* of the square perimeter.

Calling $A'B'$ the twin system of AB , a maximally entangled state $AB : A'B'$ reads

$$\begin{aligned} |\Phi\rangle_{AB:A'B'} &= \frac{1}{\sqrt{n_A n_B}} \sum_{i=1}^{n_A} \sum_{j=1}^{n_B} |ij\rangle_{AB} |ij\rangle_{A'B'} \\ &= |\Phi\rangle_{AB} |\Phi\rangle_{A'B'}. \end{aligned}$$

As a consequence, the Choi state $R_\Gamma = (\Gamma_{AB} \otimes I_{A'B'}) (\Phi_{AB:A'B'})$ becomes

$$R_\Gamma = \frac{\mathbb{1}_{ABA'B'}}{n_A n_B} + \alpha \frac{\mathbb{1}_{AA'}}{n_A} \otimes \Phi_{BB'} + \beta \Phi_{AA'} \otimes \frac{\mathbb{1}_{BB'}}{n_B} + \gamma \Phi_{AA'} \otimes \Phi_{BB'}. \quad (4.18)$$

The following result follows easily.

Theorem 4.7. *The map $\Gamma[\alpha, \beta, \gamma]$ defined by (4.2) is completely positive iff*

$$\begin{aligned} 1 + n_B \alpha &\geq 0, \\ 1 + n_A \beta &\geq 0, \\ 1 + n_B \alpha + n_A \beta + n_A n_B \gamma &\geq 0. \end{aligned} \quad (4.19)$$

Proof. Since the four addends appearing in (4.18) commute, the diagonalisation of their sum is straightforward. It is easily seen that the distinct eigenvalues of the operator in (4.18) are

$$\frac{1}{n_A n_B}, \quad \frac{1}{n_A n_B} + \frac{\alpha}{n_A}, \quad \frac{1}{n_A n_B} + \frac{\beta}{n_B}, \quad \frac{1}{n_A n_B} + \frac{\alpha}{n_A} + \frac{\beta}{n_B} + \gamma.$$

Imposing positivity of the last three expressions yields (4.19), as claimed. \square

As can be easily verified, conditions (4.19) imply that every completely positive map in the Γ class can be written as a convex combination of three points lying on the three ‘edge’ half-lines coming out from the vertex $(-\frac{1}{n_B}, -\frac{1}{n_A}, \frac{1}{n_A n_B})$. These can be represented as follows:

(1) the vertical half-line is

$$\left\{ (-1/n_B, -1/n_A, \gamma) : \gamma \geq 1/n_A n_B \right\};$$

(2) one of the other two is

$$\left\{ (-1/n_B, \beta, -\beta/n_B) : \beta \geq -1/n_A \right\},$$

(3) while the last one is symmetrically described as

$$\left\{ (\alpha, -1/n_A, -\alpha/n_A) : \alpha \geq -1/n_B \right\}.$$

4.5 Entanglement-breaking region

We now move on to the question of finding the region in the α, β, γ parameter space that defines an entanglement-breaking map through equation (4.2). Obviously, such a region must be contained in the complete positivity solid defined via (4.19). Reformulating the problem with the help of the Choi-Jamiolkowski isomorphism, we want to determine necessary and sufficient conditions for the separability of the state (4.18). This problem has already been solved in the special case $n_A = n_B$ in [203], but we will see that the most interesting phenomena appear when one considers the asymmetric case $n_A \neq n_B$.

In Subsection 4.5.1 we deduce the main result of the section, Theorem 4.9, using some abstract arguments based on the symmetry properties of the problem. This is an instance of the general procedure detailed in Subsection 2.4.2. In Subsection 4.5.2, instead, we perform a double check and exhibit explicit separable decompositions showing that the maps identified by Theorem 4.9 are indeed entanglement-breaking.

4.5.1 Computation of the entanglement-breaking region

This subsection is devoted to the determination of the region in the parameter space α, β, γ for which the map $\Gamma[\alpha, \beta, \gamma]$ in (4.2) is entanglement-breaking. As we already saw, this amounts to deciding the separability of the corresponding Choi state $(R_\Gamma)_{ABA'B'}$ defined by (4.18) with respect to the bipartition $AB : A'B'$. An important preliminary observation is that $(R_\Gamma)_{ABA'B'}$ belongs to the subspace \mathcal{V} of linear operators that commute with unitaries of the form $U_A \otimes V_B \otimes U_{A'}^* \otimes V_{B'}^*$ (for all local unitaries U, V). This subspace is a central section in the sense of Subsection 2.4.2. In fact, it constitutes a special case of Example 2.18, with $V_1 = V_2 = \mathcal{H}_{n_A n_B}$, $G = SU(n_A) \times SU(n_B)$, and the local representations given by

$$\begin{aligned}\zeta_1(U, V)(Z_{AB}) &:= U_A \otimes V_B Z_{AB} U_A^\dagger \otimes V_B^\dagger, \\ \zeta_2(U, V)(Z_{AB}) &:= U_A^* \otimes V_B^* Z_{AB} U_A^T \otimes V_B^T.\end{aligned}$$

We can therefore exploit Proposition 2.20 in order to simplify the solution of the separability problem. To this purpose, observe that under the natural identification $V_{AB} = \mathcal{H}_{n_A n_B} = V_{AB}^*$ (made possible by the Hilbert-Schmidt product) one has $\mathcal{V}^* = \mathcal{V}$.

Theorem 4.8. *If $n_A = n_B$, the map $\Gamma[\alpha, \beta, \gamma]$ in (4.2) is entanglement-breaking iff it is completely positive and PPT-binding. If $n_A < n_B$, it is entanglement-breaking iff*

$$\Gamma, \quad T\Gamma \quad \text{and} \quad \left(I_A \otimes \left(\mathbf{1}\text{Tr} - \frac{I}{n_A} \right)_B \right) \Gamma \quad (4.20)$$

are completely positive. If $n_A > n_B$ a reversed but analogous condition holds (just exchange subscripts A and B).

Proof. As we already mentioned, Γ is entanglement-breaking iff its Choi state $(R_\Gamma)_{ABA'B'}$ (4.18) is separable with respect to the bipartition $AB : A'B'$. Proposition 2.20 implies that this happens iff $\text{Tr}WR_\Gamma \geq 0$ for all entanglement witnesses W that share the same symmetry of R_Γ , i.e. that commute with unitaries of the form $U_A \otimes V_B \otimes U_{A'}^* \otimes V_{B'}^*$, for all $U \in SU(n_A)$, $V \in SU(n_B)$. In other words, W can be assumed to be the Choi matrix of a positive map $\Gamma' := \Gamma[\alpha', \beta', \gamma']$ belonging to the set defined by (4.1) (or a limit point of the form $\lim_{\kappa \rightarrow \infty} \frac{1}{\kappa} \Gamma[\kappa\alpha', \kappa\beta', \kappa\gamma']$, with $(\kappa\alpha', \kappa\beta', \kappa\gamma')$ defining a positive map for all $\kappa > 0$; this case will not be considered further because it does not introduce any new constraint).

We now show that instead of requiring the scalar condition $\text{Tr}[R_{\Gamma'}R_\Gamma] \geq 0$ to hold for all positive Γ' of the family (4.2), we can equivalently impose that $\Gamma'\Gamma$ is completely positive for all positive Γ' as in (4.2). To show this, let us write

$$\begin{aligned} \text{Tr}[R_{\Gamma'}R_\Gamma] &= \text{Tr}[(\Gamma' \otimes I)(\Phi)(\Gamma \otimes I)(\Phi)] \\ &= \text{Tr}[\Phi((\Gamma')^\dagger \Gamma \otimes I)(\Phi)] \\ &= \langle \Phi | R_{(\Gamma')^\dagger \Gamma} | \Phi \rangle \\ &= \langle \Phi | R_{\Gamma'\Gamma} | \Phi \rangle, \end{aligned}$$

where for the last step we observed that all the maps in (4.2) are self-adjoint. Now, clearly $\text{Tr}[R_{\Gamma'}R_\Gamma] \geq 0$ follows from the complete positivity of $\Gamma'\Gamma$, which indeed implies the stronger condition $R_{\Gamma'\Gamma} \geq 0$. On the other hand, if $\text{Tr}[R_{\Gamma'}R_\Gamma] \geq 0$ holds for all positive maps Γ' of the family (4.2), then we have seen that Γ must be entanglement-breaking, which in turn implies that $\Gamma'\Gamma$ is completely positive whenever Γ' is positive.

The above reasoning shows that the complete positivity of $\Gamma'\Gamma$ for all positive Γ' as in (4.2) is necessary and sufficient for Γ to be entanglement-breaking. From now on, we assume without loss of generality $n_A \leq n_B$. As detailed in the proof of Theorem 4.4, the positivity region defined by 4.1 is the convex hull of the three half-lines (4.11), (4.12) and (4.13). Consequently, it suffices to test the three families of ‘witness maps’ Γ' corresponding to those three half-lines. For each family with parameter κ ($\kappa = \gamma, \beta, \alpha$ in order), it suffices to test first the case $\kappa = 1$ and then the limit $\kappa \rightarrow \infty$. This leads to six tests corresponding to just as many positive maps:

- (i) $\Gamma[-1, -1, 1] = (\mathbf{1}\text{Tr} - I) \otimes (\mathbf{1}\text{Tr} - I)$;
- (ii) $\lim_{\kappa \rightarrow \infty} \frac{1}{\kappa} \Gamma[-1, -1, \kappa] = I$;
- (iii) $\Gamma[-1, 1, -1] = (\mathbf{1}\text{Tr} + I) \otimes (\mathbf{1}\text{Tr} - I)$;
- (iv) $\lim_{\kappa \rightarrow \infty} \frac{1}{\kappa} \Gamma[-1, \kappa, -1 - (\kappa - 1)/n_A] = I \otimes \left(\mathbf{1}\text{Tr} - \frac{1}{n_A} I \right)$;
- (v) $\Gamma[1, -1, -1] = (\mathbf{1}\text{Tr} - I) \otimes (\mathbf{1}\text{Tr} + I)$;
- (vi) $\lim_{\kappa \rightarrow \infty} \frac{1}{\kappa} \Gamma[\kappa, -1, -1 - (\kappa - 1)/n_B] = \left(\mathbf{1}\text{Tr} - \frac{1}{n_B} I \right) \otimes I$.

Now, we want to understand which one of the above six tests can be subsumed under the PPT condition that Γ and $T\Gamma$ are CP. This happens when the witness is *decomposable*, i.e. it is a positive combination of a CP and a coCP map. Let us analyse the above maps one by one:

- (i) Completely copositive, as is the tensor product of two completely copositive maps (Subsection 4.2.2).
- (ii) Obviously completely positive.
- (iii) Completely copositive, as is the tensor product of two completely copositive maps (Subsection 4.2.2).
- (iv) Only positive, but not completely positive unless $n_A = n_B$; it will be clear later that it is actually indecomposable whenever $n_A \neq n_B$; for the time being, all we can conclude is that we can not a priori discard this test.
- (v) The same as (iii).
- (vi) This is not the same as (iv), because the condition $n_A \leq n_B$ ensures that $\mathbb{1}\text{Tr} - \frac{1}{n_B}I$, and thus the entire map, are completely positive.

From the above discussion it should be clear that if $n_A = n_B$ then the PPT test is both necessary and sufficient to ensure that Γ is entanglement-breaking, while if $n_A < n_B$ the only condition that can not be absorbed in the PPT test is the one in (iv). This is the same as saying that for $n_A < n_B$ a map Γ of the class in (4.2) is entanglement-breaking iff $T\Gamma$ and $\left(I_A \otimes \left(\mathbb{1}\text{Tr} - \frac{I}{n_A}\right)_B\right)\Gamma$ are completely positive, which concludes the proof. \square

Now that necessary and sufficient conditions for Γ to be entanglement-breaking have been written down in the form (4.20), it is only a matter of finding out the shape of the corresponding solid. An interesting question, as usual, is whether we really need the second test or on the contrary the PPT condition is actually sufficient. We already saw that the suspected answer is that if $n_A < n_B$ we *do* need the second test. In other words, in that case there are PPT entangled states of the form (4.18). The main result of this section reads as follows.

Theorem 4.9. *Assume that $n_A \leq n_B$. The map $\Gamma[\alpha, \beta, \gamma]$ defined by (4.2) is entanglement-breaking iff the following conditions are met:*

$$\begin{aligned}
1 + n_B\alpha &\geq 0, \\
1 + n_B\alpha + n_A\beta + n_A n_B\gamma &\geq 0, \\
1 - \alpha + \beta - \gamma &\geq 0, \\
1 + \alpha - \beta - \gamma &\geq 0, \\
1 - \alpha - \beta + \gamma &\geq 0, \\
(n_A n_B - 1)(n_A\beta + 1) - (n_B - n_A)(\alpha + n_A\gamma) &\geq 0.
\end{aligned} \tag{4.21}$$

The last inequality can be omitted if $n_A = n_B$. The solid described by the above system is a double pyramid with triangular basis (see Figure 4.2). The basis has vertices

$$(-1/n_B, -1/n_B, 1), \quad (1, -1/n_A, -1/n_A), \quad (-1/n_B, 1, -1/n_B),$$

while the culminating vertices of the two pyramids are

$$(-1/n_B, -1/n_A, 1/n_A n_B), \quad (1, 1, 1).$$

Proof. Thanks to Theorem 4.20, we have just to impose the complete positivity of Γ , $T\Gamma$ and $(I \otimes (\mathbb{1}\text{Tr} - I/n_A)) \Gamma$.

- Γ is CP. This gives the first two conditions of (4.21) together with the requirement $\beta \geq -1/n_A$. However, the latter can be neglected because it follows from the other inequalities of the system (4.21). Indeed, multiplying the second inequality of the first line by $(n_B - n_A)/n_B$ and summing the third line produces exactly $n_A \beta + 1 \geq 0$.
- $T\Gamma$ is CP. Taking the partial transposition $T_{A'B'}$ of the Choi state (4.18) yields

$$\begin{aligned} n_A n_B (\Gamma_{AB} \otimes T_{A'B'}) (\Phi_{AB:A'B'}) &= \mathbb{1}_{ABA'B'} + \alpha \mathbb{1}_{AA'} \otimes F_{BB'} \\ &\quad + \beta F_{AA'} \otimes \mathbb{1}_{BB'} + \gamma F_{AA'} \otimes F_{BB'}, \end{aligned}$$

where F is the operator that performs the swap between two subsystems. Since the four addends in the above equation commute, finding the eigenvalues of their sum is straightforward: they are given by $1 + \alpha + \beta + \gamma$, $1 + \alpha - \beta - \gamma$, $1 - \alpha + \beta - \gamma$, $1 - \alpha - \beta + \gamma$. As is easy to see, $1 + \alpha + \beta + \gamma \geq 0$ is already implied by the complete positivity conditions; in fact, using the second inequality in (4.21) to lower bound γ gives

$$\begin{aligned} 1 + \alpha + \beta + \gamma &\geq 1 - \frac{1}{n_A n_B} + \left(1 - \frac{1}{n_A}\right) \alpha + \left(1 - \frac{1}{n_B}\right) \beta \\ &\geq \left(1 - \frac{1}{n_A}\right) \left(1 - \frac{1}{n_B}\right) \\ &\geq 0, \end{aligned}$$

where we used also $\alpha \geq -1/n_B$ and $\beta \geq -1/n_A$ in the second step. We found also the third, fourth and fifth line of (4.21).

- $(I \otimes (\mathbb{1}\text{Tr} - I/n_A)) \Gamma$ is CP. Imposing this condition requires just mechanical calculations, because the composed map under examination belongs to the same parametric class defined by (4.2), and therefore Theorem 4.7 applies. Besides the last line of (4.21), we obtain two additional inequalities:

$$n_A n_B - 1 - (n_B - n_A) \alpha \geq 0, \quad (4.22)$$

$$\alpha + (n_A n_B - 1) \beta + n_A \gamma + n_B - \frac{1}{n_A} \geq 0. \quad (4.23)$$

As is easy to see, (4.22) is redundant, because the upper bound $\alpha \leq \frac{n_A n_B - 1}{n_B - n_A}$ is weaker than $\alpha \leq 1$ which follows from third, fourth and fifth line of (4.21) combined. As a matter of fact, (4.23) is also useless, because noting that $\alpha + n_A \gamma \geq -(1 + n_A \beta)/n_B$ (thanks to the second inequality of (4.21)) gives us

$$\begin{aligned} \alpha + (n_A n_B - 1)\beta + n_A \gamma + n_B - \frac{1}{n_A} &\geq \left(n_A n_B - 1 - \frac{n_A}{n_B} \right) \left(\beta + \frac{1}{n_A} \right) \\ &\geq 0, \end{aligned}$$

where we used also $\beta \geq -1/n_A$.

The shape of the solid defined by (4.21) can be seen in Figure 4.2. Finding the vertices is now an elementary exercise. \square

It is also possible to give a more direct proof of Theorem 4.9, consisting in finding explicitly separable expressions for the vertices of the solid defined by (4.21) and represented in Figure 4.2. For details, we refer the reader to the forthcoming Subsection 4.5.2.

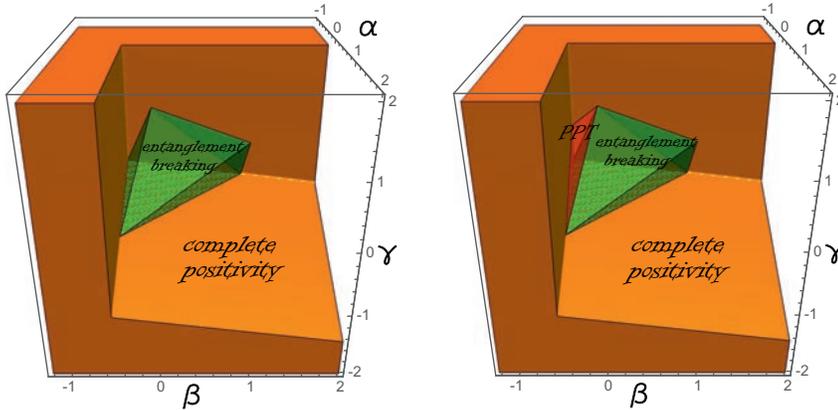


Figure 4.2: The three planes identify the complete positivity region for the maps defined in (4.2) in the α, β, γ space. Inside that, the entanglement-breaking solid (in green) is shown on the left. On the right, we added the pyramid for which this map is PPT but not entanglement-breaking (in red). Such a region exists iff $n_A \neq n_B$; here we chose the case $n_A = 2, n_B = 6$.

From the above proof of Theorem 4.9 we learnt that the PPT criterion is not sufficient for deciding separability as soon as $n_A \neq n_B$. Figure 4.2 shows the Γ maps that are PPT but not entanglement-breaking, forming another pyramid with a face of the entanglement-breaking solid as a basis and the point $(-\frac{1}{n_B}, -\frac{1}{n_A}, 1 - \frac{n_B - n_A}{n_A n_B})$ as the culminating vertex.

Remark. Several facts of independent interest can be deduced from the above discussion. Let us recall some of them.

- The state

$$R_{ABA'B'} := \mathbb{1}_{ABA'B'} - \mathbb{1}_{AA'} \otimes \Phi_{BB'} - \Phi_{AA'} \otimes \mathbb{1}_{BB'} + (n_A n_B - n_B + n_A) \Phi_{AA'} \otimes \Phi_{BB'}, \quad (4.24)$$

corresponding to the vertex of the PPT entangled pyramid depicted in red on the right of Figure 4.2, is a PPT entangled state of the bipartite system $AB : A'B'$. This construction can also be found in [200, Lemma 4].

- If $n_A < n_B$, the map

$$I_A \otimes \left(\mathbb{1} \text{Tr} - \frac{I}{n_A} \right)_B \quad (4.25)$$

acting on a bipartite system AB is positive but *indecomposable*, as can be seen by noting that it detects the PPT entangled state $R_{ABA'B'}$ defined by (4.24). As proven in [197, §IX] this remarkable fact is fully general: any k -positive but not completely positive map Λ gives rise to an indecomposable map $\Lambda \otimes I_k$ by taking a tensor product with the identity channel.

- The map

$$\Gamma[-1/n_B, -1/n_B, 1] = \mathbb{1}_{AB} \text{Tr} - \frac{1}{n_B} \mathbb{1}_A \text{Tr} \otimes I_B - \frac{1}{n_B} I_A \otimes \mathbb{1}_B \text{Tr} + I_{AB} \quad (4.26)$$

is entanglement-breaking when acting on a bipartite system AB such that $n_A \leq n_B$.

4.5.2 A more direct verification of Theorem 4.9

The proof of the above Theorem 4.9 rests crucially on the argument showing that the necessary conditions we found by composing Γ with some suitably chosen positive maps are also sufficient to ensure that Γ is entanglement-breaking. In turn, that argument is an application of the theory of symmetries as developed in Subsection 2.4.2. The curious reader might wonder, whether it is instead possible to show directly that the maps corresponding to the vertices of the solid (4.21) are entanglement-breaking. This is precisely what we are set to do now.

We start by reminding the reader that for a bipartite quantum system AA' (with $n_A = n_{A'} = n$) one can specialise Example 2.18 by taking $V_1 = V_2 = \mathcal{H}_n$, $G = SU(n)$, and

$$\begin{aligned} \zeta_1(U)(X) &:= UXU^\dagger, \\ \zeta_2(U)(X) &:= U^* X U^T. \end{aligned}$$

The corresponding *isotropic projection*, defined as in (2.57), acts as [118]

$$\Pi^{\text{iso}}(\cdot) := \int dU U \otimes U^* (\cdot) U^\dagger \otimes U^T \quad (4.27)$$

$$= |\Phi\rangle\langle\Phi| (\cdot) |\Phi\rangle\langle\Phi| + \frac{\mathbb{1} - \Phi}{n^2 - 1} \text{Tr} [(\mathbb{1} - \Phi) (\cdot)]. \quad (4.28)$$

It follows directly from (4.27) that Π^{iso} is separability-preserving as a super-operator.

Now we are ready to prove that each of the vertices of the double pyramid of Figure 4.2 corresponds to an entanglement-breaking map (or, equivalently, to a separable Choi operator (4.18)). Only one of the five vertices requires a special treatment, while the remaining four are easily seen to correspond to products of entanglement-breaking maps on A and B , which are in turn entanglement-breaking maps on AB .

- Basis vertex $\left(-\frac{1}{n_B}, -\frac{1}{n_B}, 1\right)$.

The corresponding Choi state reads

$$\begin{aligned} R_{\Gamma[-1/n_B, -1/n_B, 1]} &= \frac{\mathbb{1}_{ABA'B'}}{n_A n_B} - \frac{1}{n_A n_B} \mathbb{1}_{AA'} \otimes \Phi_{BB'} \\ &\quad - \frac{1}{n_B^2} \Phi_{AA'} \otimes \mathbb{1}_{BB'} + \Phi_{AA'} \otimes \Phi_{BB'}. \end{aligned}$$

While being aware that in general A and B do not need to have the same dimension, we can nevertheless introduce a sort of maximally entangled state

$$|\tilde{\Phi}\rangle_{AB} := \frac{1}{\sqrt{n_A}} \sum_{i=1}^{n_A} |i\rangle_A |i\rangle_B.$$

With this notation, some calculations reveal that one can write

$$R_{\Gamma[-1/n_B, -1/n_B, 1]} = \frac{n_A(n_B^2 - 1)}{n_B} (\Pi_{AA'}^{\text{iso}} \otimes \Pi_{BB'}^{\text{iso}}) \left(\tilde{\Phi}_{AB} \otimes \tilde{\Phi}_{A'B'} \right),$$

where $\tilde{\Phi} := |\tilde{\Phi}\rangle\langle\tilde{\Phi}|$. Since the right-hand side consists of the application of a map that preserves separability with respect to every cut to a state that is already $(AB : A'B')$ -separable, we must conclude that the left-hand side is indeed $(AB : A'B')$ -separable.

- Basis vertex $\left(1, -\frac{1}{n_A}, -\frac{1}{n_A}\right)$.

The corresponding map reads

$$\begin{aligned} \Gamma[1, -1/n_A, -1/n_A] &= \mathbb{1}\text{Tr} \otimes \mathbb{1}\text{Tr} + \mathbb{1}\text{Tr} \otimes I - \frac{1}{n_A} I \otimes \mathbb{1}\text{Tr} - \frac{1}{n_A} I \\ &= \left(\mathbb{1}\text{Tr} - \frac{I}{n_A} \right) \otimes (\mathbb{1}\text{Tr} + I), \end{aligned}$$

and the rightmost side, being a tensor product of two entanglement-breaking maps on A and B , is entanglement-breaking on the composite system AB .

- Basis vertex $\left(-\frac{1}{n_B}, 1, -\frac{1}{n_B}\right)$.

This case is completely analogous to the previous one:

$$\begin{aligned}\Gamma[-1/n_B, 1, -1/n_B] &= \mathbb{1}\text{Tr} \otimes \mathbb{1}\text{Tr} - \frac{1}{n_B} \mathbb{1}\text{Tr} \otimes I + I \otimes \mathbb{1}\text{Tr} - \frac{1}{n_B} I \\ &= (\mathbb{1}\text{Tr} + I) \otimes \left(\mathbb{1}\text{Tr} - \frac{I}{n_B}\right).\end{aligned}$$

- Culminating vertex $\left(-\frac{1}{n_B}, -\frac{1}{n_A}, \frac{1}{n_A n_B}\right)$.

We have

$$\begin{aligned}\Gamma[-1/n_B, -1/n_A, 1/n_A n_B] &= \mathbb{1}\text{Tr} \otimes \mathbb{1}\text{Tr} - \frac{1}{n_B} \mathbb{1}\text{Tr} \otimes I \\ &\quad - \frac{1}{n_A} I \otimes \mathbb{1}\text{Tr} + \frac{1}{n_A n_B} I \\ &= \left(\mathbb{1}\text{Tr} - \frac{I}{n_A}\right) \otimes \left(\mathbb{1}\text{Tr} - \frac{I}{n_B}\right),\end{aligned}$$

which is a tensor product of entanglement-breaking maps.

- Culminating vertex $(1, 1, 1)$.

The last case is

$$\begin{aligned}\Gamma[1, 1, 1] &= \mathbb{1}\text{Tr} \otimes \mathbb{1}\text{Tr} + \mathbb{1}\text{Tr} \otimes I + I \otimes \mathbb{1}\text{Tr} + I \\ &= (\mathbb{1}\text{Tr} + I) \otimes (\mathbb{1}\text{Tr} + I),\end{aligned}$$

again a tensor product of entanglement-breaking maps. This concludes our sanity check of the correctness of Theorem 4.9.

4.6 Entanglement-annihilating region

We now move on to the problem of characterising the triple of parameters α, β, γ for which the map $\Gamma[\alpha, \beta, \gamma]$ is entanglement-annihilating. This is done in Subsection 4.6.1, where we prove the main result of the section and perhaps of the whole chapter, Theorem 4.10. In the following Subsection 4.6.2 we then show how to apply this result to the solution of some open problems that have been raised recently [12, 13].

4.6.1 Main result

We remind the reader that an entanglement-annihilating map Λ_{AB} acting on a bipartite system is characterised by the property that $\Lambda_{AB}(\rho_{AB})$ is separable for all (pure) input states ρ_{AB} (Subsection 4.2.1). There is a naive necessary criterion that must be satisfied in order for this to be the case: if Λ_{AB} has to be entanglement-annihilating, then $(I_A \otimes T_B)\Lambda_{AB}$ must be positive (with T_B denoting partial transposition). As we mentioned in Subsection 4.2.1, maps for which the latter condition holds are called PPT-inducing in [13]. We are now in position to state and prove our main result.

Theorem 4.10. *Given a map $\Gamma[\alpha, \beta, \gamma]$ defined by (4.2), the following are equivalent:*

- (a) Γ is entanglement-annihilating ;
- (b) Γ is positive and PPT-inducing;
- (c) Γ is positive and in addition $\gamma \leq \alpha + \beta + 2$;
- (d) the following conditions are met:

$$\begin{aligned}
 \alpha + 1 &\geq 0, \\
 \beta + 1 &\geq 0, \\
 \frac{\alpha + \beta}{n} + \gamma + 1 &\geq 0, \\
 \alpha + \beta + \gamma + 1 &\geq 0, \\
 \alpha + \beta - \gamma + 2 &\geq 0.
 \end{aligned} \tag{4.29}$$

Proof. Let us break down the argument.

- (a) \Rightarrow (b). We already saw that an entanglement-annihilating map is necessarily PPT-inducing (and obviously positive).
- (b) \Rightarrow (c). Input to Γ the pure state $|\psi\rangle = \frac{|11\rangle + |22\rangle}{\sqrt{2}}$; the positivity of the partial transpose of the resulting state requires $\gamma \leq \alpha + \beta + 2$.
- (c) \Rightarrow (d). Trivially obtained by using Theorem 4.4. The region identified by these conditions is represented in Figure 4.3.
- (d) \Rightarrow (a). As one could expect, this is the only point that requires a bit of care. Observing Figure 4.3, we note that every point of the positive and PPT-inducing region is a convex combination of four points on the four half-lines forming the edges of the set. If we prove that all these four half-lines are composed entirely of entanglement-annihilating maps, we are done.

- The two lower half-lines have already been studied in the proof of Theorem 4.4. The one on the right of Figure 4.3 corresponds to (4.12), and is composed of maps of the form

$$\begin{aligned} \Gamma[-1, \beta, -1 - (\beta - 1)/n] &= (\mathbb{1}\text{Tr} + I) \otimes (\mathbb{1}\text{Tr} - I) \\ &\quad + (\beta - 1)I \otimes \left(\mathbb{1}\text{Tr} - \frac{1}{n}I\right), \end{aligned}$$

with $\beta \geq 1$. Both of the addends of the above equation are easily seen to be entanglement-annihilating, because they are tensor products of a positive and an entanglement-breaking map on the two subsystems. Consequently, their sum is entanglement-annihilating as well. The same reasoning applies to the other lower half-line (4.13).

- The two upper half-lines of Figure 4.3 are again symmetrically related and can be treated in the same way. The one on the right, for instance, is composed of maps of the form

$$\begin{aligned} \Gamma[-1, \beta, \beta + 1] &= \left(\mathbb{1}\text{Tr} - \frac{I}{2}\right) \otimes (\mathbb{1}\text{Tr} - I) \\ &\quad + \left(\beta + \frac{1}{2}\right)I \otimes (\mathbb{1}\text{Tr} + I), \end{aligned}$$

with $\beta \geq -\frac{1}{2}$. The second addend is entanglement-annihilating because it is a tensor product of the identity and an entanglement-breaking map. Proving that also the first addend is entanglement-annihilating is not completely trivial. Consider an arbitrary pure state $|\psi\rangle_{AB} = \sum_{i=1}^n \sqrt{\lambda_i} |i\rangle_A |i\rangle_B$, that we are assuming to be Schmidt decomposed (4.3) in the computational basis without loss of generality. We can also take the coefficients to be sorted in decreasing order, i.e. $\lambda_1 \geq \dots \geq \lambda_n > 0$, and such that $\sum_i \lambda_i = 1$. Define $\psi_{AB} := |\psi\rangle\langle\psi|_{AB}$, and denote the reduced state by $\rho_\psi := \text{Tr}_B \psi_{AB}$. We write

$$\begin{aligned} 2 \left(\mathbb{1}\text{Tr} - \frac{I}{2}\right) \otimes (\mathbb{1}\text{Tr} - I) (\psi) &= 2\mathbb{1} - 2\mathbb{1} \otimes \rho_\psi - \rho_\psi \otimes \mathbb{1} + \psi = \\ &= \sum_{i \neq j} \sqrt{\lambda_i \lambda_j} |ii\rangle\langle jj| \\ &\quad + \sum_{i,j} (2 - 2\lambda_j - \lambda_i + \lambda_i \delta_{ij}) |ij\rangle\langle ij|, \end{aligned}$$

where the indexes are understood to run from 1 to n . Basically, our strategy to prove that the above state is separable will consist in a comparison with a known separable state. Defining $Q := \sum_{i \neq j} |ij\rangle\langle ij|$, and denoting with Φ the projector onto the maximally entangled state (4.5), the operator

$$Q + n\Phi = \sum_{i \neq j} |ii\rangle\langle jj| + \sum_{i,j} |ij\rangle\langle ij| \quad (4.30)$$

turns out to be separable.² A first strategy could be based on a conjugation by a local diagonal matrix $D_\lambda = \text{diag}(\lambda_1, \dots, \lambda_n)$. One could write

$$\begin{aligned} D_\lambda \otimes \mathbb{1} (Q + n\Phi) D_\lambda \otimes \mathbb{1} &= \sum_{i \neq j} \sqrt{\lambda_i \lambda_j} |ii\rangle\langle jj| + \sum_{i,j} \lambda_i |ij\rangle\langle ij| = \\ &= 2\mathbb{1} - 2\mathbb{1} \otimes \rho_\psi - \rho_\psi \otimes \mathbb{1} + \psi \\ &\quad - \sum_{i,j} (2 - 2\lambda_i - 2\lambda_j + \lambda_i \delta_{ij}) |ij\rangle\langle ij|. \end{aligned}$$

If $2 - 2\lambda_i - 2\lambda_j + \lambda_i \delta_{ij} \geq 0$ for all i, j we would be done, because carrying the last addend on the left-hand side of the equation would yield a separable decomposition of the required state. However, the latter inequality fails to hold if $i = j = 1$ and $\lambda_1 > 2/3$. To include also this case, we must think of something different.

Construct $|\tilde{\psi}\rangle = \sum_i \sqrt{\lambda_i} |i\rangle$ and $\tilde{\psi} := |\tilde{\psi}\rangle\langle\tilde{\psi}|$, and use Theorem 4.6 to claim that

$$\begin{aligned} M &:= \mathbb{1} - 2D_\lambda + \tilde{\psi} \\ &= \sum_{i \neq j} \sqrt{\lambda_i \lambda_j} |i\rangle\langle j| + \sum_i (1 - \lambda_i) |i\rangle\langle i| \\ &\geq 0. \end{aligned}$$

Then, define the completely positive map ϕ_M acting as $\phi_M(X) := M \circ X$, where \circ denotes Hadamard product. One has

$$\begin{aligned} (I \otimes \phi_M)(Q + n\Phi) &= \sum_{i \neq j} \sqrt{\lambda_i \lambda_j} |ii\rangle\langle jj| + \sum_{i,j} (1 - \lambda_j) |ij\rangle\langle ij| \\ &= 2\mathbb{1} - 2\mathbb{1} \otimes \rho_\psi - \rho_\psi \otimes \mathbb{1} + \psi \\ &\quad - \sum_{i,j} (1 - \lambda_i - \lambda_j + \lambda_i \delta_{ij}) |ij\rangle\langle ij|. \end{aligned}$$

Since if $1 - \lambda_i - \lambda_j + \lambda_i \delta_{ij} \geq 0$ for all i, j (thanks to $\sum_i \lambda_i = 1$), we can conclude. □

²Although this seems to be known to experts in the field [184, Eq. (18)], we were not able to find an explicit proof in the literature, so let us provide one. Define $|+\rangle := \frac{1}{\sqrt{n}} \sum_{j=1}^n |j\rangle$, and for $\theta \in \mathbb{R}^n$ let $D_\theta := \sum_{j=1}^n e^{i\theta_j} |j\rangle\langle j|$ be the corresponding diagonal unitary. Then, one sees that

$$Q + n\Phi = \int_0^{2\pi} \frac{d^n \theta}{(2\pi)^n} D_\theta \otimes D_\theta |+\rangle\langle +| \otimes |+\rangle\langle +| D_\theta^\dagger \otimes D_\theta^\dagger,$$

implying that the left-hand side is indeed separable, as claimed.

We found particularly surprising that the state $2\mathbb{1} - 2\mathbb{1} \otimes \rho_\psi - \rho_\psi \otimes \mathbb{1} + \psi$ is separable for all the global pure states $|\psi\rangle$. Furthermore, the techniques we employed to prove this fact are to some extent original and possibly prone to be used more extensively within the context of the quantum separability problem. We will see in a moment that besides being interesting in itself, Theorem 4.10 is also useful in closing some open problems recently raised in the literature. Recently, we have learnt that the result can be employed to study the relation between the two seemingly unrelated concepts of entanglement-annihilating and entanglement-breaking channel, following [206, §IV.A].³

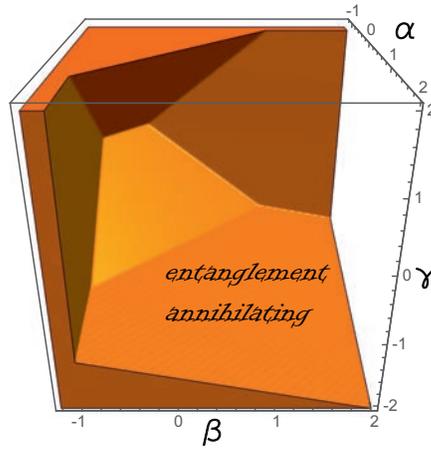


Figure 4.3: The convex region outside of the solid represents the parameter range for which the channel (4.2) is entanglement-annihilating. Here the case $n = 4$ is shown. Compare with Figure 4.1 and note the extra plane on the top identifying the additional condition $\gamma \leq \alpha + \beta + 2$.

4.6.2 Applications

Throughout this subsection, we apply the results in Section 4.6 to solve some open problems recently posed in [12, 13]. In those papers, the scenario in which a depolarising channel acts locally on two equal sides of a bipartite quantum system is considered. More specifically, using the notation in (4.1) with $-\frac{1}{n^2-1} \leq \lambda \leq 1$, it is asked what is the condition on λ_1, λ_2 under which $\Delta_{\lambda_1} \otimes \Delta_{\lambda_2}$ becomes entanglement-annihilating. In [12, Eq. (5)], a *sufficient condition* is found, that reads

$$(n^2 - 1)\lambda_1\lambda_2 \leq 1 + \frac{(n-2)(n+1)}{n+2}(\lambda_1 + \lambda_2). \quad (4.31)$$

³I am indebted to Alexander Müller-Hermes for this comment.

In particular, for the symmetric case $\lambda_1 = \lambda_2 = \lambda$, the explicit form

$$\lambda \leq \frac{n-2 + n\sqrt{\frac{2n}{n+1}}}{(n-1)(n+2)} \quad (4.32)$$

is deduced from (4.31). At the same time, the condition

$$\lambda \leq \frac{1 + \sqrt{3}}{n+1 + \sqrt{3}} \quad (4.33)$$

is shown to be necessary in order for $\Delta_\lambda \otimes \Delta_\lambda$ to be PPT-inducing. In [13] it is also conjectured, that (4.33) is also *sufficient* for enforcing the PPT-inducing behaviour (if the λ range is restricted to the complete positivity interval). Observe that:

- (i) there is a gap between the region (4.32) in which the entanglement annihilation is guaranteed and the region (4.33) outside which the global map is not even PPT-inducing;
- (ii) in [13] it is conjectured that inside that gap the map is still PPT-inducing, while no supposition is made about the entanglement annihilation.

Applying Theorem 4.10 straightforwardly solves all these problems, showing that (4.33) is indeed a *necessary and sufficient* condition for the map $\Delta_\lambda \otimes \Delta_\lambda$ to be PPT-inducing or, equivalently, entanglement-annihilating. In particular, this proves the conjecture in [13]. Before discussing the details of the above solution, we stress that it is not necessary to assume the complete positivity of the local maps (as done in both [12] and [13]) in order for our problem to make sense. On the contrary, it is enough to demand that the product $\Delta_{\lambda_1} \otimes \Delta_{\lambda_2}$ is positive. A straightforward application of Theorem 4.4 shows that this is the case iff

$$-\frac{1}{n-1} \leq \lambda_1, \lambda_2 \leq 1 \quad \text{and} \quad \lambda_1 \lambda_2 \geq -\frac{1}{n^2-1}. \quad (4.34)$$

By comparison, remind that the complete positivity conditions are

$$-\frac{1}{n^2-1} \leq \lambda_1, \lambda_2 \leq 1. \quad (4.35)$$

Now, we are in position to show our solution to the aforementioned open questions.

Corollary 4.11. *The product $\Delta_{\lambda_1} \otimes \Delta_{\lambda_2}$ of two local depolarising maps (4.1) is entanglement-annihilating iff, besides the positivity conditions (4.34), the inequality*

$$(n^2 + 2n - 2)\lambda_1 \lambda_2 \leq 2 + (n-2)(\lambda_1 + \lambda_2) \quad (4.36)$$

holds. In the symmetric case $\lambda_1 = \lambda_2 = \lambda$, (4.36) becomes simply

$$-\frac{\sqrt{3}-1}{n+1-\sqrt{3}} \leq \lambda \leq \frac{1+\sqrt{3}}{n+1+\sqrt{3}}. \quad (4.37)$$

Proof. We can apply Theorem 4.10 in the form of condition 3. Writing

$$\Delta_{\lambda_1} \otimes \Delta_{\lambda_2} = \lambda_1 \lambda_2 \Gamma \left[\frac{n\lambda_2}{1-\lambda_2}, \frac{n\lambda_1}{1-\lambda_1}, \frac{n^2\lambda_1\lambda_2}{(1-\lambda_1)(1-\lambda_2)} \right],$$

we have only to impose the further constraint

$$\frac{n^2\lambda_1\lambda_2}{(1-\lambda_1)(1-\lambda_2)} \leq \frac{n\lambda_2}{1-\lambda_2} + \frac{n\lambda_1}{1-\lambda_1} + 2,$$

which becomes (4.36) after elementary algebraic manipulations. Deducing (4.37) is now a simple exercise. Observe that the lower bound on λ expressed by (4.37) is superfluous if the complete positivity condition is imposed (and in that case we end up with (4.33)), but must be retained if only the positivity is imposed. \square

We want to answer another question that is left open in [12]. Besides the local depolarising noise, in that paper a global depolarizing noise of the form $(\Delta_\lambda)_{AB} = \lambda I_{AB} + (1-\lambda) \frac{\mathbb{1}_{AB}}{n_A n_B} \text{Tr}$ is also considered, in the simplest case $n_A = n_B = n$. It is observed that $(\Delta_\lambda)_{AB}$ is not PPT-inducing if $\lambda > \frac{2}{n^2+2}$, but an explicit entanglement-annihilating construction is provided only for $\lambda \leq \frac{n+2}{(n+1)(n^2-n+2)}$.

It is very simple to observe that $(\Delta_{2/(n^2+2)})_{AB} = \frac{1}{n^2+2}(\mathbb{1} \text{Tr} + 2I)_{AB}$ is already entanglement-annihilating, indeed. This follows easily from Theorem 4.10, but is also a consequence of the well-known fact [125] that $\mathbb{1} + 2\rho$ is separable for all normalised density matrices ρ on a bipartite system.

4.7 Conclusion

This concludes the characterisation of the extended depolarising channel. As we have shown, this natural generalisation to bipartite systems of the paradigmatic noisy channel displays a rich structure that nonetheless admits analytical solutions. While explicitly working out the parameter regions for which the channel is positive, completely positive, entanglement-breaking and entanglement-annihilating, we have developed some new techniques to deal with the separability problem in quantum information. In particular, we have made some detailed observations on the Hadamard product in relation to entanglement theory, and showed how these tricks can be useful in proving separability of special states or classes of states. This led us to the solution of some outstanding problems in entanglement annihilation. Along the way, we also retrieved a simple example of a positive indecomposable map and constructed a notable entanglement-annihilating map.

The maps we consider naturally emerge in situations where two separate systems are subject to white noise and thus provide a useful tool in predicting the physical impact of noise on entanglement. Furthermore they provide new ways to reveal bound entanglement and an analytical characterisation of entanglement for a natural class of states. We hope that the observations we point

out elucidate certain facts in a way that is useful also for future applications. The solution of the depolarising channel now provides a promising basis for studying natural generalisations. First of all one should look at straightforward extensions to the multipartite case, which will involve many more parameters, but keep the local unitary symmetries (in fact the number of parameters scales exponentially in the number of parties only). Another potential path to pursue is to study the case of coloured noise. While this would in general lead to a number of parameters growing in the dimension, thermal noise in equidistant Hamiltonians would still give a physically important five parameter family that may yield tractable solutions.

Chapter 5

Gaussian entanglement revisited

5.1 Introduction

Quantum entanglement and the problem of its identification and characterisation has guided our investigation in Chapter 2 (see especially Section 2.4), as well as in Chapter 4, where we had to analyse the entanglement transformation properties of certain physically relevant maps. In the present chapter, instead, we give new insights into the separability problem for Gaussian states. These are special states that arise naturally, from both the theoretical and the experimental point of view, when one deals with continuous variable quantum systems, i.e. collections of quantum oscillators such as electromagnetic modes.

This chapter is organised as follows. As usual, the rest of this section is devoted to introducing the problem (Subsection 5.1.1) and to pointing the reader to our main original contributions (Subsection 5.1.2). In Section 5.2 we recall some basic technical facts concerning Schur complements and matrix means, and fix our notation for the continuous variable formalism. Section 5.3 is then devoted to establishing a simplified separability condition for Gaussian states (Theorem 5.5), something we will make extensive use of throughout in rest of the chapter. In Section 5.4 we revisit some known results in Gaussian entanglement theory by employing the tools introduced in the previous sections. In particular, we provide a new compact proof of the sufficiency of the PPT condition for separability of 1 vs n -mode Gaussian states (Theorem 5.6). Section 5.5 contains most of our novel findings, including an extension of the PPT-separability equivalence to bipartite Gaussian states that are invariant under the exchange of any two modes pertaining to one of the two parties (Theorem 5.12), and the solution to the problem [16] of characterising those Gaussian states that can not be made entangled by means of any passive operation (Theorem 5.14). Finally, Section 5.6 concludes the chapter with a brief summary and some future perspectives related to this work.

5.1.1 Gaussian separability problem

Continuous variable quantum systems [207, 208, 209, 210] play a major role in quantum information science, which should not come as a surprise for at least two good reasons. First, they are somewhat privileged by Nature, since the electromagnetic field is essentially a collection of harmonic oscillators, one for each frequency. This is of outstanding practical other than theoretical importance, since sending light via – for instance – an optical fibre is right now (and perhaps for many decades to come) the only feasible way to engineer quantum communication among distant facilities. Second, harmonic oscillators are among the few quantum systems that we know how to solve analytically (essentially since the very early days of quantum mechanics), which makes them good candidates for accurate experimental control.

There is another lesson that we can learn from the choice Nature made for the laws that govern the electromagnetic field. Namely, the Hamiltonian of said field turns out to be *quadratic in the field operators*. By employing relatively simple optical devices one can explore the whole set of such Hamiltonians, but it is not that easy to go out of this family and make the photons interact with each other directly. For this reason, it seems natural to expect that thermal states of quadratic Hamiltonians, also called *Gaussian states*, will be among the most common quantum states one can access in a laboratory. This intuition turns out to be correct, at least loosely speaking, and in fact Gaussian states are among the most widely employed quantum resources.

From the information theoretic standpoint, those states and the quantum channels that preserve their structure, called *Gaussian channels*, have been used to successfully implement paradigmatic platforms for continuous variable quantum information processing. Among the others, let us mention unconditional quantum teleportation in optical and atomic domains [211, 212, 213], quantum cryptography with coherent states [214], and sub-shot-noise interferometry in gravitational wave detectors [215, 216].

The epithet ‘Gaussian’ comes from the fact that, exactly as for their classical counterparts, i.e. multivariate Gaussian random variables, also quantum Gaussian states are entirely described by the expected values of the field operators, which together form a real vector, and the associated covariance matrix, called a *quantum covariance matrix* (QCM). For a system made of n electromagnetic modes, the QCM is a $2n \times 2n$ real matrix. The study of Gaussian states and of their entanglement properties is entirely ascribed to the characterisation of QCMs. For this reason, the corresponding separability problem can be studied with methods of matrix analysis, linear algebra and symplectic geometry, which come directly from *classical* mechanics [217].

The main question we address in this chapter is as follows. *Given a QCM V_{AB} , can we determine whether the corresponding bipartite Gaussian state ρ_{AB}^G is entangled or separable?* It is known [15] that this problem can be rephrased as a semidefinite programming (SDP), and thus admits an efficient algorithmic solution. In fact, an explicit and provably efficient algorithm that decides the problem has been put forward [218]. This is in stark contrast to what happens

for the separability problem in standard finite-dimensional quantum systems, which is known to be NP-hard under certain assumptions on the required accuracy [114, 115].

When the two local systems have a limited number of modes or when the state displays some further symmetry there are even closed solution available, in the form of necessary and sufficient separability criteria. The most famous of such criteria is undoubtedly the positive partial transpose (PPT) criterion [116], which turns out to be a powerful tool also in the analysis of continuous variable systems. In fact, in the seminal paper [14] it was proven that two-mode Gaussian states are separable iff they are PPT, while the year later the same result was shown to hold for Gaussian states of 1 vs n modes [15]. Since PPT entangled Gaussian states exist already for 2 vs 2 modes [15], this closes the problem of determining all the pairs (m, n) such that the PPT criterion is both necessary and sufficient in m vs n modes.

However, if further symmetries are imposed on the state, things may change. For instance, later it has been discovered that PPT-ness is again sufficient for ‘bi-symmetric’ Gaussian states, characterised by their invariance under local permutations of any two modes within any of the two subsystems [219], and for ‘isotropic’ Gaussian states, whose QCM has a fully degenerate ‘symplectic spectrum’ [220, 221, 222].

Finally, let us mention that in the Gaussian world there are more compelling reason than just mathematical convenience to look at the PPT criterion. Namely, it is known that for Gaussian states, while entanglement can never be distilled by Gaussian operations alone [223, 224, 225], entanglement distillability [226, 227] under general local operations assisted by classical communications is equivalent to violation of the PPT condition [218, 228].

5.1.2 Our contributions

The material presented in this chapter is part of the homonymous paper [229]:

- L. Lami, A. Serafini, and G. Adesso. Gaussian entanglement revisited. *Preprint arXiv:1612.05215*, 2016.

The objective I had in mind when I started to think about the problems discussed in this chapter was to simplify the presentation of the beautiful theory of Gaussian entanglement, which developed rapidly at the beginning of this century thanks to the works of many authors [14, 15, 218, 228, 230, 223, 224, 225, 222, 221, 16, 219, 231].¹ In particular, one of my immediate goals was to develop a mathematical formalism that would help shortening the somehow heterogeneous proof of the equivalence between separability and PPT-ness for 1 vs n -mode Gaussian states. Indeed, the way it developed historically, the argument is composed of two separate parts: first, Simon [14] tackled the case

¹Let me take the chance to thank Alessio Serafini for the incredibly stimulating course in quantum optics he taught at the Scuola Normale Superiore in Pisa in the academic year 2013-2014, which was of inspiration for the work we present here.

$n = 1$ by means of a direct construction; later, Werner and Wolf [15] were able to reduce the case of arbitrary n to that of $n = 1$ thanks to an ingenious application of symplectic geometry techniques.

This problem is our first testing ground. In Section 5.4 (Subsections 5.3 and 5.4.1) we show how to unify the two parts of the proof in single compact reasoning that is technically much simpler and furthermore works directly for the case of generic n (Theorem 5.6). Key to our proof is the intensive use of *Schur complements* [232], which have enjoyed applications in various areas of (Gaussian) quantum information theory [218, 222, 223, 224, 225, 233, 17, 234, 235, 236], and – as further reinforced by this work – may be appreciated as a mathematical cornerstone for continuous variable quantum theory. These mathematical tools are instrumental in establishing a simplified separability condition (Theorem 5.5) that is more easily handled than the original one in [15], at least in this context. In fact, the former requires convex optimisation over marginal covariance matrices on one subsystem only, yielding a significant simplification over the latter, which instead require optimisation on both. Among the other things, this new condition allows us to retrieve immediately the recently proven result that Gaussian states are separable iff they are completely extendible with Gaussian extensions [237].

Subsection 5.4.2 contains instead an alternative proof of the sufficiency of the PPT criterion for ‘isotropic’ Gaussian states, characterised by the property that their covariance matrix has a single distinct symplectic eigenvalue (Theorem 5.9). In the traditional approach, the sufficiency of PPT for their separability follows from a well known ‘mode-wise’ decomposition of pure-state covariance matrices [220, 221, 222], and from the fact that the covariance matrix of an isotropic state is just a multiple of the covariance matrix of a pure Gaussian state. On the contrary, main ingredients of this novel proof are advanced matrix analysis tools such as the operator geometric mean, already found to be useful in the context of quantum optics [235, 236]. Along the way, we prove a curious result on matrix means (Lemma 5.3) that to the best of our knowledge was not known before.

In Section 5.5 we show how to apply some of the methods discussed in this chapter to deduce some new results, rather than just new proofs of already established results. In Subsection 5.5.1 we show how to apply our simplified separability condition (Theorem 5.5) to prove that Gaussian states invariant under partial transposition are necessarily separable (Corollary 5.11), a result previously known only for the partial transposition of qubit subsystems [238]. We then show in Subsection 5.5.2, Theorem 5.9, that the 1 vs n -mode PPT-separability equivalence can be further extended to a class of arbitrary bipartite multimode Gaussian states that we call ‘mono-symmetric’, i.e. invariant under local exchanges of any two modes on one of the two subsystems (see Figure 5.1 for a pictorial representation). This result, which (to the best of our knowledge) is observed and proven here for the first time, generalises the case of bi-symmetric states studied in [219], providing as a by-product a simplified proof for the latter as well.

Finally, in Subsection 5.5.3 we consider the well-known class of Gaussian passive operations (i.e. the ones that preserve the average number of excitations of the input state, such as beam splitters and phase shifters), which play a central role in quantum optics [239, 209, 210], and we prove that a Gaussian state that always remains PPT under such a set of operations must also always stay separable. This novel result complements the seminal study of [16], in that the latter only considered the possibility of turning a PPT state into a non-PPT one through passive operations – essentially, the question of generating distillable entanglement – which is not the same as the question of generating inseparability, because as we mentioned Gaussian PPT bound entangled states do exist [15]. Here we settle the latter, more general and fundamental question. All the previous results enable us to substantially extend the range of equivalence between Gaussian separability and PPT in contexts of strong practical relevance. Last but not the least, we address the separability problem directly, and derive a novel simplified necessary and sufficient condition for Gaussian separability.

5.2 The toolbox: Schur complements, matrix means and Gaussian states

This section has the purpose of acquainting the reader with some of the technical tools to be employed later in the chapter, and in fact throughout the rest of the thesis. For this reason, we take the time to introduce some definitions and properties that will be useful later, even if they are not strictly needed for the derivation of this chapter’s results. In Subsection 5.2.1 we discuss Schur complements extensively, while Subsection 5.2.2 is dedicated to an essential introduction to the theory of matrix means. Subsection 5.2.3 is then devoted to Gaussian states and their fundamental properties.

5.2.1 Schur complements

One of the messages of the present paper is to lend further support to the fact that methods based on Schur complements can be successfully employed to derive fundamental results in continuous variable quantum information, following a streak of applications to various contexts including separability, distillability, steerability, entanglement monogamy, characterisation of Gaussian maps, and related problems [218, 222, 223, 224, 225, 233, 17, 234, 235, 236]. As a divertimento to set the stage, let us present a compact, essential compendium of such methods. Useful references on Schur complements are [198] or the monograph [232].

Given a square matrix M partitioned into blocks as

$$M = \begin{pmatrix} A & X \\ Y & B \end{pmatrix}, \tag{5.1}$$

the **Schur complement** of its (square, invertible) principal submatrix A , denoted by M/A , is defined as

$$M/A := B - YA^{-1}X. \quad (5.2)$$

We observe that the following factorisation formula holds:

$$M = \begin{pmatrix} \mathbf{1} & 0 \\ YA^{-1} & \mathbf{1} \end{pmatrix} \begin{pmatrix} A & 0 \\ 0 & B - YA^{-1}X \end{pmatrix} \begin{pmatrix} \mathbf{1} & A^{-1}X \\ 0 & \mathbf{1} \end{pmatrix}. \quad (5.3)$$

From this the determinant formula

$$\det M = (\det A)(\det M/A) \quad (5.4)$$

and the additivity of ranks

$$\text{rk } M = \text{rk } A + \text{rk } (M/A). \quad (5.5)$$

follow straightforwardly. Moreover, observe that if M is symmetric then (5.3) identifies a kind of normal form under congruence.

From a point of view of matrix analysis, Schur complements arise naturally when one wants to express the inverse of a block matrix in a compact form. Namely, for a matrix M partitioned as in (5.1) one can prove the useful formula [232]

$$M^{-1} = \begin{pmatrix} A^{-1} + A^{-1}X(M/A)^{-1}YA^{-1} & -A^{-1}X(M/A)^{-1} \\ -(M/A)^{-1}YA^{-1} & (M/A)^{-1} \end{pmatrix}. \quad (5.6)$$

Naturally, an analogous expression holds when one exchanges the roles of A and B and similarly those of X and Y . Incidentally, from this latter fact many useful matrix identities can be easily derived. For instance, observe that $(M/A)^{-1}$ is a submatrix of M^{-1} . A direct computation shows that its Schur complement within M^{-1} is $M^{-1}/(M/A)^{-1} = A^{-1}$.

Another useful property is *congruence covariance*. In fact, if $N = \begin{pmatrix} N_1 & 0 \\ 0 & N_2 \end{pmatrix}$ is decomposed conformally to the partition in (5.3), we find

$$(NMN^T) / (N_1^T AN_1) \geq N_2 (M/A) N_2^T, \quad (5.7)$$

for all square N_1, N_2 of appropriate size, with equality if N_1 is invertible.

The Schur complement notation is useful mainly because computations are more easily carried out when employing it. One of the most useful rules that this formalism obeys is the *quotient property*, that allows us to simplify concatenation of Schur complements as if they were ordinary fractions. Consider a matrix M as in (5.1), and take a square invertible sub-block A_1 of its first block A . Then A/A_1 turns out to be a square, invertible submatrix of M/A_1 , and moreover

$$M/A = (M/A_1)/(A/A_1). \quad (5.8)$$

On a different line, Schur complements are the answer to a number of questions that arise pretty naturally in linear algebra. Many of these applications stem from the fact that the positivity conditions of 2×2 Hermitian block matrices can be easily written in terms of Schur complements.

Lemma 5.1. *Consider a Hermitian matrix*

$$H = \begin{pmatrix} A & X \\ X^\dagger & B \end{pmatrix}. \quad (5.9)$$

Then H is strictly positive definite ($H > 0$) if and only if $A > 0$ and $H/A = B - X^\dagger A^{-1} X > 0$. Then, by taking suitable limits, H is semidefinite positive ($H \geq 0$) if and only if $A \geq 0$ and $B - X^\dagger(A + \epsilon \mathbf{1})^{-1} X \geq 0$ for all $\epsilon > 0$.

A consequence of this result that will be relevant to us is the following.

Corollary 5.2. *Let H be a Hermitian matrix partitioned as in (5.9). Then, if $A > 0$,*

$$H/A = \sup \left\{ C = C^\dagger : H > 0 \oplus C \right\}. \quad (5.10)$$

Here we mean that the matrix set on the right-hand side has a supremum (i.e. a minimum upper bound) with respect to the positive semidefinite partial order, and that this supremum is given by the Schur complement on the left-hand side.

We note in passing that from the above variational representation it follows immediately that H/A is monotonically non-decreasing and concave in $H > 0$. Consequently, $(H/A)^{-1}$ is monotonically non-increasing and convex in $H > 0$.

5.2.2 Matrix means

Somehow related to Schur complements are the so-called matrix means. As one might expect from their name, these are functions taking two positive matrices as inputs and yielding another positive matrix as output. For an excellent introduction to this topic, we refer the reader to [240, IV]. Given two strictly positive matrices $A, B > 0$, the simplest mean one can define is the **arithmetic mean** $(A + B)/2$, whose generalisation from scalars to matrices does not present difficulties. Another easily defined object is the **harmonic mean** [241, 242], denoted by $A!B$ and given by

$$A!B := \left(\frac{A^{-1} + B^{-1}}{2} \right)^{-1}. \quad (5.11)$$

Incidentally, the harmonic mean can be also defined as a Schur complement, with the help of the identity

$$A!B = A - A(A + B)^{-1}A = \begin{pmatrix} A & A \\ A & A + B \end{pmatrix} / (A + B), \quad (5.12)$$

which immediately implies that $A!B$ is monotone and jointly concave in A and B , i.e. concave in the pair (A, B) .

The least trivially defined among the elementary means is undoubtedly the **geometric mean** $A\#B$ between strictly positive matrices $A, B > 0$ [243, 242], which can be constructed as

$$A\#B := \max\{X = X^\dagger : A \geq XB^{-1}X\} \quad (5.13)$$

$$= \max\left\{X = X^\dagger : \begin{pmatrix} A & X \\ X & B \end{pmatrix} \geq 0\right\}, \quad (5.14)$$

where the above maximisation is as usual with respect to the positive semidefinite partial order (the fact that the particular set of matrices we chose admits an absolute maximum is already nontrivial), and (5.14) is equivalent to (5.13) thanks to Lemma 5.1. From the variational characterisation (5.14) we see that the geometric mean $A\#B$ is strictly monotonic in $A, B > 0$. Furthermore, with a bit of work one can show that $A\#B$ is explicitly given by [243]

$$A\#B = A^{1/2} \left(A^{-1/2} B A^{-1/2} \right)^{1/2} A^{1/2}. \quad (5.15)$$

Having multiple expressions for a single matrix mean is always useful, as some properties that are not easy to prove within one formulation may become apparent when a different approach is taken. For instance, the fact that $A\#B$ is covariant under congruences, i.e. [242, Corollary 2.1]

$$(MAM^\dagger)\#(MBM^\dagger) = M(A\#B)M^\dagger \quad \forall \text{ invertible } M, \quad (5.16)$$

is far from transparent if one looks at (5.15), while it becomes almost obvious when (5.13) is used. On the contrary, the fact that $A\#B = (AB)^{1/2}$ when $[A, B] = 0$ is not easily seen from (5.13), but it is readily verified employing (5.15). Incidentally, these two latter requirements (congruence invariance and reduction to standard geometric mean for commuting matrices) identify uniquely the expression (5.15) as the ‘correct’ matrix geometric mean. The expression for the determinant of the geometric mean, given by

$$\det(A\#B) = \sqrt{(\det A)(\det B)}, \quad (5.17)$$

is another instance of a property that is easily readable from the explicit formula (5.15).

The geometric mean is jointly concave in $A, B > 0$, which amounts to saying that the map $(A, B) \mapsto A\#B$ is concave. This is most easily seen from (5.14). Another property of the geometric mean that is relatively easy to derive using the variational principle (5.14) but not obvious at all from all the other expressions concerns its behaviour under positive maps. Namely, one can show that for all positive maps $\Lambda : \mathcal{H}_n \rightarrow \mathcal{H}_m$ (see Subsection 4.2.1) it holds that [242, Theorem 3]

$$\Lambda(A\#B) \leq \Lambda(A)\#\Lambda(B). \quad (5.18)$$

Now, let us discuss an interesting geometric interpretation of (5.15) that has drawn considerable attention in the mathematical community. For more

details we refer the reader to [240, VI]. The set of positive definite matrices of fixed size d can be seen as a manifold in the real space $\mathbb{R}^{d \times d}$. We can turn it into a Riemannian manifold by introducing on the tangent space the metric $ds^2 := \text{Tr}[(A^{-1}dA)^2]$ (sometimes called **trace metric**). It turns out the geodesic connecting two positive matrices A and B in this metric, parametrised by $t \in [0, 1]$, is given by

$$\gamma(t) = A^{1/2} \left(A^{-1/2} B A^{-1/2} \right)^t A^{1/2} =: A \#_t B, \quad (5.19)$$

sometimes called the **weighted geometric mean**. Confronting (5.15) and (5.19) we see in particular that $A \# B$ is nothing but the geodesic midpoint between A and B .

A consequence of this observation is that the monotonicity of the geometric mean under positive maps (5.18) extends to the weighted case as well, reading

$$\Lambda(A \#_t B) \leq \Lambda(A) \#_t \Lambda(B) \quad (5.20)$$

for all matrices $A, B > 0$, positive maps Λ , and $t \in [0, 1]$. To show the above inequality, just apply (5.18) iteratively to conclude that (5.20) holds at least when t is a dyadic rational. Then, continuity implies that it must hold indeed for all $t \in [0, 1]$. This standard reasoning is totally analogous to the one normally used to show that mid-point convexity and convexity are equivalent for continuous functions. Also the determinantal identity (5.17) can be easily generalised to this more general case. In fact, employing (5.19) it is readily seen that

$$\det(A \#_t B) = (\det A)^{1-t} (\det B)^t. \quad (5.21)$$

Of course, we can ask ourselves, how the geometric mean compares to the other means (arithmetic and harmonic) in the matrix setting. As it happens with scalars, the inequality

$$A!B \leq A \# B \leq \frac{A+B}{2} \quad (5.22)$$

holds true for all $A, B > 0$ [242, Corollary 2.1(iv)]. In view of the above inequality, it could be natural to wonder, whether the the geometric mean between the leftmost and rightmost sides of (5.22) has some relation with $A \# B$. That this could be a fruitful thought is readily seen by asking the same question for real numbers. In fact, when $0 < a, b \in \mathbb{R}$ it is elementary to verify that

$$\sqrt{ab} = \sqrt{\frac{a+b}{2} \left(\frac{1/a + 1/b}{2} \right)^{-1}}.$$

Our first result is a little lemma extending this to the non-commutative case. We were not able to find a proof in the literature, so we provide one.

Lemma 5.3. *For $A, B > 0$ strictly positive matrices, the identity*

$$A \# B = \left(\frac{A+B}{2} \right) \# (A!B) \quad (5.23)$$

holds true.

Proof. We start by defining

$$\begin{aligned}\tilde{A} &:= (A+B)^{-1/2} A (A+B)^{-1/2}, \\ \tilde{B} &:= (A+B)^{-1/2} B (A+B)^{-1/2}.\end{aligned}$$

It is easy to see that $[\tilde{A}, \tilde{B}] = 0$, for instance because

$$\tilde{A} + \tilde{B} = (A+B)^{-1/2} (A+B) (A+B)^{-1/2} = \mathbf{1}.$$

Therefore, the identity $\tilde{A}\#\tilde{B} = (\tilde{A}\tilde{B})^{1/2}$ holds. Now, on the one hand the congruence covariance of the geometric mean implies that

$$\begin{aligned}\tilde{A}\#\tilde{B} &= \left((A+B)^{-1/2} A (A+B)^{-1/2} \right) \# \left((A+B)^{-1/2} B (A+B)^{-1/2} \right) \\ &= (A+B)^{-1/2} (A\#B) (A+B)^{-1/2}.\end{aligned}$$

On the other hand,

$$\begin{aligned}\tilde{A}\tilde{B} &= (A+B)^{-1/2} A (A+B)^{-1} B (A+B)^{-1/2} \\ &= (A+B)^{-1/2} (B^{-1}(A+B)A^{-1})^{-1} (A+B)^{-1/2} \\ &= \frac{1}{2} (A+B)^{-1/2} (A!B) (A+B)^{-1/2}.\end{aligned}$$

Putting all together, we see that

$$\begin{aligned}(A+B)^{-1/2} (A\#B) (A+B)^{-1/2} &= \tilde{A}\#\tilde{B} \\ &= (\tilde{A}\tilde{B})^{1/2} \\ &= \frac{1}{\sqrt{2}} \left((A+B)^{-1/2} (A!B) (A+B)^{-1/2} \right)^{1/2}.\end{aligned}$$

Conjugating by $(A+B)^{1/2}$, we obtain

$$\begin{aligned}A\#B &= \frac{1}{\sqrt{2}} (A+B)^{1/2} \left((A+B)^{-1/2} (A!B) (A+B)^{-1/2} \right)^{1/2} (A+B)^{1/2} \\ &= \left(\frac{A+B}{2} \right) \# (A!B),\end{aligned}$$

where the last step is an application of (5.15). \square

5.2.3 Gaussian formalism

In the remainder of this section, we provide a brief introduction to the main concepts of the Gaussian formalism. Not all of what we say here will be needed for the sake of the present chapter. However, we will refer to this introduction for the rest of the thesis when dealing with Gaussian states. For further details we refer the reader to the textbook [210], whose notation and conventions we follow.

Canonical commutation relations

Quantum continuous variables describe quantum mechanics applied to an infinite-dimensional Hilbert space equipped with position and momentum operators x_i, p_j ($i, j = 1, \dots, n$) satisfying the so-called canonical commutation relations $[x_i, p_j] = i\delta_{ij}$ (in natural units, $\hbar = 1$). Such a Hilbert space describes, for instance, a collection of n quantum harmonic oscillators, or modes of the electromagnetic radiation field, with x_i, p_j being the non-commutative analogues of the classical electric and magnetic fields.

The operators x_i, p_j are often grouped together to form a single vector of $2n$ operators $r := (x_1, \dots, x_n, p_1, \dots, p_n)$. The canonical commutation relations then take the form

$$[r, r^T] = i\Omega := i \begin{pmatrix} 0 & \mathbf{1} \\ -\mathbf{1} & 0 \end{pmatrix}. \tag{5.24}$$

We can also consider a different block decomposition of the system, corresponding to the ordering $r = (x_1, p_1, \dots, x_n, p_n)$. In this new convention, one has

$$[r, r^T] = \omega^{\oplus n}, \quad \omega := \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}. \tag{5.25}$$

From now on, we will prefer the former to the latter convention, unless otherwise specified. We will refer to expressions written according to (5.25) as decomposed **mode-wise**.

An important object one can form is the displacement operator. For any $z \in \mathbb{R}^{2n}$, we define

$$D_z := e^{iz^T \Omega r}. \tag{5.26}$$

Gaussian states

As we said, Nature has a special preference for quadratic Hamiltonians. A prominent example is the free-field Hamiltonian $\mathcal{H}_0 = \frac{1}{2}r^T r$. For this reason, thermal states of quadratic Hamiltonians are extremely easily produced in the lab, in fact so easily that they deserve a special name, **Gaussian states**. As the name suggests, they can be fully described by a real displacement vector $w \in \mathbb{R}^{2n}$ and a real, $2n \times 2n$ **quantum covariance matrix** (QCM) V , defined respectively as

$$w := \text{Tr} [\rho r], \tag{5.27}$$

$$V := \text{Tr} [\rho \{r - w, r^T - w^T\}], \tag{5.28}$$

where the anticommutator $\{H, K\} := HK + KH$ is needed in the quantum case in order to make the above expression real, and $w := w \cdot \text{id}$ as operators on the Hilbert space.² As it turns out, quantum covariance matrices that come

²It is customary not to divide by 2 when defining the covariance matrix in the quantum case, the reason being that in this way (5.29) looks simpler.

out of (5.28) are exactly those real positive definite matrices that moreover satisfies the Heisenberg uncertainty relation [244]

$$V + i\Omega \geq 0. \quad (5.29)$$

The reason why some lower bound like (5.29) should hold is that, unlike in the classical case, in the quantum case not all the covariances can be small at the same time, and in particular not covariances that pertain to conjugate variables. Note that (5.29) can equivalently be written as $V - i\Omega \geq 0$ upon applying transposition (as $V^T = V$, $\Omega^T = -\Omega$). From now on we will often refer to real matrices satisfying (5.29) as QCMs.

The Gaussian state $\rho^G(V, w)$ with QCM V and displacement vector w admits the representation

$$\rho^G(V, w) = \int \frac{d^{2n}u}{(2\pi)^n} e^{-\frac{1}{4}u^T V u - iw^T u} D_{\Omega u}, \quad (5.30)$$

which justifies the alternative definition of Gaussian states as the continuous variable states associated with a Gaussian characteristic function. For more on representations related to (5.30), see [210, IV].

Bipartite systems

Here we are mostly interested in the entanglement properties of Gaussian states, so we need to consider bipartite systems. The QCM V_{AB} of a Gaussian state ρ_{AB}^G pertaining to a $(m+n)$ -mode bipartite system AB can be naturally written in block form according to the splitting between the subsystems A and B :

$$V_{AB} = \begin{pmatrix} V_A & X \\ X^T & V_B \end{pmatrix}. \quad (5.31)$$

According to the same splitting, the matrix Ω appearing in (5.24) takes the form

$$\Omega_{AB} = \begin{pmatrix} \Omega_A & 0 \\ 0 & \Omega_B \end{pmatrix} = \Omega_A \oplus \Omega_B, \quad (5.32)$$

with $\Omega_A = \omega^{\oplus m}$ and $\Omega_B = \omega^{\oplus n}$.

Deterministic dynamics

Clearly, linear transformations $r \rightarrow Sr$ that preserve the commutation relations (5.24) play a special role within this framework. Any such transformation is described by a **symplectic** matrix, i.e. a matrix S with the property that $S\Omega S^T = \Omega$. Symplectic matrices form a non-compact, connected Lie group that is additionally closed under transposition, and is typically denoted by $\text{Sp}(2n, \mathbb{R})$ [245]. The importance of these operations arises from the fact that for any symplectic S there is a unitary evolution U_S on the Hilbert space – that we call **symplectic unitary** – such that $U_S^\dagger r U_S = Sr$. Most importantly, such

a unitary is the product of a finite number of factors³ of the form $e^{i\mathcal{H}_Q}$, where \mathcal{H}_Q is a quadratic Hamiltonian, and as such it can be easily implemented in laboratory. Under conjugation by U_S , Gaussian states transform as

$$U_S^\dagger \rho^G(V, w) U_S = \rho^G(SV S^T, Sw) . \quad (5.33)$$

Unitary evolutions model the dynamics of closed quantum systems, but most systems one deals with in the laboratory interact with the environment, hence undergo some more general dynamics. A very general scenario that fits our Gaussian framework is as follows: first, the input state is put in contact with an external system in a reference Gaussian state; second, the joint system evolves with a symplectic unitary; and third, the ancillary system, or part of it, is traced away, i.e. physically discarded. What is left is known as a **Gaussian channel** [210, §5.3]. When acting on Gaussian states, the action of any such channel \mathcal{N} can be cast as an action on covariance matrix and displacement vector as follows:

$$\mathcal{N} : \begin{cases} V & \mapsto XVX^T + Y \\ w & \mapsto Xw + s \end{cases} , \quad (5.34)$$

where X, Y are matrices of appropriate size, and s is a real vector. In order for the channel \mathcal{N} to be completely positive, X and Y should satisfy the matrix inequality $Y + i\Omega \geq iX\Omega X^T$.

Non-deterministic dynamics

Of course, our action on a quantum system are not limited to letting it evolve. We can also access it through a measurement, and in our case, not surprisingly, **Gaussian measurements** are the most relevant. A Gaussian measurement can be thought of as a collection $\{E_t\}_{t \in \mathbb{R}^{2n}}$ of positive Gaussian operators of the form $E_t := \rho^G(\gamma, t)$, where γ is a QCM obeying (5.29), referred to as the **seed** of the measurement, and the notation is that of (5.30). As is easy to see, these operators form a valid quantum measurement, as they satisfy the normalisation condition $\int \frac{d^{2n}t}{(2\pi)^n} E_t = \mathbb{1}$. We can then ask ourselves what happens when one performs said measurement on the second subsystem B of a bipartite system AB that is initially in a Gaussian state $\rho_{AB} = \rho^G(V_{AB}, w_{AB})$. One can show that [210, §5.4.5]

$$\mathrm{Tr}_B [\mathbb{1} \otimes E_t \rho_{AB}] = p(t) \rho^G(\tilde{V}_A, \tilde{w}_A(t)) , \quad (5.35)$$

where $p(t)$ is the probability of getting the outcome t . Using the notation of (5.31), one has

$$p(t) = \frac{2^{n_B} e^{-(t-w_B)^T (V_B + \gamma)^{-1} (t-w_B)}}{\sqrt{\det(V_B + \gamma)}} , \quad (5.36)$$

$$\tilde{V}_A = V_A - X(V_B + \gamma)^{-1} X^T , \quad (5.37)$$

$$\tilde{w}_A(t) = w_A + X(t - w_B) . \quad (5.38)$$

³In fact, up to two such factors. I thank Uther P.F. Shackerley-Bennet and Alessio Serafini for bringing this to my attention.

Observe that (5.37) can be equivalently written as

$$\tilde{V}_A = (V_{AB} + 0 \oplus \gamma)/(V_B + \gamma) \quad (5.39)$$

with the help of the Schur complement formalism. Remarkably, the final QCM \tilde{V}_A does *not* depend on the outcome of the measurement t , as long as this is known. On the contrary, forgetting it would correspond to a partial trace over the system B , and in this case we would have $\tilde{V}_A = V_A$.

The most general Gaussian map we want to consider is an arbitrary concatenation of Gaussian channels and Gaussian partial measurements. Maps of this form act as follows on the covariance matrix and displacement vector of an input Gaussian state:

$$\Gamma_{1 \rightarrow 2} : \begin{cases} V_1 & \mapsto \gamma_2 - \delta_{12}^T (\gamma_1 + \Sigma V_1 \Sigma)^{-1} \delta_{12} \\ w_1 & \mapsto s_2 + \delta_{12}^T (\gamma_1 + \Sigma V_1 \Sigma)^{-1} \Sigma (w_1 + s_1) \end{cases}, \quad (5.40)$$

where $\gamma_{12} = \begin{pmatrix} \gamma_1 & \delta_{12} \\ \delta_{12}^T & \gamma_2 \end{pmatrix}$ is a QCM of the joint system 12, i.e. it satisfies (5.29) with $\Omega = \Omega_{12}$, s_{12} is a real vector, and Σ is the matrix that changes the sign of all the momenta, i.e.

$$\Sigma = \begin{pmatrix} \mathbb{1} & 0 \\ 0 & -\mathbb{1} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}^{\oplus n}, \quad (5.41)$$

with the rightmost expression being with respect to a mode-wise decomposition. The QCM γ_{12} that appears in (5.40) turns out to be that of the Choi state corresponding to the map $\Gamma_{1 \rightarrow 2}$. We will then refer to it as the **Choi covariance matrix** of $\Gamma_{1 \rightarrow 2}$. Observe that the first line of (5.40) inherits the Schur complement structure from (5.39), and can be rewritten accordingly as

$$\Gamma_{1 \rightarrow 2} : V_1 \mapsto (\gamma_{12} + \Sigma V_2 \Sigma)/(\gamma_1 + \Sigma V_1 \Sigma). \quad (5.42)$$

Although it may not be apparent from the above expressions, the action of any Gaussian channel, specified by (5.34), can also be written as the limit of expressions of the form (5.40), which makes (5.40) the most general physically allowed transformation we can think of within the Gaussian realm. For details and a derivation of these formulae, see [225, Eq. (10a) and (10b)] or [210, §5.5].

Williamson normal form

It turns out that all Gaussian states can be brought into a remarkably simple normal form via unitary transformations induced by quadratic Hamiltonians. In fact, a theorem by Williamson [246, 247] implies that for all strictly positive matrices $V > 0$ there is a symplectic transformation S such that

$$SVS^T := \begin{pmatrix} \Lambda & 0 \\ 0 & \Lambda \end{pmatrix}, \quad \Lambda = \text{diag}(\nu_1, \dots, \nu_n) > 0. \quad (5.43)$$

The diagonal elements $\nu_i > 0$, each taken with multiplicity one, are called symplectic eigenvalues of V , and are uniquely determined by V (up to their order,

which can be assumed decreasing by convention with no loss of generality). Accordingly, we will refer to $\vec{\nu} = (\nu_1, \nu_2, \dots, \nu_n)$ as the **symplectic spectrum** of V . Notably, Heisenberg uncertainty relation (5.29) can be conveniently restated as $\Lambda \geq \mathbb{1}$, or equivalently $\nu_i \geq 1$ for all $i = 1, \dots, n$.

A Gaussian state $\rho^G(V, w)$ can be shown to be pure if and only if all of its symplectic eigenvalues are equal to 1, which corresponds to the matrix equality $V\Omega V\Omega = -\mathbb{1}$. Correspondingly, a QCM V satisfying $\nu_j = 1$ for all $j = 1, \dots, n$ will be called a **pure QCM**. Equivalent conditions for a QCM V to be pure are: (i) $\det V = 1$; or (ii) $\text{rk}(V + i\Omega) = n$ (i.e. half the maximum). Note that pure QCMs V are themselves symplectic matrices, $V = (S^T S)^{-1} \in \text{Sp}(2n, \mathbb{R})$, and are the extremal elements in the convex set of QCMs. Mixed Gaussian states can be thought of as reduced states of a global Gaussian pure state by the addition of a (fictional) auxiliary system. The global pure state is called a **purification** of the original state. At the level of covariance matrices, this means that all QCMs can be seen as submatrices of symplectic matrices acting on a higher number of modes. For details, see Lemma 7.9.

Finally, note that for the sake of our problems the displacement vector w is often irrelevant since it can be made to vanish by local unitaries, resulting from the action of the displacement operator of (5.26) on each individual mode. Since non-local properties such as entanglement are invariant under local unitaries, all the results we are going to present will not depend on the first moments. Therefore, in what follows, we will completely specify any Gaussian state under our investigation as $\rho^G(V)$ in terms of its QCM V alone.

5.3 Simplified separability criterion for Gaussian states

One of the main motivations behind the work we present in this chapter was to revisit some standard results in the theory of Gaussian entanglement, possibly simplifying their proofs with the help of the methods of matrix analysis we discussed above. In order to do this, we first need to develop a new necessary and sufficient separability condition for Gaussian states, which is the purpose of this section.

As is apparent from the discussion above, the entanglement properties of a bipartite Gaussian state should admit a convenient translation at the level of QCMs. Recall that, in general, a bipartite quantum state ρ_{AB} is separable if and only if it can be written as a convex mixture of product states as in (4.6). For a Gaussian state ρ_{AB}^G of a bipartite continuous variable system, we have then the following.

Lemma 5.4. [15, Proposition 1]. *A Gaussian state $\rho_{AB}^G(V_{AB})$ with $(m + n)$ -mode QCM V_{AB} is separable if and only if there exist an m -mode QCM $\gamma_A \geq i\Omega_A$ and an n -mode QCM $\gamma_B \geq i\Omega_B$ such that*

$$V_{AB} \geq \gamma_A \oplus \gamma_B. \tag{5.44}$$

In view of the above result, a QCM V_{AB} satisfying (5.44) for some marginal QCMs γ_A, γ_B will itself be called *separable* from now on. The criterion in (5.44)

is necessary and sufficient for separability of QCMs, and can be evaluated numerically via convex optimisation [218, 233]. However, such optimisation runs over both marginal QCMs, hence can be sometimes difficult to handle theoretically.

The first main result of this paper is to show that the necessary and sufficient separability condition (5.44), for any m and n , can be further simplified. This result is quite neat and of importance in its own right. In particular, it allows us to recast the Gaussian separability problem as a convex optimisation over the marginal QCM of *one* subsystem only (say A without loss of generality), presumably resulting also in an appreciable reduction of computational resources, especially in case party A comprises a much smaller number of modes than party B .⁴

Theorem 5.5 (Simplified separability condition for an arbitrary QCM).
A QCM V_{AB} of $m+n$ modes is separable if and only if there exists an m -mode QCM $\gamma_A \geq i\Omega_A$ such that

$$V_{AB} \geq \gamma_A \oplus i\Omega_B. \quad (5.45)$$

In terms of the block form (5.31) of V_{AB} , when $V_B > i\Omega_B$ the above condition is equivalent to the existence of a real matrix γ_A satisfying

$$i\Omega_A \leq \gamma_A \leq V_A - X(V_B - i\Omega_B)^{-1}X^T. \quad (5.46)$$

If $V_B - i\Omega_B$ is not invertible, we require instead $i\Omega_A \leq \gamma_A \leq V_A - X(V_B + \epsilon\mathbb{1}_B - i\Omega_B)^{-1}X^T$ for all $\epsilon > 0$.

Proof. Since both sets of QCMs V_{AB} defined by (5.44) and (5.45) are clearly topologically closed, we can just show without loss of generality that their interiors coincide. This latter condition can be rephrased as an equivalence between the two following statements: (a) $V_{AB} > \gamma_A \oplus \gamma_B$ for some QCMs γ_A, γ_B ; and (b) $V_{AB} > \gamma_A \oplus i\Omega_B$ for some QCM γ_A .

Now, once $\gamma_A < V_A$ is fixed, the supremum of all the matrices γ_B satisfying $V_{AB} > \gamma_A \oplus \gamma_B$ is given by the Schur complement $(V_{AB} - (\gamma_A \oplus 0_B)) / (V_A - \gamma_A)$, as the variational characterisation (5.10) reveals. Therefore, statement (i) is equivalent to the existence of $i\Omega_A \leq \gamma_A < V_A$ such that $(V_{AB} - (\gamma_A \oplus 0_B)) / (V_A - \gamma_A) > i\Omega_B$. This is the same as to require $V_{AB} > \gamma_A \oplus i\Omega_B$, as the positivity conditions of Lemma 5.1 immediately show.

Until now, we have proven that the separability of V_{AB} can be restated as $V_{AB} \geq \gamma_A \oplus i\Omega_B$ for some appropriate QCM γ_A . Employing Lemma 5.1, we see that this is turn equivalent to (5.46), or to its ϵ -modified version when $V_B - i\Omega_B$ is not invertible. \square

It is worth noticing that both Lemma 5.4 and Theorem 5.5 extend straightforwardly to encompass the case of full separability of multipartite Gaussian

⁴We do not engage in a thorough comparison of the computational resources needed to decide (5.44) and (5.45), since our interest in (5.45) is mainly theoretical, as we shall see.

states. In the case of Lemma 5.4, this extension was already formulated in [15, 230]. As for Theorem 5.5, the corresponding necessary and sufficient condition for the full separability of a k -partite QCM $V_{A_1 \dots A_k}$ would read $V_{A_1 \dots A_k} \geq \gamma_{A_1} \oplus \dots \oplus \gamma_{A_{k-1}} \oplus i\Omega_{A_k}$ for appropriate QCMs $\gamma_{A_1}, \dots, \gamma_{A_{k-1}}$.

Remark. It has been recently observed [237] that condition (5.46) is equivalent to the corresponding Gaussian state $\rho_{AB}^G(V_{AB})$ with QCM V_{AB} being completely extendible with Gaussian extensions. We remind the reader that a bipartite state ρ_{AB} is said to be ‘completely extendible’ (Subsection 2.4.1) if for all k there exists a state $\rho_{AB_1 \dots B_k}$ that is: (i) symmetric under exchange of any two B_i systems; and (ii) an extension of ρ_{AB} in the sense that $\text{Tr}_{B_2 \dots B_k} \rho_{AB_1 \dots B_k} = \rho_{AB}$. When the original state ρ_{AB}^G is Gaussian, it is natural to consider extensions $\rho_{AB_1 \dots B_k}^G$ of Gaussian form as well. Interestingly enough, the above Theorem 5.5 provides a simple alternative proof of the remarkable fact (also proven in [237]) that Gaussian states are separable if and only if completely extendible with Gaussian extensions.

5.4 Sufficiency of the PPT condition – Revisited

In this section we revisit some of the classic results in Gaussian entanglement theory with the help of the mathematical tools we have developed so far, in particular Theorem 5.5. Namely, in Subsection 5.4.1 we present an alternative (and simpler) proof of the equivalence between separability and PPT-ness for 1 vs n -mode Gaussian states, while in Subsection 5.4.2 we tackle the same problem for isotropic Gaussian states.

5.4.1 PPT equals separability for 1 vs n -mode Gaussian states – Revisited

Most of the present chapter is devoted to the investigation of known and new conditions under which separability becomes equivalent to PPT for Gaussian states, so that the problem of deciding whether a given QCM is separable or not admits a handy formulation. For any bipartite state ρ_{AB} , recall that the PPT criterion (Subsection 4.2.1) provides a useful necessary condition for separability [116]: if a bipartite state ρ_{AB} is separable, then $\rho_{AB}^{T_B} \geq 0$, where the suffix T_B denotes transposition with respect to the degrees of freedom of subsystem B only. As already mentioned, in finite-dimensional systems, PPT is also a sufficient condition for separability in 2×2 and 2×3 quantum systems [117]. In continuous variable systems, the PPT criterion turns out to be also sufficient for separability of QCMs when either A or B is composed of one mode only [14, 15].

Theorem 5.6 (PPT is sufficient for Gaussian states of 1 vs n modes).

Let V_{AB} be a bipartite QCM such that either A or B are composed of one mode

only. Then V_{AB} is separable if and only if

$$V_{AB} \geq \begin{pmatrix} i\Omega_A & 0 \\ 0 & \pm i\Omega_B \end{pmatrix} = i\Omega_A \oplus (\pm i\Omega_B), \quad (5.47)$$

which amounts to the corresponding Gaussian state being PPT, $(\rho_{AB}^G)^{T_B} \geq 0$.

For completeness, we recall that the partial transpose of an $(m+n)$ -mode QCM V_{AB} , i.e. the covariance matrix of the partially transposed density operator $(\rho_{AB}^G)^{T_B}$, is given by

$$V_{AB}^{T_B} = \Theta V_{AB} \Theta, \quad (5.48)$$

where

$$\Theta := \mathbf{1}_A \oplus \Sigma_B \quad (5.49)$$

with the notation of (5.41) [14]. Accordingly, we can say that the QCM V_{AB} is PPT if and only if $V_{AB}^{T_B}$ is a valid QCM obeying (5.29), which is equivalent to (5.47) since $\Theta \Omega_{AB} \Theta = i\Omega_A \oplus (-i\Omega_B)$.

The original proof of Theorem 5.6 came in two steps. First, Simon [14] proved it in the particular case when both A and B are made of one mode only by performing an explicit analysis of the symplectic invariants of V_{AB} ; this seminal analysis is quite straightforward to follow and particularly instructive, but eventually a bit cumbersome, since it requires to distinguish between three cases, according to the sign of $\det X$, where X is the off-diagonal block of the QCM V_{AB} partitioned as in (5.31). Later on, Werner and Wolf [15] reduced the problem for the 1 vs n -mode case with arbitrary n to the 1 vs 1-mode case; the proof of this reduction is geometric in nature and rather elegant, but also relatively difficult.

Our purpose here is to use Schur complements to provide the reader with a simple, direct proof of Theorem 5.6. Before coming to that, there is a preliminary lemma we want to discuss.

Lemma 5.7. *Let M, N be 2×2 Hermitian matrices. There is a real symmetric matrix R satisfying $M \leq R \leq N$ if and only if $M \leq N, N^*$, where $*$ denotes complex conjugation.*

Proof. The necessity of the condition $M \leq N, N^*$ is easily verified, so we just prove sufficiency. The only complex entry in a 2×2 Hermitian matrix is in the off-diagonal element. Suppose without loss of generality that $\Im M_{12} \geq 0$ and $\Im N_{12} \leq 0$ (both conditions in the statement are in fact symmetric under complex conjugation of M or N). It is easy to verify that a p such that $0 \leq p \leq 1$ and $\Im(pM + (1-p)N)_{12} = 0$ always exists, and we see that $R := pM + (1-p)N$ is a real symmetric matrix. Moreover, since R belongs to the segment joining M and $N \geq M$ we conclude that $M \leq R \leq N$. \square

Remark. Lemma 5.7 admits an appealing physical interpretation which also leads to an intuitive proof. This interpretation is based on the fact that 2×2

Hermitian matrices can be seen as events in 4-dimensional Minkowski space-time through the correspondence $x_0\mathbf{1} + \vec{x} \cdot \vec{\sigma} \leftrightarrow (x_0, \vec{x})$. Furthermore, $M \leq N$ translates in Minkowski space-time to ‘ N is in the absolute future of M ’, since the remarkable determinantal identity $\det(x_0\mathbf{1} + \vec{x} \cdot \vec{\sigma}) = x_0^2 - \vec{x}^2$ holds true. Now, the complex conjugation at the matrix level becomes nothing but a spatial reflection with respect to a fixed spatial plane in Minkowski space-time. Thus, our original question is: is it true that whenever both an event N and its spatial reflection N^* are in the absolute future of a reference event M then there is another event R which is: (i) in the absolute future of M ; (ii) in the absolute past of both N and N^* ; and (iii) lies right on the reflection plane? The answer is clearly yes, and there is a simple way to obtain it. Start from M and shoot a photon to the location of that event between N and N^* that will happen on the other side of the reflection plane. After some time the photon hits the plane, and this event R clearly satisfies all requirements.

Now we are ready to give our direct proof of the equivalence between PPT and separability for 1 vs n -mode Gaussian states, leveraging the simplified separability condition of Theorem 5.5.

Proof of Theorem 5.6. Suppose without loss of generality that A is composed of one mode only. As in the proof of Theorem 5.5, since both sets of QCMs V_{AB} defined by (5.44) and (5.47) are topologically closed, we can assume that V_{AB} is in the interior of the PPT set, i.e. that $V_{AB} > i\Omega_A \oplus (\pm i\Omega_B)$. Our goal will be to show that in this case V_{AB} belongs to the separable set, as characterized by Theorem 5.5. Since $V_B - i\Omega_B$ is taken to be invertible, the PPT condition reads

$$V_A - X(V_B \mp i\Omega_B)^{-1}X^T \geq i\Omega_A.$$

Now, define $M = i\Omega_A$ and $N = V_A - X(V_B + i\Omega_B)^{-1}X^T$, and observe that $N^* = V_A - X(V_B - i\Omega_B)^{-1}X^T$. Thanks to Lemma 5.7, we can find a real matrix γ_A such that

$$V_A - X(V_B \mp i\Omega_B)^{-1}X^T \geq \gamma_A \geq i\Omega_A.$$

Choosing the negative sign in the above inequality, we see that the second condition (5.46) in Theorem 5.5 is met, and therefore V_{AB} is separable. \square

5.4.2 PPT equals separability for multimode isotropic Gaussian states – Revisited

It is well known that the PPT criterion is in general sufficient, as well as obviously necessary, for pure bipartite states to be separable [116]. This may be seen by a direct inspection of the Schmidt decomposition (4.3) of a pure state. Let us note, incidentally, that a stronger statement holds, namely that any bound entangled state (in any dimension) must have at least rank 4 [248].

The Schmidt decomposition theorem is in fact so important that a Gaussian version of it, that is, the determination of a normal form of pure QCMs under

local symplectic operations, is of central importance in continuous variable quantum information. As can be shown at the covariance matrix level [220, 222] or at the density operator level [221], every pure bipartite Gaussian state $\rho_{AB}^G(V_{AB})$ can be brought into a tensor product of two-mode squeezed vacuum states and single-mode vacuum states by means of local unitaries with respect to the A vs B partition. In particular, by acting correspondingly with local symplectic transformations, any pure QCM V_{AB} (where pure means $\det V_{AB} = 1$) can be transformed into a direct sum of (pure) two-mode squeezed vacuum QCMs and (pure) single-mode vacuum QCMs. More precisely, at the level of QCMs, one can formulate this fundamental result as follows.

Theorem 5.8 (Mode-wise decomposition of pure Gaussian states [220, 221, 222]). *Let V_{AB} be a bipartite QCM of $m + n$ modes $A_1, \dots, A_m, B_1, \dots, B_n$, assuming $m \leq n$ (with no loss of generality). If V_{AB} is a pure QCM, i.e. all its symplectic eigenvalues are equal to 1 (which amounts to $\det V_{AB} = 1$), then there exist local symplectic transformations $S_A \in \text{Sp}(2m, \mathbb{R})$, $S_B \in \text{Sp}(2n, \mathbb{R})$ mapping V_{AB} into the following normal form:*

$$(S_A \oplus S_B)V_{AB}(S_A^T \oplus S_B^T) = \bigoplus_{j=1}^m \bar{V}_{A_j B_j}(r_j) \oplus \bigoplus_{k=m+1}^n \mathbb{1}_{B_k}, \quad (5.50)$$

where $\bar{V}_{A_j B_j} = \begin{pmatrix} c_j \mathbb{1} & s_j \zeta \\ s_j \zeta & c_j \mathbb{1} \end{pmatrix}$ with $c_j = \cosh(2r_j)$ and $s_j = \sinh(2r_j)$, for a real squeezing parameter r_j , is the pure QCM of a two-mode squeezed vacuum state of modes A_j and B_j , and $\mathbb{1}_{B_k}$ is the pure QCM of the single-mode vacuum state of mode B_k . In particular, with respect to the block form (5.31), for any pure QCM V_{AB} the marginal QCMs V_A and V_B have matching symplectic spectra, given by $\vec{\nu}_A = (c_1, \dots, c_m)$ and $\vec{\nu}_B = (c_1, \dots, c_m, \underbrace{1, \dots, 1}_{n-m})$.

Leaving apart its far-reaching applications, in the context of the present paper this result is mainly instrumental for assessing the separability of so-called *isotropic* multimode Gaussian states. The QCM of any such state of $m + n$ modes is characterised by the property of having a completely degenerate symplectic spectrum, i.e. formed of only one distinct symplectic eigenvalue $\nu \geq 1$ (repeated $m + n$ times). This means that the QCM V_{AB} of any isotropic state is proportional by a factor ν to a pure QCM. Hence, Theorem 5.8 tells us that V_{AB} can be brought into a direct sum of two-mode QCMs via a local symplectic congruence (local with respect to any partition into groups of modes A and B), as first observed in [220]. Thanks to Theorem 5.6, this guarantees the following.

Theorem 5.9. *The PPT criterion is necessary and sufficient for separability of all isotropic Gaussian states of an arbitrary number of modes.*

However, notwithstanding the importance of Theorem 5.8 per se, one could strive to seek a more direct way to obtain Theorem 5.9. Our purpose in this

subsection is in fact to provide an alternative proof of this result, which does not appeal to the mode-wise decomposition theorem at all, and uses directly Lemma 5.4 instead, leveraging matrix analysis tools such as the notions of matrix means introduced in Section 5.2.2.

Let us start with a preliminary result, equivalent to [249, Proposition 12] or to [235, Lemma 13]. We include a proof for the sake of completeness.

Lemma 5.10. *Let $V > 0$ be a positive matrix. Then*

$$\gamma_V^\# := V \# (\Omega V^{-1} \Omega^T) \quad (5.51)$$

is a pure QCM. Furthermore, the following are equivalent:

- (a) *the matrix V satisfies Heisenberg's uncertainty principle (5.29);*
- (b) *$V \geq \Omega V^{-1} \Omega^T$; and*
- (c) *$V \geq \gamma_V^\#$.*

Moreover, V is a pure QCM iff condition (b) (equivalently, (c)) is saturated with equality, and $V > i\Omega$ iff condition (b) (equivalently, (c)) is a strict inequality.

Proof. Let the Williamson form of V be given by (5.43). Then we can write

$$\begin{aligned} \Omega V^{-1} \Omega^T &= \Omega S^{-T} (\Lambda^{-1} \oplus \Lambda^{-1}) S^{-1} \Omega^T \\ &\stackrel{(1)}{=} S \Omega (\Lambda^{-1} \oplus \Lambda^{-1}) \Omega^T S^T \\ &\stackrel{(2)}{=} S (\Lambda^{-1} \oplus \Lambda^{-1}) \Omega \Omega^T S^T \\ &\stackrel{(3)}{=} S (\Lambda^{-1} \oplus \Lambda^{-1}) S^T, \end{aligned}$$

where we used in order: (1) the identities $\Omega S^{-T} = S \Omega$, $S^{-1} \Omega^T = \Omega^T S^T$, all consequences of the defining symplectic identity $S \Omega S^T = \Omega$; (2) the fact that Ω commutes with $\Lambda^{-1} \oplus \Lambda^{-1}$; and (3) the orthogonality relation $\Omega \Omega^T = \mathbb{1}$. Now the first claim becomes obvious, since $\Lambda = \Lambda^{-1} = \mathbb{1}$ if and only if V is a pure QCM. In general, as it can be seen from the above expression, V and $\Omega V^{-1} \Omega^T$ are brought in Williamson form by simultaneous congruences with the same symplectic matrix S .

Observe that Heisenberg's uncertainty principle (5.29), which can be rephrased as $\Lambda \geq \mathbb{1}$, translates directly to $\Lambda \geq \Lambda^{-1}$ since $\Lambda > 0$. Via the above calculation, this leads to $V = S(\Lambda \oplus \Lambda)S^T \geq S(\Lambda^{-1} \oplus \Lambda^{-1})S^T = \Omega V^{-1} \Omega^T$. We have then shown the equivalence between claims (a) and (b).

As for (c), the covariance of the geometric mean under congruence ensures that

$$\begin{aligned} V \# (\Omega V^{-1} \Omega^T) &= \left(S \begin{pmatrix} \Lambda & 0 \\ 0 & \Lambda \end{pmatrix} S^T \right) \# \left(S \begin{pmatrix} \Lambda^{-1} & 0 \\ 0 & \Lambda^{-1} \end{pmatrix} S^T \right) \\ &= S \left(\begin{pmatrix} \Lambda & 0 \\ 0 & \Lambda \end{pmatrix} \# \begin{pmatrix} \Lambda^{-1} & 0 \\ 0 & \Lambda^{-1} \end{pmatrix} \right) S^T \\ &= S S^T, \end{aligned}$$

where the last passage is an easy consequence of the fact that $A\#A^{-1} = \mathbb{1}$ for all $A > 0$. Again, upon congruence by S Heisenberg's uncertainty relation $\Lambda \geq \mathbb{1}$ becomes $V = S(\Lambda \oplus \Lambda)S^T \geq SS^T = V\#(\Omega V^{-1}\Omega^T)$, which proves the equivalence between (a) and (c). Finally, the last claims follow easily from the above reasoning. \square

Now we are ready to explain our direct argument to show separability of PPT isotropic Gaussian states, alternative to the use of the mode-wise decomposition.

Proof of Theorem 5.9. Thanks to (5.48) and (5.49), the PPT condition (5.47) for a QCM V_{AB} takes the form $\Theta V \Theta \geq i\Omega$, where Θ is defined in (5.49). Thanks to Lemma 5.10, this becomes in turn

$$\Theta V \Theta \geq (\Theta V \Theta)\#(\Omega \Theta V^{-1} \Theta \Omega^T)$$

and finally

$$V \geq V\#(\Theta \Omega \Theta V^{-1} \Theta \Omega^T \Theta) = (gV)\#(Z \Omega (gV)^{-1} \Omega^T Z)$$

after conjugating by Θ , applying once more the covariance of the geometric mean under congruences, introducing a real parameter $g > 0$ (to be fixed later), and defining $Z := \mathbb{1}_A \oplus (-\mathbb{1}_B) = \Theta \Omega \Theta^T$. Now, we apply Lemma 5.3 to the above expression, obtaining

$$V \geq \left(\frac{gV + Z \Omega (gV)^{-1} \Omega^T Z}{2} \right) \# ((gV)! (Z \Omega (gV)^{-1} \Omega^T Z))$$

Although it is not yet transparent, we are done, as the right-hand side of the above inequality is exactly of the form $\gamma_A \oplus \gamma_B$ when V is the QCM of an isotropic Gaussian state. In fact, let $g > 0$ be such that gV is a pure QCM, satisfying $gV = \Omega (gV)^{-1} \Omega^T = \begin{pmatrix} P & Y \\ Y^T & Q \end{pmatrix}$, where the block decomposition is with respect to the A vs B splitting. Then on the one hand since $Z = \mathbb{1}_A \oplus (-\mathbb{1}_B)$ we find

$$\frac{gV + Z \Omega (gV)^{-1} \Omega^T Z}{2} = \begin{pmatrix} P & 0 \\ 0 & Q \end{pmatrix},$$

while on the other hand

$$\begin{aligned} (gV)! (Z \Omega (gV)^{-1} \Omega^T Z) &= 2 \left((gV)^{-1} + Z \Omega (gV) \Omega^T Z \right)^{-1} \\ &= 2 \Omega \left(\Omega (gV)^{-1} \Omega^T + Z (gV) Z \right)^{-1} \Omega^T \\ &= 2 \Omega (gV + Z (gV) Z)^{-1} \Omega^T \\ &= \begin{pmatrix} \Omega P^{-1} \Omega^T & 0 \\ 0 & \Omega Q^{-1} \Omega^T \end{pmatrix}, \end{aligned}$$

where we used the definition (5.11) of harmonic mean and the easily verified fact that $[Z, \Omega] = 0$. Putting all together, we find

$$\begin{aligned} V &\geq \begin{pmatrix} P & 0 \\ 0 & Q \end{pmatrix} \# \begin{pmatrix} \Omega P^{-1} \Omega^T & 0 \\ 0 & \Omega Q^{-1} \Omega^T \end{pmatrix} \\ &= \begin{pmatrix} P \# \Omega P^{-1} \Omega^T & 0 \\ 0 & Q \# \Omega Q^{-1} \Omega^T \end{pmatrix} \\ &= (\gamma_P^\#)_A \oplus (\gamma_Q^\#)_B, \end{aligned}$$

where the notation is as in (5.51). Since Lemma 5.10 ensures that $\gamma_P^\# = P \# (\Omega P^{-1} \Omega^T)$ is a QCM since $P > 0$ (and analogously for Q), a direct invocation of Lemma 5.4 allows us to conclude the proof. \square

5.5 Novel results on Gaussian entanglement

While in the previous Section 5.4 we reviewed classic results and provided new proof of them, our purpose in this section is to apply some of our methods to gain some new insights into the nature of Gaussian entanglement. Namely, in Subsection 5.5.1 we focus on Gaussian states that are not only PPT but even invariant under partial transposition, and show that they are necessarily separable, which mimics an analogous result found in the finite-dimensional setting in [238]. Next, Subsection 5.4.2 is devoted to extending Theorem 5.6 to the more general case of a bipartite Gaussian state of m vs n modes which is invariant under the exchange of any two modes of one of the two subsystems. Finally, in Subsection 5.5.3 we close an open problem of [16], by showing that bipartite Gaussian states that remain PPT when an arbitrary passive operation is applied to them are necessarily separable.

5.5.1 Gaussian states that are invariant under partial transpose are separable

As an example of straightforward application of Theorem 5.5, we study here the separability of a special class of PPT Gaussian states, i.e. those that are *invariant* under partial transposition of one of the subsystems. This problem has an analogue in finite-dimensional quantum information, already studied in [238], where it was shown that bipartite states on $\mathbb{C}^2 \otimes \mathbb{C}^N$ that are invariant under partial transpose on the first system are necessarily separable.⁵ Here we show that for Gaussian states an even stronger statement holds, in that

⁵The proof reported in [238] is rather long, so here we provide a shorter one, again based on Schur complements. A state on $\mathbb{C}^2 \otimes \mathbb{C}^N$ that is invariant under partial transposition on the first subsystem can be represented in block form as $\rho = \begin{pmatrix} A & X \\ X & B \end{pmatrix}$. By a continuity argument, we can suppose without loss of generality that $A > 0$. Rewrite $\rho = \begin{pmatrix} A & X \\ X & X A^{-1} X \end{pmatrix} + |1\rangle\langle 1| \otimes (B - X A^{-1} X)$. Both terms are positive by Lemma 5.1. Since the second one is separable, let us deal only with the first one, call it $\tilde{\rho}$. We have $\tilde{\rho} = A^{1/2} \otimes A^{1/2} \begin{pmatrix} \mathbf{1} & Y \\ Y & Y^2 \end{pmatrix} A^{1/2} \otimes A^{1/2}$, where

invariance under partial transposition implies separability for any number of local modes.

Corollary 5.11. *A bipartite Gaussian state ρ_{AB}^G that is invariant under partial transposition of one of the two subsystems is necessarily separable.*

Proof. Without loss of generality, we can assume that the partial transpose on the B system leaves the state invariant. We now show that under this assumption the separability condition (5.46) is immediately satisfied, since the rightmost side is already a real, symmetric matrix. In fact, equating the original QCM (5.31) with the one obtained after partial transpose on the B system, according to (5.48) and (5.49), we get the identities $X = X\Sigma$ and $V_B = \Sigma V_B \Sigma$, with Σ defined by (5.41). As a consequence,

$$\begin{aligned} X(V_B - i\Omega_B)^{-1}X^T &= X\Sigma(V_B - i\Omega_B)^{-1}\Sigma X^T \\ &= X(\Sigma(V_B - i\Omega_B)\Sigma)^{-1}X^T \\ &= X(V_B + i\Omega_B)^{-1}X^T, \end{aligned}$$

where we used also $\Sigma\Omega_B\Sigma = -\Omega_B$. This shows that $X(V_B - i\Omega_B)^{-1}X^T$ is equal to its complex conjugate, and is therefore (despite appearances) a real symmetric matrix. Hence the separability condition (5.46) is satisfied with $\gamma_A = V_A - X(V_B - i\Omega_B)^{-1}X^T$. \square

5.5.2 PPT equals separability for multimode mono-symmetric Gaussian states

Throughout this Section, we show how the PPT criterion is also necessary and sufficient for deciding the separability of bipartite Gaussian states of m vs n modes that are symmetric under the exchange of any two among the first m modes. These states will be referred to as **mono-symmetric** (with respect to the first party A). As can be easily seen, this novel result (see Figure 5.1 for a graphical visualisation) is a generalisation of both Theorem 5.6 and of one of the main results in [219], where the subclass of bi-symmetric states was considered instead, bi-symmetric meaning that they are invariant under swapping any two modes either within the first m or within the last n (that is, they are mono-symmetric in both A and B).

Theorem 5.12. *Let $\rho_{AB}^G(V_{AB})$ be a mono-symmetric Gaussian state of $m+n$ modes, i.e. specified by a QCM V_{AB} that is symmetric under the exchange of any two of the m modes of subsystem A . Then there exists a local unitary operation on A corresponding to a symplectic transformation $S_A \in \text{Sp}(2m, \mathbb{R})$ that*

$Y := A^{-1/2}X A^{-1/2}$ is Hermitian. Denoting by $Y = \sum_i y_i |e_i\rangle\langle e_i|$ its spectral decomposition, we obtain the following manifestly separable representation of $\tilde{\rho}$:

$$\tilde{\rho} = A^{1/2} \otimes A^{1/2} \left(\sum_i \begin{pmatrix} 1 & y_i \\ y_i & y_i^2 \end{pmatrix} \otimes |e_i\rangle\langle e_i| \right) A^{1/2} \otimes A^{1/2}.$$

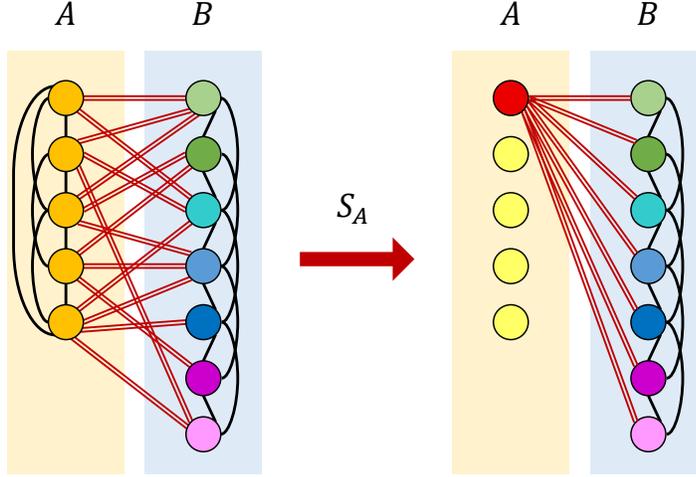


Figure 5.1: Mono-symmetric Gaussian states of two parties A (with m modes) and B (with n modes) are invariant under exchange of any two modes within party A . By means of a suitable symplectic transformation on subsystem A , these states can be reduced to a 1 vs n -mode Gaussian state and a collection of $m - 1$ uncorrelated single-mode states on A 's side. Since PPT is equivalent to separability for 1 vs n -mode Gaussian states, it follows that PPT is necessary and sufficient for separability of all m vs n -mode mono-symmetric Gaussian states. In the schematics, entanglement between pairs of modes from the same party is depicted as a single solid (black) line, while entanglement across a mode from A and a mode from B is depicted as a double (dark red) line.

transforms ρ_{AB}^G into the tensor product of $m-1$ uncorrelated single-mode Gaussian states $\tilde{\rho}_{A_j}^G(\tilde{V}_{A_j})$ ($j = 2, \dots, m$) and a bipartite Gaussian state $\tilde{\rho}_{A_1B}^G(\tilde{V}_{A_1B})$ of 1 vs n modes. At the QCM level, this reads

$$(S_A \oplus \mathbb{1}_B)V_{AB}(S_A^T \oplus \mathbb{1}_B) = \left(\bigoplus_{j=2}^m \tilde{V}_{A_j} \right) \oplus \tilde{V}_{A_1B}, \quad (5.52)$$

the above direct sum being with respect to a mode-wise decomposition. The separability properties of V_{AB} and \tilde{V}_{A_1B} are equivalent, in particular $\rho_{AB}^G(V_{AB})$ is separable if and only if it is PPT.

Proof. Throughout this proof, we resort to a mode-wise decomposition, which is more apt to capture the symmetry of the problem. We will prove (5.52) directly at the QCM level, by constructing a suitable local symplectic S_A . By virtue of the symmetry under the exchange of any two modes of subsystem A , if we decompose V_{AB} as in (5.31), the submatrices V_A and X have the following

structure (with respect to a mode-wise decomposition):

$$V_A = \begin{pmatrix} \alpha & \varepsilon & \dots & \varepsilon \\ \varepsilon & \alpha & & \vdots \\ \vdots & & \ddots & \varepsilon \\ \varepsilon & \dots & \varepsilon & \alpha \end{pmatrix}, \quad X = \begin{pmatrix} \kappa_1 & \kappa_2 & \dots & \kappa_n \\ \kappa_1 & \kappa_2 & \dots & \kappa_n \\ \vdots & & & \vdots \\ \kappa_1 & \kappa_2 & \dots & \kappa_n \end{pmatrix}, \quad (5.53)$$

where each one of the blocks $\alpha, \varepsilon, \kappa_j$ in (5.53) is a 2×2 real matrix, with α and ε symmetric [250].

We can now decompose the real space of the first m modes as $\mathbb{R}^{2m} = \mathbb{R}^m \otimes \mathbb{R}^2$. According to this decomposition, we may rewrite V_A and X as follows:

$$V_A = \mathbf{1}_m \otimes (\alpha - \varepsilon) + m |+\rangle\langle+| \otimes \varepsilon, \quad X = \sqrt{m} \sum_{j=1}^n |+\rangle\langle j| \otimes \kappa_j, \quad (5.54)$$

where $|+\rangle = \frac{1}{\sqrt{m}} \sum_{i=1}^m |i\rangle$, with $\{|i\rangle\}_{i=1}^m$ denoting the standard basis for \mathbb{R}^m . Observe that the symplectic form Ω_A on subsystem A decomposes accordingly as $\Omega_A = \mathbf{1}_m \otimes \omega$. If O is an $m \times m$ orthogonal matrix such that $O|+\rangle = |1\rangle$, we easily see that on the one hand $O \otimes \mathbf{1}_2 \Omega_A O^T \otimes \mathbf{1}_2 = \Omega_A$, i.e. $O \otimes \mathbf{1}_2$ is symplectic, while on the other hand

$$\begin{aligned} O \otimes \mathbf{1}_2 V_A O^T \otimes \mathbf{1}_2 &= |1\rangle\langle 1| \otimes (\alpha + (m-1)\varepsilon) + \sum_{i=2}^m |i\rangle\langle i| \otimes (\alpha - \varepsilon) \\ &= \begin{pmatrix} \alpha + (m-1)\varepsilon & 0 & \dots & 0 \\ 0 & \alpha - \varepsilon & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \dots & 0 & \alpha - \varepsilon \end{pmatrix} \end{aligned}$$

and

$$\begin{aligned} O \otimes \mathbf{1}_2 X &= \sqrt{m} \sum_{j=1}^n |1\rangle\langle j| \otimes \kappa_j \\ &= \begin{pmatrix} \sqrt{m} \kappa_1 & \sqrt{m} \kappa_2 & \dots & \sqrt{m} \kappa_n \\ 0 & 0 & \dots & 0 \\ \vdots & & & \vdots \\ 0 & 0 & \dots & 0 \end{pmatrix}. \end{aligned}$$

Therefore, the initial QCM V_{AB} has been decomposed as a direct sum of $m-1$ one-mode QCMs $\tilde{V}_{A_j} = \alpha - \varepsilon$, and of one $(1+m)$ -mode QCM $\tilde{V}_{A_{1B}}$, via a local symplectic operation on subsystem A , i.e. $S_A = O \otimes \mathbf{1}_2$. This proves (5.52) constructively. Applying Theorem 5.6, one then gets immediately that the PPT condition is necessary and sufficient for separability in this case. \square

This original result yields a substantial enlargement to the domain of validity of PPT as a necessary and sufficient criterion for separability of multimode Gaussian states, reaching beyond any existing literature. In practice, Theorem 5.12 tells us that, in any mono-symmetric Gaussian state, all the correlations (including and beyond entanglement) shared among the whole m modes of A and the whole n modes of B can be localised onto correlations between a single mode A_1 of A vs the whole B , by means of a local unitary (symplectic at the QCM level) operation at A 's side only. Being unitary, this operation is fully reversible, meaning that the correlations with B can be redistributed back and forth between A_1 and the whole set of A modes with no information loss. This also means that quantitative results on any measure of such correlations between A and B encoded in V_{AB} can be conveniently evaluated in the much simpler 1 vs n -mode normal form \tilde{V}_{A_1B} constructed in the proof Theorem 5.12, ignoring the $m - 1$ uncorrelated modes.

In the special case of V_{AB} being the QCM of a bi-symmetric state, i.e. with full permutation symmetry within both A and B , it is immediate to observe that applying a similar construction by means of a local unitary at B 's side as well fully reduces V_{AB} to a two-mode QCM $\tilde{V}_{A_1B_1}$, with equivalent entanglement properties as the original V_{AB} , plus a collection of $m + n - 2$ uncorrelated single modes. This reproduces the findings of [219].

Similarly to what discussed in the Remark at the end of Subsection 5.3, the results of Theorem 5.12 can also be straightforwardly extended to characterise full separability and, conversely, multipartite entanglement of arbitrary multimode Gaussian states which are partitioned into k subsystems, with the requirement of local permutation invariance within some of these subsystems. It is clear that, by suitable local symplectic transformations, each of those locally symmetric parties can be localised onto a single mode correlated with the remaining parties, thus removing the redundancy in the QCM. Gaussian states of this sort generalise the so-called multi-symmetric states studied in [251], where local permutation invariance was enforced within all of the subsystems, resulting in a direct multipartite analogue of bi-symmetric states.

5.5.3 Entangling Gaussian states via passive optical operations

Throughout this subsection, we finally complete the solution of a problem posed in [16] and there addressed under some additional constraints. Let us start by recalling that symplectic operations can be divided into two main categories, namely those such as squeezers that require an exchange of energy between the system and the apparatus, called **active**, and those that can be implemented using only beam splitters and phase plates, called **passive**. A symplectic matrix K represents a passive transformation if and only if it is also *orthogonal*, meaning that $KK^T = \mathbb{1}$ (it may be worth adding that symplectic orthogonal transformations form the maximal compact subgroup of the symplectic group). As it turns out, symplectic orthogonal matrices can be represented in an especially simple form if we resort to a position-momentum block decomposition.

Namely, one has the parametrisation [210]

$$K = W^\dagger \begin{pmatrix} U & \\ & U^* \end{pmatrix} W, \quad (5.55)$$

where

$$W := \frac{1}{\sqrt{2}} \begin{pmatrix} \mathbf{1} & i\mathbf{1} \\ \mathbf{1} & -i\mathbf{1} \end{pmatrix}$$

and U is a generic, $n \times n$ unitary matrix, with U^* denoting its complex conjugate (as usual, n is the number of modes).

Since the implementation of passive operations is so inexpensive in quantum optics and entangled states so useful for quantum technologies, the question first posed in [16] was a natural one: *what bipartite Gaussian states are such that they can be entangled via a global, passive operation?* However, in this full generality the problem was left unanswered in [16]. Instead, another related question was investigated and answered there, namely whether *distillable* Gaussian entanglement can be produced in the same fashion. For Gaussian states, as mentioned in Subsection 5.1.1, distillability is well known to be equivalent to non-positivity of the partial transpose [218, 228], so the authors of [16] proceeded to identify the class of Gaussian states that can be made to violate the PPT condition by means of a passive transformation. However, it is important to realise that since PPT and separability are not the same for general multi-mode Gaussian states, the two questions are a priori different. Here we show that the answer to the original question above turns out to be yet another situation where the PPT condition is necessary and sufficient to ensure separability of Gaussian states. In other words, we will prove that a bipartite Gaussian state that can not be made distillable (i.e. non-PPT) via passive operations is necessarily separable, and thus it stays separable under the application of said passive operations. Let us start with a technical lemma that we deduce from recent results obtained in [252].

Lemma 5.13. *Let $A > 0$ be a strictly positive $2n \times 2n$ matrix. Let $\nu_i(A)$ and $\lambda_i(A)$ denote its symplectic and ordinary (orthogonal) eigenvalues, respectively, arranged in nondecreasing order. Then*

$$\nu_1(A)^2 \geq \lambda_1(A)\lambda_2(A). \quad (5.56)$$

In particular, every positive matrix whose two smallest eigenvalues obey the inequality $\lambda_1\lambda_2 \geq 1$ is automatically a legitimate QCM.

Proof. From [252, Eq. (17)] we deduce $\prod_{j=i}^k \nu_j(A)^2 \leq \prod_{j=1}^{2k} \lambda_j(A)$ for all $k = 1, \dots, n$. Choosing $k = 1$ yields the claim. \square

Now, we are ready to present our strengthening of [16, Proposition 1].

Theorem 5.14. *Let V be a bipartite QCM of an n -mode system. Then the following are equivalent:*

- (a) KVK^T is separable for all Gaussian passive transformations K ;
 (b) KVK^T is PPT for all Gaussian passive transformations K ; and
 (c) the two smallest eigenvalues of V satisfy $\lambda_1(V)\lambda_2(V) \geq 1$.

Proof. The implication (a) \Rightarrow (b) is obvious, while (c) \Rightarrow (b) already follows from Lemma 5.13 together with the fact that the partial transpose at the level of QCMs is a congruence by orthogonal transformation and thus does not change the ordinary spectrum. One of the main contributions of [16] is the proof that (b) and (c) are in fact equivalent. In view of this discussion, we have just to show that (c) \Rightarrow (a). To this end, we will assume that V satisfies $\lambda_1(V)\lambda_2(V) \geq 1$ and construct two local QCMs γ_A, γ_B that satisfy the hypothesis of the original separability criterion given by Lemma 5.4. Call $\lambda_1(V) = k$ and observe that if $k \geq 1$ then $V \geq \mathbb{1} = \mathbb{1}_A \oplus \mathbb{1}_B$ and we are done. Otherwise, assume $k < 1$ and denote by $|x\rangle$ the normalised eigenvector corresponding to the minimal eigenvalue of V , i.e. $V|x\rangle = k|x\rangle$ and $\langle x|x\rangle = 1$. Since $\lambda_2(V) \geq \frac{1}{k}$ and a fortiori $\lambda_i(V) \geq \frac{1}{k}$ for all $i \geq 2$, we can write

$$V \geq k|x\rangle\langle x| + \frac{1}{k}(\mathbb{1} - |x\rangle\langle x|).$$

Now, decompose the vector $|x\rangle$ into its A and B components as $|x\rangle = \begin{pmatrix} \sqrt{p}|y\rangle_A \\ \sqrt{1-p}|z\rangle_B \end{pmatrix}$, where $0 \leq p \leq 1$ and $\langle y|y\rangle = 1 = \langle z|z\rangle$. Then, Lemma 5.13 guarantees that the matrices

$$\begin{aligned} \gamma_A &:= k|y\rangle\langle y| + \frac{1}{k}(\mathbb{1} - |y\rangle\langle y|) \\ \gamma_B &:= k|z\rangle\langle z| + \frac{1}{k}(\mathbb{1} - |z\rangle\langle z|) \end{aligned}$$

are legitimate QCMs. For this reason, showing that $V_{AB} - \gamma_A \oplus \gamma_B \geq 0$ would complete our proof. By direct computation, we find

$$\begin{aligned} V_{AB} - \gamma_A \oplus \gamma_B &= -\left(\frac{1}{k} - k\right) \begin{pmatrix} p|y\rangle\langle y| & \sqrt{p(1-p)}|y\rangle\langle z| \\ \sqrt{p(1-p)}|z\rangle\langle y| & (1-p)|z\rangle\langle z| \end{pmatrix} \\ &\quad + \frac{1}{k} \begin{pmatrix} \mathbb{1} & 0 \\ 0 & \mathbb{1} \end{pmatrix} \\ &\quad - \begin{pmatrix} -\left(\frac{1}{k} - k\right)|y\rangle\langle y| + \frac{1}{k}\mathbb{1} & 0 \\ 0 & -\left(\frac{1}{k} - k\right)|z\rangle\langle z| + \frac{1}{k}\mathbb{1} \end{pmatrix} \\ &= \left(\frac{1}{k} - k\right) \begin{pmatrix} (1-p)|y\rangle\langle y| & -\sqrt{p(1-p)}|y\rangle\langle z| \\ -\sqrt{p(1-p)}|z\rangle\langle y| & p|z\rangle\langle z| \end{pmatrix} \\ &= \left(\frac{1}{k} - k\right) (\sqrt{1-p}|y\rangle \quad \sqrt{p}|z\rangle)^T (\sqrt{1-p}|y\rangle \quad \sqrt{p}|z\rangle) \\ &\geq 0, \end{aligned}$$

and we are done. \square

Remark. In some sense, one can think of the question posed in [16] and answered here in Theorem 5.14 as a continuous variable analogue of the *absolute separability* problem in finite-dimensional quantum information, which asks for the characterisation of those spectra $\sigma = (\lambda_1, \dots, \lambda_{mn})$ such that every bipartite quantum state on $\mathbb{C}^m \otimes \mathbb{C}^n$ with spectrum σ is separable [253]. For a recent review of the state of the art, we refer the reader to [254]. A suggestive argument concerning this analogy goes as follows. An arbitrary unitary transformation $\rho \mapsto U\rho U^\dagger$ corresponds to an internal time evolution according to some unknown Hamiltonian. Then, the absolutely separable states are exactly those bipartite states whose correlations are so weak that they can not be made entangled by any internal evolution. In the case of continuous variable quantum systems, one may hold the free-field Hamiltonian $\mathcal{H} = \frac{1}{2}r^T r$ as the privileged one, so that it makes sense to restrict oneself to those unitary evolutions that preserve this particular Hamiltonian. If the original state is Gaussian and the unitaries are generated by quadratic Hamiltonians, so that they are associated with symplectic matrices, preserving the free-field Hamiltonian is the defining feature of passive transformations, and one obtains exactly the problem we solved here.

As is often the case, the technical details and the nature of the solution are simpler in the Gaussian realm. We found that the condition for being ‘absolutely separable’ in the Gaussian sense is expressed by a simple inequality involving only the two smallest ordinary eigenvalues of the QCM, and that there are no ‘absolutely PPT’ states that are not ‘absolutely separable’ too. This latter equivalence has been conjectured to hold for the original problem in discrete-variable systems as well, but so far only partial answers are available. Namely, the conditions for absolute PPT-ness can be written explicitly [255], but whether or not they imply absolute separability is in general unknown. However, the answer to this latter question has been shown to be affirmative for the case of two qubits [256] and more recently for qubit-qudit systems [257].

5.6 Summary and outlook

Throughout this chapter we presented some significant advances in the mathematical and physical study of separability and entanglement distillability in Gaussian states of continuous variable quantum systems. Based on the properties of Schur complements and other matrix analysis tools, we obtained a simplified necessary and sufficient condition for the separability of all multi-mode Gaussian states, requiring optimisation over the set of local covariance matrices of one subsystem only (Theorem 5.5). Exploiting this result, we presented a compact proof of the equivalence between PPT and separability for 1 vs n -mode Gaussian states (Theorem 5.6), a seminal result in continuous variable quantum information theory [14, 15], as well as extended the criterion to multimode classes of so-called mono-symmetric (Theorem 5.6) and isotropic Gaussian states (Theorem 5.9), through novel derivations. Furthermore, we completed the investigation of entanglement generation under passive opera-

tions by extending seminal results [16] to consider the generation of any, possibly PPT, Gaussian entangled state: in this context we showed that, if passive operations cannot turn an initial Gaussian state into a non-PPT one, then no PPT entanglement can be generated through them either (Theorem 5.14). This can be interpreted as establishing the equivalence between absolute separability and absolute PPT-ness in the Gaussian world. Side results of our analysis include a novel proof that Gaussian states invariant under partial transposition are separable (Corollary 5.11), as well as an independent proof of the equivalence between Gaussian separability and complete extendibility with Gaussian extensions [237].

In the context of this study, and with the methods illustrated above, it would be interesting to research more general combinations of symmetries and conditions on the symplectic spectra of quantum covariance matrices whereby the sufficiency of the PPT separability criterion might be further extended. For instance, is it possible to obtain a Gaussian analogue of the results in [248], whereby bound entangled Gaussian states can only exist given some simple condition on their symplectic spectrum? For example, both for mono-symmetric and isotropic states, large degeneracies in their symplectic spectra (for the marginal covariance matrix of one subsystem, and for the global covariance matrix of the bipartite system, respectively) were responsible for the sufficiency of the PPT condition for separability. It would be desirable to provide a full systematic characterisation of such requirements, possibly drawing inspiration from and/or shedding new insight on the Gaussian quantum marginal problem [258].

Finally, let us stress how matrix analysis tools such as those heavily hampered in this chapter have already been proven useful for qualitative and quantitative analysis of entanglement and other correlations in general states of continuous variable systems [218, 222, 223, 224, 225, 233, 259, 260, 261, 234, 235, 236].

Equipped with our renewed mathematical tools, we are now prepared to climb up to the stars.

Part III

Quantum correlations in continuous variable systems

Chapter 6

Schur complement inequalities and monogamy of quantum correlations

6.1 Introduction

The sun is setting on our investigation of quantum entanglement. However, this does not relieve us from the challenge of studying more general kinds of correlations that bipartite quantum states exhibit, which we are now prepared to take up. In fact, until now the expression ‘non-classical correlations’ in the title of the present thesis has been interpreted mainly in two distinct senses. The first of them is connected with the notion of entanglement, a concept whose importance in quantum information is hardly overestimated, and to the study of which we devoted most of Chapter 2. The second, stronger meaning has to do with violation of Bell inequalities, leading to the phenomenon dubbed ‘non-locality’ in Chapter 2. As it turns out, these are two extremes of a wider range of meanings one can give to the above words. For instance, an intermediate scenario can lead to the notion of *steering* [5, 259], which is equally important for some applications. If one were to summarise the content of this chapter in a single sentence, one could say that it aims to establish properties of and relations among various forms of non-classical correlations in bipartite quantum systems, and more specifically in continuous variable quantum systems whose state is Gaussian.

The present chapter is organised as follows. The rest of this section is devoted to introducing and motivating our investigation (Subsection 6.4.1) and to pointing the reader to our contributions (Subsection 6.1.2). In Section 6.2 we discuss the concept of steering and the associated resource theory, with particular regard to the Gaussian setting. The aim of Section 6.3 is to prove some matrix inequalities that involve quantum covariance matrices and to use them to derive properties of Gaussian steerability measures. Section 6.4 is then dedicated to the proof of a new hierarchical relation among the quantifiers of

Gaussian steerability, entanglement and total correlations based on the Rényi-2 entropies (Theorem 6.14). Several consequences are drawn from this fact. Finally, Section 6.5 contains the conclusions and a discussion of the outlook of this work.

6.1.1 Quantum correlations in Gaussian states

We spent most part of Chapter 2 studying mainly two types of non-classicality exhibited by composite systems in general probabilistic theories. The first and most fundamental phenomenon connected with non-classicality is the emergence of entanglement, i.e. the existence of bipartite states that can not be expressed as a convex combination of products of local states. This notion is intrinsically algebraic and as such does not bear a direct operational interpretation. However, one can turn it into something more concrete by using local measurement to generate from a state of a physical system – whatever its mathematical description is – concrete correlations among the measurement outcomes. When the correlations can not be explained by a classical way of thinking, i.e. by a ‘hidden variable model’, we say that they are Bell-nonlocal, or simply nonlocal. This second concept has nothing to do with the way we describe the system with a mathematical model, and on the contrary is fully intrinsic.

A useful way to think of nonlocality – especially in an information theoretical sense – is in terms of a task. Assume that two parties, Alice and Bob want to convince a third party, Charlie, that the state they share is entangled. Charlie does not know the details of the physical system under examination and can not access it, but he can nevertheless set up a test as follows. First, he forbids all communication between Alice and Bob. Then, he demands they provide him with a list of labels identifying possible measurements on their respective sides. In a multiple-round protocol, he picks labels from this list, asks them to perform the corresponding measurements (one for each side), and records the outcomes. For certain states (for this reason called nonlocal) and appropriately chosen measurements, Alice and Bob might be able to succeed and convince Charlie they do have shared entanglement, as otherwise he would not be able to explain the correlations he sees.

The above scenario features two *untrusted* parties, since Charlie believes both Alice and Bob could try to cheat him. There is however an alternative scenario where only one of the two parties (say Alice) is untrusted, and Bob’s good faith is instead acknowledged. In this case, we can identify Charlie with Bob, and the task becomes the following: first, Alice provides Bob with a list of labels identifying possible measurements on her side and with the states of Bob’s share of the system that each outcome of those measurements will yield; then, over multiple rounds, Bob picks labels at random, asks Alice to perform the corresponding measurements and to tell him the outcomes, and tests whether his state is what Alice claimed to be beforehand. When the state and the measurements are carefully chosen, Bob will have the feeling that Alice is able to ‘steer’ his system (which is isolated from the external environment) into

very different states, hence the name *steering* that this phenomenon bears [259]. Incidentally, let us mention that this peculiar feature of the theory was already identified by Schrödinger in person, who felt somewhat uncomfortable with it [5]. Anyway, if Bob is unable to find a classical explanation for this steering capability, he has to conclude that Alice possesses a system that is entangled with his. The global state is then said to be *steerable* from A to B with respect to the chosen measurements. For a formal definition, we refer the reader to Subsection 6.2.

In the case of continuous variable bipartite quantum systems, arguments similar to that in Subsection 5.1.1 suggest that a special case of the above scenario is of particular relevance, namely when the global state is Gaussian and the set of available measurements on Alice's side consists of all Gaussian measurements.¹ In this case, steerability can be characterised in closed form by means of a condition on the global covariance matrix analogous to that for separability, 5.44 [259]. Inspired by this characterisation, a measure of *Gaussian steerability* that quantifies the amount by which this condition is violated was introduced in [17]. We will devote some time discussing this measure and its properties, see the forthcoming Subsection 6.1.2.

The Gaussian steerability measure of [17] is only one of the many correlation measures one can construct by looking at the covariance matrix of the state. Other important examples are provided by the Gaussian entanglement of formation [263, 264] and by its Rényi-2 version [260]. The immediate advantage of using the Rényi-2 entropy instead of the more natural von Neumann entropy (3.30) lies in the fact that it respects the quadratic nature of Gaussian states. As we will see, this enables us to prove a lot of properties that these quantifiers obey. However, we do not want to give the reader the impression that mathematical convenience is the only reason to prefer Rényi-2 quantifiers in the Gaussian framework. In fact, such quantifiers can be more operationally relevant when one deals with *measured* quantities. The reason why this is the case will become clearer in the forthcoming Chapter 7.

6.1.2 Our contributions

This chapter is based on the (almost) homonymous paper [235]:

- L. Lami, C. Hirche, G. Adesso, and A. Winter. Schur complement inequalities for covariance matrices and monogamy of quantum correlations. *Phys. Rev. Lett.*, 117(22):220502, 2016.

Our first goal in this chapter will be to establish a collection of results for the Schur complement of a covariance matrix of a Gaussian state. As we said, covariance matrices bear a direct impact on the quantitative characterization

¹The reader could wonder what happens when one considers instead nonlocality in the Gaussian framework. It turns out that because of the representation (5.30) Gaussian measurements applied to bipartite Gaussian states can never lead to a violation of a Bell inequality, a fact that was already clear to Bell himself [262, XXI]. For further details on ways around this problem, see [210, IV].

of various forms of quantum correlations in continuous variable systems, and in turn on their usefulness for quantum technologies. Our analysis is inspired by recent works [260, 265, 266, 234], in which an inequality sharing the same formal structure as the strong subadditivity of entropy was obtained, by purely algebraic methods, for the log-determinant of positive semidefinite matrices $V_{ABC} \geq 0$:

$$\log \det V_{ABC} + \log \det V_C \leq \log \det V_{AC} + \log \det V_{BC}. \quad (6.1)$$

In the rest of the thesis, we will call (6.1) **strong subadditivity (SSA) of log-det entropy**. More reasons why the name is fitting will become apparent in Chapter 7. If one identifies V_{ABC} with the QCM of a $(n_A + n_B + n_C)$ -mode tripartite *quantum* system, which requires the extra condition corresponding to Heisenberg’s uncertainty principle (5.29), i.e. $V_{ABC} + i\Omega_{ABC} \geq 0$, then the scalar inequality (6.1) has relevant implications, yielding alternative quantifiers of correlations [260], monogamy constraints for Gaussian entanglement [260], and limitations for joint steering of single-mode states in a multipartite scenario [266, 234, 261].

In Section 6.3 we show, *inter alia*, that such an inequality admits a powerful *operator* strengthening directly at the level of covariance matrices (Theorems 6.1 and 6.5). This paves the way to several applications. First, we take the Gaussian steerability measure introduced in [17] and prove it is a monotone under the relevant sets of free operations in the Gaussian subtheory of the recently established resource theory of steering [267] (Theorem 6.10). Second, we tackle monogamy questions related to this quantifier. Since steerability is intrinsically asymmetric, there is more than one possible monogamy inequality to be investigated. Our findings show that the measure is monogamous with respect to the *steered* party, but not in general with respect to the *steering* party, unless some further assumptions are made (Theorem 6.13). The theory of Schur complements (Subsection 5.2.1) plays a crucial role in the proofs of all the main results of Section 6.3, some of which are highly nontrivial, and require the use of other sophisticated matrix analysis techniques.

Next, in Section 6.4 we study different forms of correlations in bipartite Gaussian states, especially when measured through quantifiers based on the Rényi-2 entropy. Our main result there is Theorem 6.14, which tells us that a strong inequality characterising ‘well-behaved’ theories of bipartite correlations [18] holds for Gaussian entanglement of formation and mutual information of Gaussian states when measured by means of the Rényi-2 entropy. The importance of this result lies also in the fact that an analogous inequality does *not* hold for the same quantifiers based on the standard von Neumann entropy (3.30). Moreover, a straightforward corollary of Theorem 6.14 yields the monogamy of the Rényi-2 Gaussian entanglement of formation for arbitrary many modes per party (Corollary 6.15), a fairly general statement which goes beyond all previous results. Let us add that the proof of Theorem 6.14 we present, surprisingly short and direct, relies heavily on the properties of the matrix geometric mean we already discussed in Subsection 5.2.2.

6.2 Quantum steering

The aim of this section is to provide a basic introduction to the quantum phenomenon known as steering. In Subsection 6.2.1 we discuss steering in general, as presented in [259], and examine the Gaussian special case more closely. Subsection 6.2.2 is then devoted to reviewing the basics of the subsequently developed resource theory of steering [267]. Finally, in Subsection 6.3.3 we introduce the Gaussian steerability measure of [17].

6.2.1 Generalities

Among the various surprising phenomena that occur in the quantum world, steering was among the first ones to be recognised. In a famous paper [5], Schrödinger himself introduced the concept and term, and noted with disappointment that ‘it is rather discomfoting that the theory should allow a system to be steered or piloted into one or the other type of state at the experimenter’s mercy in spite of his having no access to it.’ A more modern approach to the problem can be found in [259], where steering is seen from an information theoretical point of view as a *resource*. The question then becomes: *are there tasks than can be accomplished thanks to this capability of quantum systems of being steered by distant parties?*

As shown in [5], the answer turns out to be affirmative, and we already sketched the description of such a task in Subsection 6.4.1. Let us repeat it here in a slightly more general – and more dramatised – form. The scenario features two distant parties, Alice and Bob, connected with a one-way quantum channel that goes from Alice to Bob. Alice can send as many copies as he wants of a quantum state. Bob has a quantum laboratory where he can perform arbitrary quantum measurements with his powerful equipment, and he can test that this is actually the case. Alice maintains that the state she is sending is actually one of the two shares of a bipartite entangled state. How can she prove her claim to Bob, if he can not access her lab and does not trust her?

Alice maintains that her lab is equipped with some instruments with which she can implement some quantum measurements. In what follows, we will index such measurements with the label x . The possible values of x are publicly declared by Alice. The outcomes of those measurements, also made public by Alice, are indexed by a . The protocol is made by several identical rounds, all starting with Bob choosing a random label x and then asking Alice to output an (alleged) measurement outcome a . The crux of the game consists in Alice’s ability to foresee the state that Bob will have after she claims to have performed the measurement x and recorded the outcome a . Let us denote such state by $\rho_{a,x}$. We assume that the experiment can be repeated as many times as Bob wants, so that he can then test his quantum system using his powerful equipment and make sure that this is indeed the case. Another thing Bob will learn over multiple rounds is the probability of Alice outputting a when he asks for the measurement x .

It is useful to combine these two objects (states and probabilities) in a single operator $\rho(a|x) := p(a|x)\rho_{a,x}$. With this notation, the only thing Bob can access is the collection $\{\rho(a|x)\}_{a,x}$, called an **assemblage** [267]. In what follows, the two variables specifying the outcome of the measurement and the label identifying what measurement we are making will be always added as subscripts, in this order, when specifying the assemblage we consider. Any assemblage corresponding to a physically relevant situation must be *non-signalling*, i.e. it must satisfy the condition

$$\sum_a \rho(a|x) = \sum_a \rho(a|x') =: \rho_B \quad \forall x, x'. \quad (6.2)$$

Of course, very natural assemblages are those generated by a bipartite quantum state via measurements, i.e. such that

$$\rho(a|x) = \text{Tr}_A[E_a^x \otimes \mathbf{1} \rho_{AB}] \quad (6.3)$$

for some state ρ_{AB} and some collection of measurements $\{E_a^x\}_{a,x}$, whose elements will obey $E_a^x \geq 0$ for all a, x and $\sum_a E_a^x = \mathbf{1}_A$ for all x . As is easy to realise, such assemblages correspond to situations where Alice is really trying to distribute to Bob a share of a bipartite quantum state. However, it is known that not all non-signalling assemblages are compatible with the requirement of being generated by an original bipartite quantum state [268].

Let us come back to the main question here: given an assemblage $\{\rho(a|x)\}_{a,x}$, can Bob find an alternative classical explanation for what he is observing, or is any explanation going to involve entanglement? An explanation of the former kind will look like this: he will posit that Alice is trying to cheat him, as she is in fact sending not shares of entangled states but merely random quantum states that are generated according to a ‘hidden variable’ λ that she can access. In order to be consistent with the observations, this explanation must be such that

$$\rho(a|x) = \sum_\lambda p(\lambda) p(a|x\lambda) \xi_\lambda \quad \forall a, x, \quad (6.4)$$

where $p(\lambda)$ is some a priori probability, ξ_λ some collection of states, and $p(a|x\lambda)$ a classical decision rule to output a label a based on x and λ . Equation (6.4) defines the $A \rightarrow B$ **unsteerability** of the assemblage $\{\rho(a|x)\}_{a,x}$. Given a quantum state ρ_{AB} and a set of measurements \mathcal{M} on A , if the collection of measurements $\{E_a^x\}_{a,x}$, where $x \in \mathcal{M}$, is such that the associated assemblage (6.3) is $A \rightarrow B$ unsteerable, we say that ρ_{AB} itself is $A \rightarrow B$ unsteerable with measurements in \mathcal{M} .

A scenario that is of particular relevance to us is when the state $\rho_{AB} = \rho_{AB}^G(V_{AB}, w_{AB})$ is a Gaussian state of a bipartite continuous variable system, and the set of measurements \mathcal{M} comprises all Gaussian measurements on A . The question of what the steering properties of such state are in this setting boils down to a condition that involves only its covariance matrix V_{AB} , partitioned as in (5.31). In fact, it can be shown [259] that ρ_{AB} is $A \rightarrow B$ unsteerable

with Gaussian measurements iff

$$V_{AB}/V_A \geq i\Omega_B, \quad (6.5)$$

which can be equivalently written as

$$V_{AB} \geq 0_A \oplus i\Omega_B \quad (6.6)$$

by virtue of Corollary 5.2.

6.2.2 Resource theory of steering

If one wants to build a resource theory of steering, in which steerable states (or better, assemblages) are seen as resources that allow us to perform certain tasks, a prerequisite is the identification not only of free (unsteerable) states, but of free operations as well. A basic requirement free operations have to meet is that they can not create steerable states out of unsteerable ones, i.e. they have to preserve unsteerability. A thorough discussion of the free operations that are operationally relevant for applications can be found in [267]. We will limit ourselves to introducing some natural candidates for the role of free operations, and to verifying that they do not change unsteerability of assemblages.

Note. In what follows, we will indicate all probability distributions with the same letter p , which is very convenient for limiting the amount of symbols in a single equation. What distinguishes one from the other is the labels they carry, that identifies not only the particular realisation, but also the relevant random variable (together with the associated probability distribution) uniquely. For instance, we will write Bayes' rule like $p(a|b) = p(b|a) \frac{p(a)}{p(b)}$, since it is understood that b can not be another realisation of the random variable A that produces a , otherwise we would have denoted it by a' .

There are at least three elementary operations Alice and Bob can perform on their assemblages without changing its unsteerability. Let us discuss them one by one. Then, the free operations of the resource theory of steering will be defined as arbitrary concatenations of those elementary transformations.

- (1) Alice can measure Bob's share with a quantum instrument $\{\mathcal{E}_\omega\}_\omega$ before handing it over to Bob. The outcome ω is recorded by Alice, and now the game begins. What is the assemblage associated with the problem? Well, it is clearly $\left\{ \frac{\mathcal{E}_\omega(\rho(a|x))}{p(\omega)} \right\}_{a,x}$, where $p(\omega) = \text{Tr}[\mathcal{E}_\omega(\rho_B)]$ is the prior probability of getting the outcome ω . Let us stress that although this new assemblage does depend on ω , it does not feature ω as a *variable*. It turns out that by doing all this Alice can never make steerable an assemblage

that was previously unsteerable. In fact, if (6.4) holds then

$$\begin{aligned}\frac{\mathcal{E}_\omega(\rho(a|x))}{p(\omega)} &= \sum_\lambda p(\lambda)p(a|x\lambda)\frac{\mathcal{E}_\omega(\xi_\lambda)}{p(\omega)} \\ &= \sum_\lambda p'(\lambda)p(a|x\lambda)\xi'_\lambda,\end{aligned}$$

where

$$\begin{aligned}p'(\lambda) &:= \frac{p(\lambda)}{p(\omega)}\text{Tr}[\mathcal{E}_\omega(\xi_\lambda)] = \frac{p(\lambda)}{p(\omega)}p(\omega|\lambda) = p(\lambda|\omega), \\ \xi'_\lambda &:= \frac{\mathcal{E}_\omega(\xi_\lambda)}{\text{Tr}[\mathcal{E}_\omega(\xi_\lambda)]}.\end{aligned}$$

- (2) Alice can pre-process classical information, that is, she can: (i) ask Bob to start the protocol by outputting a new classical label y instead of x ; (ii) use the former to produce a label x according to some classical decision rule $p(x|y)$; and (iii) run the original protocol with this x , outputting a and forgetting x . Bob's final state when the input is y and the outcome is a becomes

$$\begin{aligned}\rho_{a,y} &= \sum_x p(x|a,y)\rho_{a,x} \\ &= \sum_x \frac{p(a,x,y)}{p(a,y)}\rho_{a,x} \\ &= \sum_x \frac{p(a,x|y)}{p(a|y)}\rho_{a,x} \\ &= \sum_x \frac{p(x|y)p(a|x,y)}{p(a|y)}\rho_{a,x}.\end{aligned}$$

Using the fact that $p(a|x,y) = p(a|x)$, valid since Bob's input y is fed into the original device $p(a|x)$ only through x , we obtain

$$\begin{aligned}\rho(a|y) &= p(a|y)\rho_{a,y} \\ &= \sum_x p(x|y)p(a|x)\rho_{a,x} \\ &= \sum_x p(x|y)\rho(a|x).\end{aligned}$$

From this we see that the original assemblage is unsteerable whenever the

old one was such, since

$$\begin{aligned}\rho(a|y) &= \sum_x p(x|y)\rho(a|x) \\ &= \sum_{x,\lambda} p(x|y)p(\lambda)p(a|x,\lambda)\xi_\lambda \\ &= \sum_\lambda p(\lambda)p(a|y,\lambda)\xi_\lambda,\end{aligned}$$

where $p(a|y,\lambda) = \sum_x p(x|y)p(a|x,\lambda)$.

- (3) Alice can post-process all her classical information before outputting the final label. Namely, instead of a she can choose to output a new label b , which is produced with probability $p(b|ax)$, and later delete a . According to Bayes' rule, Bob's state corresponding to an outcome b will be

$$\begin{aligned}\rho_{b,x} &= \sum_a p(a,x|b)\rho_{a,x} \\ &= \sum_a \frac{p(a,b,x)}{p(b,x)}\rho_{a,x} \\ &= \sum_a \frac{p(a,b|x)}{p(b|x)}\rho_{a,x} \\ &= \sum_a \frac{p(a|x)p(b|a,x)}{p(b|x)}\rho_{a,x},\end{aligned}$$

i.e. $\rho(b|x) = \sum_a p(b|a,x)\rho(a|x)$. If the original assemblage was unsteerable, then we have

$$\begin{aligned}\rho(b|x) &= \sum_a p(b|a,x)\rho(a|x) \\ &= \sum_{a,\lambda} p(b|a,x)p(\lambda)p(a|x,\lambda)\xi_\lambda \\ &= \sum_\lambda p(b|x,\lambda)\xi_\lambda,\end{aligned}$$

where $p(b|x,\lambda) = \sum_a p(a|x,\lambda)p(b|a,x)$, and hence the new assemblage $\{\rho(b|x)\}_{b,x}$ is unsteerable, too.

Remark. Although we did not say it explicitly, all of the above operations preserve the fundamental requirement of being non-signalling.

The most general operation on assemblages that can be implemented using as building blocks the above three elementary steps is called in [267] a **deterministic one-way LOCC**, or **1W-LOCC** for short. Of course, as we defined them these operations are specific to assemblages, while the set $\text{LOCC}_{\rightarrow}$ considered in Chapter 3 comprised protocols acting on quantum states.

According to [267], any valid *quantifier* of $A \rightarrow B$ steerability should be mandatorily: (i) vanishing on unsteerable assemblages and (ii) nonincreasing on average under 1W-LOCC operations, and optionally (iii) convex. A quantity satisfying (i), (ii), and (iii) is referred to as a **convex steering monotone**; examples of such monotones are discussed in [267].

What happens if we apply a 1W-LOCC to a quantum assemblage such as (6.3)? Do we get another quantum assemblage? The answer to this question is readily seen to be affirmative: the quantum state that generates the new assemblage is of the form $\frac{(I \otimes \mathcal{E})(\rho_{AB})}{\text{Tr} \mathcal{E}(\rho_B)}$, where \mathcal{E} is a trace non-increasing completely positive map (i.e. an element of a quantum instrument), which corresponds to a specific outcome as selected by Alice. This is a useful observation to keep in mind.

Finally, let us consider the resource theory of steering in the Gaussian scenario. When the underlying states are Gaussian, it is very natural to restrict ourself to Gaussian free operations. In the case of 1W-LOCC, this means the following. First, the quantum instrument considered in point (1) of the above list has to be obtained by composing: (i) additions of Gaussian ancillas; (ii) symplectic unitaries; and (iii) Gaussian partial measurements (Subsection 5.2.3). In particular, each of the maps \mathcal{E}_ω is a Gaussian completely positive map whose action is expressed by (5.40). Second, the classical channels in points (2) and (3) of the above discussion have to be classical Gaussian channels. Assemblages deriving from Gaussian states together with **Gaussian 1W-LOCC operations** as we described them define a **resource theory of Gaussian steering** that is a or sub-theory of the standard resource theory of steering we discussed in this subsection.

6.2.3 Gaussian steerability measure

We now turn our attention to the problem of *quantifying* steerability in a meaningful way. Mathematically, this amounts to identifying real measures on assemblages that are *monotone* (say, monotonically non-increasing) under all free operations. We are mostly interested in the case of steerability of Gaussian states by Gaussian measurements², so quantifiers can be equivalently thought of as being defined on the set of bipartite QCMs V_{AB} , which we assume partitioned with respect to the $A : B$ splitting as in (5.31).

In [17], a Gaussian steerability measure was introduced by leveraging the condition (6.5). In order to discuss it, we need to introduce some definitions first. In what follows, denote by $\nu_i(A)$ the i -th smallest symplectic eigenvalue of a strictly positive $2n \times 2n$ matrix $A > 0$. Construct the two functions

$$g_{\pm}(A) = \sum_{i=1}^n \max \{ \pm \log \nu_i(A), 0 \} . \quad (6.7)$$

²Some of the results presented in this chapter hold also when either the state or the measurements are Gaussian, but we focus here on the simplest case of all, i.e. when both the state and the measurement are Gaussian.

Clearly, by the Williamson decomposition theorem (5.43), $g_-(A) = 0$ is equivalent to A satisfying (5.29), i.e. being a legitimate QCM. Let us mention in passing that the function g_- finds many applications in continuous variable quantum information. For instance, the logarithmic negativity [269, 270] of a bipartite state ρ_{AB} , defined as $E_N(\rho_{AB}) = \log \|\rho_{AB}^{T_B}\|_1$ (where T_B denotes partial transposition), takes the form $E_N(\rho_{AB}) = g_-(V_{AB}^{T_B}) = g_-(\Theta V_{AB} \Theta)$ if $\rho_{AB} = \rho_{AB}^G(V_{AB}, w_{AB})$ is a Gaussian state with QCM V_{AB} ; here, the partial transpose of the QCM is expressed according to (5.48). More details on the properties of the functions in (6.7) will be provided in Subsection 6.3.2.

Now, let us come back to our measure of Gaussian steerability, which on a Gaussian state ρ_{AB}^G with QCM V_{AB} takes the form [17]

$$\mathcal{G}(A)B)_V = g_-(V_{AB}/V_A), \quad (6.8)$$

in the case of party A steering party B . As we mentioned, we will consider the function $\mathcal{G}(A)B)_V$, from now on dubbed **Gaussian steerability**, as defined on bipartite covariance matrices instead of states, but this is no loss of generality. Observe that by virtue of the properties of g_- discussed above, (6.8) measures the amount by which the unsteerability condition (6.5) is violated, and is therefore a good candidate for a valid quantifier.

A point to stress is that the definition (6.8) can be extended to every state that possesses a covariance matrix, even if it is not Gaussian. While $\mathcal{G}(A)B)_V > 0$ is necessary and sufficient for steerability from A to B of a Gaussian state with QCM V_{AB} by means of Gaussian measurements on A [17, 259], it can be shown that it is only sufficient if either the state [271] or the measurements [272, 273] are non-Gaussian. This extends significantly the range of applicability of many of our results.

6.3 Schur complement inequalities and Gaussian steerability

This aim of this section is to show that the Gaussian steerability of (6.8) is a valid monotone of the Gaussian resource theory of steering. In order to achieve our goal, in Subsection 6.3.1 we first derive some fundamental inequality that Schur complements, and in particular the Schur complements of quantum covariance matrices, must obey. After a brief technical detour to study some mathematical properties of Subsection 6.3.2, we apply these inequalities to prove the main results of the section, Theorems 6.10 and 6.13.

6.3.1 Schur complement inequalities

For an introduction to Schur complements, we refer the reader to the dedicated monograph [232]. Here, we follow the notation and conventions of Subsection 5.2.1, where a succinct list of properties of Schur complements is also provided.

Our starting point is the inequality (6.1), from which the authors of [234] deduce that

$$\log \det V_{AC} + \log \det V_{BC} - \log \det V_A - \log \det V_B \geq 0 \quad (6.9)$$

for all QCMs V_{ABC} .³ This is the form of (6.1) that is particularly relevant in Gaussian quantum information.

For the sake of the presentation, let us provide a one-line proof of (6.9). We already observed that Corollary 5.2 implies that the function $H \mapsto H/A$ is concave in $H > 0$. Applying the logarithm (which is a matrix monotone and concave) and taking the trace one obtains the concavity of $\text{Tr} \log(H/A) = \log \det(H/A)$ in $H > 0$. Then, rewrite the left-hand side of (6.9) as

$$\log \det(V_{AC}/V_A) + \log \det(V_{BC}/V_B).$$

Since this is a concave function of V_{ABC} , one can evaluate it on pure QCMs, which are easily seen to saturate it. As shown in [234], inequalities (6.1) and (6.9) are equivalent expressions of strong subadditivity for the log-determinant, and can be converted into each other by ‘purifying’ V_{ABC} .

This suggests that the Schur complement of QCMs can define a natural notion of quantum conditional covariance, as previously noted for classical Gaussian variables [274]. Hence we will study the Schur complement V_{AB}/V_B , thereby proving that many well-known properties of the standard conditional entropy $H(A|B) = H(AB) - H(B)$, where H denotes respectively Shannon or von Neumann entropy for a classical or quantum system, have a straightforward equivalent within this framework.

We start by recalling that a canonical formulation of strong subadditivity in classical and quantum information theory is $H(A|BC) \leq H(A|C)$, i.e. partial trace on the conditioning system increases the conditional entropy [159, 160, 109]. Guided by our formal analogy, our first result is thus a generalization of (6.1).

Theorem 6.1 (Partial trace in the denominator increases Schur complement). *If $V_{ABC} \geq 0$ is any tripartite positive semidefinite matrix, then*

$$V_{ABC}/V_{BC} \leq V_{AC}/V_C. \quad (6.10)$$

Proof. Since $V_{ABC} \geq W_A \oplus 0_{BC}$ implies $V_{AC} \geq W_A \oplus 0_C$, employing Corollary 5.2 we find

$$\begin{aligned} V_{ABC}/V_{BC} &= \max \{W_A : V_{ABC} \geq W_A \oplus 0_{BC}\} \\ &\leq \max \{W_A : V_{AC} \geq W_A \oplus 0_C\} \\ &= V_{AC}/V_C, \end{aligned}$$

which proves the claim. □

³We stress that while (6.1) holds for all strictly positive matrices, (6.9) requires V_{ABC} to be a QCM, i.e. to satisfy (5.29).

Clearly, taking the determinant of (6.10) and applying the factorization property (5.4) of the Schur complement under determinants, equation (5.4), yields (6.1) immediately. Notice further that the invariance of V_{AB}/V_B under symplectic operations on B , implied by the congruence covariance property of the Schur complement, equation (5.7), and its monotonicity under partial trace, suffice to guarantee its monotonicity under general Gaussian channels \mathcal{N}_B on B .

But there is more: perhaps surprisingly, the Schur complement is also monotonically increasing under general non-deterministic *classical* Gaussian operations on B . These maps act at the level of covariance matrices as in (5.40), but the ‘Choi covariance matrix’ of the map, i.e. $\gamma = \begin{pmatrix} \gamma_1 & \delta_{12} \\ \delta_{12}^T & \gamma_2 \end{pmatrix}$, is only required to be a strictly positive definite matrix, instead of a legitimate QCM, which would yield a completely positive Gaussian map. We formulate this statement as follows.

Theorem 6.2 (Classical Gaussian maps in the denominator increase Schur complement). *If $\Gamma_{B \rightarrow B'}$ is a non-deterministic classical Gaussian map acting as in (5.40), with the corresponding ‘Choi covariance matrix’ $\gamma_{BB'}$ only required to be strictly positive definite, then*

$$\Gamma_{B \rightarrow B'}(V_{AB})/\Gamma_{B \rightarrow B'}(V_B) \geq V_{AB}/V_B.$$

Proof. We employ the formula (5.42), which in our case reads $\Gamma_{B \rightarrow B'} : V_B \mapsto (\gamma_{BB'} + V_B)/(\gamma_B + V_B)$. Then,

$$\begin{aligned} & \Gamma_{B \rightarrow B'}(V_{AB})/\Gamma_{B \rightarrow B'}(V_B) \\ &= ((\gamma_{BB'} + V_{AB})/(\gamma_B + V_B)) / ((\gamma_{BB'} + V_B)/(\gamma_B + V_B)) \\ &\stackrel{(1)}{=} (\gamma_{BB'} + V_{AB})/(\gamma_{BB'} + V_B) \\ &\stackrel{(2)}{\geq} V_{AB}/V_B. \end{aligned}$$

The justification of these steps is as follows: (1) is an instance of the quotient property (5.8); (2) follows from the elementary bound

$$\begin{pmatrix} A & X \\ X^T & B + \sigma \end{pmatrix} / (B + \sigma) \geq \begin{pmatrix} A & X \\ X^T & B \end{pmatrix} / B,$$

in turn an immediate consequence of the monotonicity of the matrix inversion on strictly positive definite matrices. \square

Next, we would like to obtain from (6.10) an operator generalization of (6.9) by applying the symplectic purification trick. This requires a certain amount of work, so we state a couple of preliminary lemmas. From now on, we will always assume that the V matrices are not only positive semidefinite but actual QCMs, which obey (5.29).

Lemma 6.3. *Let V_{AB} be a bipartite pure QCM. Then*

$$V_{AB}/V_A = \Omega_B^T V_B^{-1} \Omega_B. \quad (6.11)$$

Proof. Since V_{AB} is symplectic, it satisfies $V_{AB}^{-1} = \Omega_A^T V_{AB}^T \Omega_{AB} = \Omega_{AB}^T V_{AB} \Omega_{AB}$. Comparing the B blocks thanks to (5.6) yields (6.11). \square

Lemma 6.4. *Let V_{ABC} be a tripartite pure QCM. Then*

$$V_{AB}/V_B = \Omega_A^T (V_{AC}/V_C)^{-1} \Omega_A. \quad (6.12)$$

Proof. Applying Lemma 6.3 to the bipartite pure QCM $V_{AC:B}$ (i.e. the same matrix V_{ABC} but seen as representing a pure QCM pertaining to the bipartite system $AC : B$), one obtains

$$V_{ABC}/V_B = \Omega_{AC}^T V_{AC}^{-1} \Omega_{AC}.$$

Projecting onto the A component thanks to the identities

$$\begin{aligned} V_{AB}/V_B &= \Pi_A (V_{ABC}/V_B) \Pi_A^T, \\ \Pi_A \Omega_{AC}^T V_{AC}^{-1} \Omega_{AC} \Pi_A^T &= \Omega_A^T (V_{AC}/V_C)^{-1} \Omega_A, \end{aligned}$$

the latter following from the block inversion formula (5.6), one deduces (6.12). \square

We then get the following for any tripartite quantum covariance matrix.

Theorem 6.5 (Schur complement of QCMs is monogamous). *If $V_{ABC} \geq i\Omega_{ABC}$ is any tripartite QCM, then*

$$V_{AC}/V_A \geq \Omega_C^T (V_{BC}/V_B)^{-1} \Omega_C. \quad (6.13)$$

Proof. Consider a symplectic purification V_{ABCD} of the system ABC . Applying first (6.10) and then (6.12) yields (6.13): $V_{AC}/V_A \geq V_{ACD}/V_{AD} = \Omega_C^T (V_{BC}/V_B)^{-1} \Omega_C$. Alternatively, observe that the difference between right- and left-hand side of (6.13) is concave in V_{ABC} (as V_{AC}/V_A is concave and $(V_{BC}/V_B)^{-1}$ is convex), and it vanishes on pure QCMs by Lemma 6.4. \square

The following special case of the above theorem will be useful later.

Corollary 6.6. *If $V_{AB} \geq i\Omega_{AB}$ is any bipartite QCM, then*

$$V_{AB}/V_B \geq \Omega_A^T V_A^{-1} \Omega_A. \quad (6.14)$$

Proof. When the B subsystem is uncorrelated with the rest, (6.13) becomes $V_{AC}/V_A \geq \Omega_C^T V_C^{-1} \Omega_C$, which yields the claim after relabelling. \square

We remark that the operator inequalities (6.10) and (6.13) are significantly stronger than the scalar ones (6.1) and (6.9) reported in [260, 265, 266, 234], as the former establish algebraic limitations directly at the level of covariance matrices, in a similar spirit to the marginal problem [258], for arbitrary multipartite states. These powerful tools will be instrumental in the investigation of quantum correlations, especially steering and entanglement, conducted in the forthcoming Subsection 6.3.3.

6.3.2 Mathematical detour: properties of the functions g_{\pm}

Before we delve into the study of the properties of the Gaussian steerability (6.8), we need to gain some knowledge of the basic properties of the functions g_{\pm} defined in (6.7). A decisive ingredient of our analysis is a version of the Courant-Fischer-Weyl variational principle for symplectic eigenvalues proven recently in [252].

First of all, observe that

$$g_{\pm}(A) = g_{\pm}(SAS^T) \quad \forall \text{ symplectic } S, \tag{6.15}$$

since those functions are defined only in terms of symplectic eigenvalues. Moreover, for all $A, B > 0$ it holds that

$$g_{\pm}(A^{-1}) = g_{\mp}(A), \tag{6.16}$$

$$g_+(A) - g_-(A) = \frac{1}{2} \log \det A, \tag{6.17}$$

$$g_{\pm}(A \oplus B) = g_{\pm}(A) + g_{\pm}(B). \tag{6.18}$$

The next lemma will prove helpful many times.

Lemma 6.7. *The function g_- is monotonically non-increasing and convex, while g_+ is monotonically non-decreasing but neither convex nor concave.*

Proof. The fact that g_+, g_- are monotone in their inputs can be seen as an easy consequence of the symplectic equivalent of Weyl’s monotonicity theorem first proven in [222, Lemma 2] (see also [275, Theorem 8.15]). This states that if $A, B > 0$ are $2n \times 2n$ real matrices that satisfy $A \geq B$, then their ordered symplectic eigenvalues are such that $\nu_i(A) \geq \nu_i(B)$ for all $i = 1, \dots, n$. The claim follows by performing elementary algebraic manipulations.

Now, let us prove the convexity of $g_-(A)$. Our proof employs the recently found variational expression

$$\prod_{i=1}^k \nu_i(A) = \min \left\{ \sqrt{\det(S^T AS)} : S^T \Omega_{2n} S = \Omega_{2k} \right\} \tag{6.19}$$

for the product of the k smallest symplectic eigenvalues [252, Theorem 5]. In the above expression, Ω_{2k} denotes the standard symplectic form on k modes. We find

$$g_-(A) = \max_{1 \leq k \leq n} \sum_{i=1}^k (-\log \nu_i(A)) \tag{6.20}$$

$$= -\frac{1}{2} \min \left\{ \log \det(S^T AS) : S^T \Omega_{2n} S = \Omega_{2k}, 1 \leq k \leq n \right\}. \tag{6.21}$$

Since $\log \det$ is well-known to be concave on positive matrices [276], and $F(x) := \min_{y \in Y} f(x, y)$ is always concave in x if $f(x, y)$ was concave in x for all fixed $y \in Y$, we infer that g_- is indeed convex. Finally, in order to see that g_+ is neither convex nor concave it suffices to test it on positive multiples of the identity. \square

Remark. Why Lemma 6.7 does not imply that the logarithmic negativity is convex. The formula $E_N = g_-(\Theta V_{AB} \Theta)$ for the logarithmic negativity, discussed in Subsection 6.3.3, together with the convexity of g_- , could lead us to think that the logarithmic negativity is convex in the input state, which is known to be false [269, 270]. The reason why this reasoning does not work is that E_N is expressible in terms of the covariance matrix only for Gaussian states, that do not constitute a convex set. However, it is true that if $\{\rho_i\}_i$ is a family of Gaussian states such that their convex combination $\sum_i p_i \rho_i$ is again Gaussian, then

$$E_N \left(\sum_i p_i \rho_i \right) \leq \sum_i p_i E_N(\rho_i). \quad (6.22)$$

One could dub this property **Gaussian-convexity**. The logarithmic negativity is thus an example of a Gaussian-convex function which is in general non-convex.

Lemma 6.7 can be used to prove that $g_-(V_{AB})$ decreases if the coherences between subsystems A and B are erased.

Lemma 6.8. Let $V_{AB} > 0$ be a bipartite positive definite matrix (not necessarily a QCM). Then

$$g_-(V_{AB}) \geq g_-(V_A) + g_-(V_B). \quad (6.23)$$

Proof. We will give a straightforward proof based on the convexity of g_- , but an alternative argument can be deduced directly from the variational expression (6.21). Observing that $\mathbf{1}_A \oplus (-\mathbf{1}_B)$ is a symplectic operation, and using the symplectic invariance of g_- (6.15), one finds

$$g_-(V_{AB}) = g_- \left((\mathbf{1}_A \oplus (-\mathbf{1}_B)) V_{AB} (\mathbf{1}_A \oplus (-\mathbf{1}_B)) \right), \quad (6.24)$$

from which we infer

$$\begin{aligned} g_-(V_{AB}) &= \frac{1}{2} \left(g_-(V_{AB}) + g_- \left((\mathbf{1}_A \oplus (-\mathbf{1}_B)) V_{AB} (\mathbf{1}_A \oplus (-\mathbf{1}_B)) \right) \right) \\ &\geq g_- \left(\frac{1}{2} V_{AB} + \frac{1}{2} (\mathbf{1}_A \oplus (-\mathbf{1}_B)) V_{AB} (\mathbf{1}_A \oplus (-\mathbf{1}_B)) \right) \\ &= g_-(V_A \oplus V_B) \\ &= g_-(V_A) + g_-(V_B). \end{aligned}$$

thanks to Lemma 6.7. □

Remark. One could be tempted to conjecture inequalities linking $g_-(V_{AB})$ with $g_-(V_A)$ and $g_-(V_{AB}/V_A)$. However, in general on the one hand $g_-(V_{AB}) \not\geq g_-(V_A) + g_-(V_{AB}/V_A)$ (counterexample: bipartite QCM that is steerable from A to B) and on the other hand $g_-(V_{AB}) \not\leq g_-(V_A) + g_-(V_{AB}/V_A)$ (numerical counterexamples are easily found).

We have seen that the function g_- admits a variational expression (6.21) in terms of a maximum (or the negative of a minimum). Now, we explore an alternative variational principle for g_- that yields it directly as a minimum instead. The following lemma completes our toolbox.

Lemma 6.9. *The functions g_{\pm} admit the following representations:*

$$g_+(W) = \min_{W \leq Z \geq i\Omega} \frac{1}{2} \log \det Z, \quad (6.25)$$

$$g_-(W) = \min_{\Omega^T W^{-1} \Omega \leq Z \geq i\Omega} \frac{1}{2} \log \det Z. \quad (6.26)$$

Proof. Since (6.16) holds, we can limit ourselves to showing the first. The monotonicity of g_+ implies that

$$g_+(W) \leq g_+(Z) = \frac{1}{2} \log \det Z,$$

whenever $W \leq Z \geq i\Omega$, where the last equality holds because Z is a QCM. This shows that $g_+(W) \leq \min_{W \leq Z \geq i\Omega} \frac{1}{2} \log \det Z$. On the other hand, we can perform a Williamson decomposition (5.43) on W and write it as $W = S \begin{pmatrix} \nu & 0 \\ 0 & \nu \end{pmatrix} S^T$, with S symplectic and $\nu \geq \mathbb{1}$ diagonal. This allows us to construct the ansatz $\tilde{Z} := S \begin{pmatrix} \tilde{\nu} & 0 \\ 0 & \tilde{\nu} \end{pmatrix} S^T$, where $\tilde{\nu}_i := \max\{\nu_i, 1\}$, which satisfies $W \leq \tilde{Z} \geq i\Omega$ and

$$\frac{1}{2} \log \det \tilde{Z} = \sum_i \max\{0, \log \nu_i(W)\} = g_+(W), \quad (6.27)$$

hence we are done. \square

Remark. We remind the reader that for any $W > 0$ the condition $Z \geq \Omega^T W^{-1} \Omega$ is equivalent to

$$\begin{pmatrix} W & \Omega \\ \Omega^T & Z \end{pmatrix} \geq 0, \quad (6.28)$$

by virtue of Lemma 5.1.

6.3.3 Monotonicity of Gaussian steerability

We are now in position to employ the mathematical results obtained in Subsections 6.3.1 and 6.3.2 to our problem, that is, showing that the Gaussian steerability in (6.8) is a valid monotone within the resource theory of Gaussian steering. The requirements that such a monotone should satisfy, as put forward in [267], are reported in the discussion at the end of Subsection 6.2.2. Let us translate them into the Gaussian framework.

- (i) In order for $\mathcal{G}(A)B$ to vanish on Gaussian unsteerable states, it must be that $\mathcal{G}(A)B_V = 0$ whenever $V_{AB}/V_A \geq i\Omega_B$ (6.5). This is automatically guaranteed by construction, thanks to (6.7).

- (ii) The monotonicity under 1W-LOCC is more difficult. As we said at the end of Subsection 6.2.2, the measure needs to be monotonically non-increasing under the transformation

$$\rho_{AB} \mapsto \frac{(I \otimes \mathcal{E})(\rho_{AB})}{\text{Tr}[\mathcal{E}(\rho_B)]},$$

where \mathcal{E} is an arbitrary completely positive map on the B system. Since here we care only about the Gaussian resource theory, we can assume \mathcal{E} to be a Gaussian CP map, whose action is given by (5.40).

- (iii) We move on to the third, optional requirement, namely the convexity, or in our case Gaussian-convexity (see the Remark after Lemma 6.7). A good starting point is the observation that the QCM \bar{V} of a convex combination $\sum_i p_i \rho_i$ of states ρ_i with covariance matrices V_i satisfies [15]

$$\bar{V}_i \geq \sum_i p_i V_i.$$

In particular, convexity in the quantum state is automatically guaranteed by: (iii.a) anti-monotonicity in the QCM, and (iii.b) convexity in the QCM.

The following result shows that the Gaussian steerability measures meets all requirements (i)-(iii), hence fully validating it within the Gaussian resource theory of steering. This settled a question left open in [17, 267]. As a by-product, we also establish further properties such as the monotonicity under non-deterministic Gaussian maps on the *steering* party (Theorem 6.10(c)). Our findings extend those of [17] considerably, for in [17] (some of) the stated properties were only proven in the special case of one-mode steered subsystem, i.e. $n_B = 1$.

Theorem 6.10 (Properties of Gaussian steerability). *The Gaussian steerability measure, defined on Gaussian states by (6.8), satisfies the following properties:*

- (a) $\mathcal{G}(A)B)_V$ is convex and non-increasing in the QCM V_{AB} , hence Gaussian-convex in the state ρ_{AB}^G , which meets condition (iii) above;
- (b) $\mathcal{G}(A)B)$ is additive under tensor products, i.e. under direct sums of QCMs:
- $$\mathcal{G}(A_1 A_2)B_1 B_2)_{V_{A_1 B_1} \oplus W_{A_2 B_2}} = \mathcal{G}(A_1)B_1)_{V_{A_1 B_1}} + \mathcal{G}(A_2)B_2)_{W_{A_2 B_2}}; \quad (6.29)$$
- (c) $\mathcal{G}(A)B)$ is non-increasing under general, non-deterministic Gaussian maps on the steering party A ;
- (d) $\mathcal{G}(A)B)$ is non-increasing under general, non-deterministic Gaussian maps on the steered party B , meeting requirement (ii) above;
- (e) for any QCM V_{ABC} , it holds $\mathcal{G}(A)C)_V \leq g_+(V_{BC}/V_B)$.

Before delving into the proof of the above theorem, let us comment a bit on the various claims. As mentioned before, (a) and (d) serve the main purpose of the section, i.e. showing that the Gaussian steerability measure (6.8) is a valid monotone of the Gaussian resource theory of steering. Identity (6.29) in (b) is desirable – even if not mandatory – for a well-behaved measure of quantum correlations. As a matter of fact, most standard measures of entanglement do not satisfy it (a notable exception is the squashed entanglement (3.28), whose additivity was shown in [11]).

As for claim (c), one could expect monotonicity under *Gaussian channels* on the steering party, because of the following reasoning. Taking a quantum assemblage as in (6.3) and applying before the measurement a quantum channel \mathcal{N} on A one gets

$$\mathrm{Tr}_A [E_a^x \otimes \mathbb{1} (\mathcal{N} \otimes I)(\rho_{AB})] = \mathrm{Tr}_A [\mathcal{N}^\dagger(E_a^x) \otimes \mathbb{1} \rho_{AB}],$$

where the adjoint map \mathcal{N}^\dagger is defined by $\mathrm{Tr}[X\mathcal{N}(Y)] = \mathrm{Tr}[\mathcal{N}^\dagger(X)Y]$, exactly as in (B.1). The operators $\mathcal{N}^\dagger(E_a^x)$ form a valid measurement because \mathcal{N} is trace-preserving. If $\{E_a^x\}_a$ covers over all Gaussian measurements when x runs over its range, and \mathcal{N} is a Gaussian channel, then $\{\mathcal{N}^\dagger(E_a^x)\}_a$ constitutes a subset of all Gaussian measurements. Naturally, we do not want the steerability to increase when the set of measurements that are available to Alice is *smaller*, which leads to (c).

Claim (e) establishes an upper bound that can be useful in practical situations. An example of such scenario would feature a malicious agent A who is trying to steer system C . Thanks to inequality (e), B and C can estimate the maximum steerability A can achieve by measuring properties of their covariance matrix V_{BC} alone.

Proof of Theorem 6.10. We prove the claims one by one.

(a) $\mathcal{G}(A)B)_V$ is convex and non-increasing as a function of the QCM $V_{AB} > 0$.

Both properties follow straightforwardly by combining concavity and monotonicity of the Schur complement (deduced from Corollary 5.2) with the convexity of g_- (Lemma 6.7). Let us prove convexity for instance. Since the Schur complement is concave, for any $V_{AB}, W_{AB} > 0$ and $0 \leq p \leq 1$ we obtain

$$(pV_{AB} + (1-p)W_{AB}) / (pV_A + (1-p)W_A) \geq pV_{AB}/V_A + (1-p)W_{AB}/W_A.$$

Applying the fact that g_- is monotonically non-increasing and convex gives

$$\begin{aligned} \mathcal{G}(A)B)_{pV_{AB}+(1-p)W_{AB}} &= g_-((pV_{AB} + (1-p)W_{AB}) / (pV_A + (1-p)W_A)) \\ &\leq g_-(pV_{AB}/V_A + (1-p)W_{AB}/W_A) \\ &\leq pg_-(V_{AB}/V_A) + (1-p)g_-(W_{AB}/W_A) \\ &= p\mathcal{G}(A)B)_V + (1-p)\mathcal{G}(A)B)_W. \end{aligned}$$

(b) $\mathcal{G}(A)B$ is additive under tensor products.

Elementary, since

$$\begin{aligned} \mathcal{G}(A_1 A_2)B_1 B_2)_{V_{A_1 B_1} \oplus W_{A_2 B_2}} &= g_-((V_{A_1 B_1} \oplus W_{A_2 B_2})/(V_{A_1} \oplus W_{A_2})) \\ &= g_-(V_{A_1 B_1}/V_{A_1} \oplus W_{A_2 B_2}/W_{A_2}) \\ &= g_-(V_{A_1 B_1}/V_{A_1}) + g_-(W_{A_2 B_2}/W_{A_2}) \\ &= \mathcal{G}(A_1)B_1)_{V_{A_1 B_1}} + \mathcal{G}(A_2)B_2)_{W_{A_2 B_2}}, \end{aligned}$$

where the third equality follows from (6.18).

(c) $\mathcal{G}(A)B$ is monotonically non-increasing under general, non-deterministic Gaussian maps on the steering party A .

Using the monotonicity of the Schur complement under general Gaussian maps, as expressed by Theorem 6.2, one gets

$$\Gamma_{A \rightarrow A'}(V_{AB})/\Gamma_{A \rightarrow A'}(V_A) \geq V_{AB}/V_A.$$

Applying g_- to both sides and reversing the inequality as prescribed by Lemma 6.7, we obtain

$$\mathcal{G}(A')B)_{\Gamma_{A \rightarrow A'}(V_{AB})} \leq \mathcal{G}(A)B)_{V_{AB}}, \quad (6.30)$$

which is the claim.

(d) $\mathcal{G}(A)B$ is monotonically non-increasing under general, non-deterministic Gaussian maps on the steered party B .

This is the most difficult claim to prove. First of all, we recall that any general, non-deterministic Gaussian map can always be obtained by: (i) adding an uncorrelated ancillary system in a Gaussian state; (ii) performing a global symplectic operation; and (iii) measuring some of the modes by means of a Gaussian measurement. Clearly, $\mathcal{G}(A)B$ is invariant under the addition of an ancillary steered system in an uncorrelated state because of the additivity proven in point (b). Furthermore, the invariance under symplectic operations on B is guaranteed by (6.15). Thus, we are only left to prove that the Gaussian steerability decreases when a partial Gaussian measurement is performed on the steered system.

According to (5.39), given a composite system ABC in a Gaussian state with QCM V_{ABC} , when one makes a Gaussian measurement on the subsystem C that has seed γ_C , the reduced post-measurement state of subsystem AB is Gaussian and with a QCM given by $\tilde{V}_{AB} = (V_{ABC} + \gamma_C)/(V_C + \gamma_C)$ (independently of the outcome), where we adopted the shorthand notation $\gamma_C = 0_{AB} \oplus \gamma_C$. Bearing that in mind, we are claiming that for all QCMs V_{ABC} one has

$$g_-(\tilde{V}_{AB}/\tilde{V}_A) \leq g_-(V_{ABC}/V_A). \quad (6.31)$$

Call $W_{BC} := V_{ABC}/V_A$. Then, a simple calculation that uses the quotient property of the Schur complement (5.8) shows that

$$\begin{aligned} \tilde{V}_{AB}/\tilde{V}_A &= ((V_{ABC} + \gamma_C)/(V_C + \gamma_C)) / ((V_{AC} + \gamma_C)/(V_C + \gamma_C)) \\ &= (V_{ABC} + \gamma_C)/(V_{AC} + \gamma_C) \\ &= ((V_{ABC} + \gamma_C)/V_A) / ((V_{AC} + \gamma_C)/V_A) \\ &= (V_{ABC}/V_A + \gamma_C) / (V_{AC}/V_A + \gamma_C) \\ &= (W_{BC} + \gamma_C)/(W_C + \gamma_C). \end{aligned}$$

Thus, (6.31) takes the form

$$g_-((W_{BC} + \gamma_C)/(W_C + \gamma_C)) \leq g_-(W_{BC}), \quad (6.32)$$

to be proven for all $W_{BC} > 0$. Now, since the measured matrix $(W_{BC} + \gamma_C)/(W_C + \gamma_C)$ is concave in γ_C , and g_- is decreasing and convex by Lemma 6.7, we can restrict ourselves to prove inequality (6.32) only in the case in which γ_C is a pure QCM, i.e. when it is symplectic.

Now we apply the above Lemma 6.9 (together with the remark immediately below it). Suppose we found a matrix $Z_{BC} \geq i\Omega_{BC}$ such that

$$\begin{pmatrix} W_{BC} & \Omega_{BC} \\ \Omega_{BC}^T & Z_{BC} \end{pmatrix} \geq 0, \quad \frac{1}{2} \log \det Z_{BC} = g_-(W_{BC}). \quad (6.33)$$

Then, consider the matrix

$$\begin{pmatrix} 0_B \oplus \gamma_C & 0_B \oplus \Omega_C^T \\ 0_B \oplus \Omega_C & 0_B \oplus \gamma_C \end{pmatrix} \geq 0, \quad (6.34)$$

where the last inequality holds because $\gamma \geq \Omega^T \gamma^{-1} \Omega$ for all $\gamma \geq i\Omega$, as Williamson's decomposition immediately reveals. Adding (6.34) to (6.33) we get

$$\begin{pmatrix} W_{BC} + \gamma_C & \Omega_B \oplus 0_C \\ \Omega_B^T \oplus 0_C & Z_{BC} + \gamma_C \end{pmatrix} \geq 0. \quad (6.35)$$

Taking the Schur complement with respect to the two C components, thanks to the two crucial zero blocks we have just formed, we obtain

$$\begin{pmatrix} (W_{BC} + \gamma_C)/(W_C + \gamma_C) & \Omega_B \\ \Omega_B^T & (Z_{BC} + \gamma_C)/(Z_C + \gamma_C) \end{pmatrix} \geq 0. \quad (6.36)$$

Remarkably, since $Z_{BC} \geq i\Omega_{BC}$ one finds easily $(Z_{BC} + \gamma_C)/(Z_C + \gamma_C) \geq i\Omega_B$. Therefore, the same Lemma 6.9 gives us

$$g_-((W_{BC} + \gamma_C)/(W_C + \gamma_C)) \leq \frac{1}{2} \log \det(Z_{BC} + \gamma_C)/(Z_C + \gamma_C). \quad (6.37)$$

The proof is ended once we show that

$$\det(Z_{BC} + \gamma_C)/(Z_C + \gamma_C) \leq \det Z_{BC}, \quad (6.38)$$

for all QCMs $Z_{BC} \geq i\Omega_{BC}$ and for all pure QCMs γ_C . This rather surprising fact is hard to prove at the level of QCMs, but it becomes more tractable once we go back to the Hilbert space picture behind. This can be done thanks to the identity $\text{Tr}[\rho^2] = 1/\sqrt{\det V}$, relating the *purity* $\text{Tr}[\rho^2]$ of a Gaussian state ρ to the determinant of its QCM V [210, §3.5]. Such an identity allows us to restate (6.38) as the claim that *purity of Gaussian states increases when pure Gaussian measurements are applied*.

Let us now translate also the measurement into the Hilbert space picture. Since γ_C is pure, the Gaussian measurement will be represented by a collection of rank-one (unnormalized) operators $\{\psi(t)_C\}_t$ such that $\int \frac{d^{2n_C}t}{(2\pi)^{n_C}} \psi(t)_C = \mathbb{1}_C$ (Subsection 5.2.3). Furthermore, it follows from (5.35), (5.36), (5.37), and (5.38) that the outcomes of this measurement on a Gaussian state ρ_{BC} with covariance matrix Z_{BC} are Gaussian states of the form $\tilde{\rho}(t)_B = U(t)^\dagger \tilde{\rho}_B U(t)$, where $U(t)$ are displacement unitaries depending on t and $\tilde{\rho}_B$ is a Gaussian state independent of t with QCM $(Z_{BC} + \gamma_C)/(Z_C + \gamma_C)$. With these hypotheses, we now see that Lemma 6.12 below allows us to conclude that $\text{Tr}[\tilde{\rho}_B^2] \geq \text{Tr}[\rho_{BC}^2]$, which immediately yields (6.38) since both $\tilde{\rho}_B$ and ρ_{BC} are Gaussian states.

(e) *The upper bound $\mathcal{G}(A)C)_V \leq g_+(V_{BC}/V_B)$ holds for any QCM V_{ABC} .*

Taking (6.13), applying g_- thanks to Lemma 6.7, and using the symplectic invariance of g_- (6.15) yields exactly

$$\mathcal{G}(A)C) = g_-(V_{AC}/V_A) \leq g_+(V_{BC}/V_B). \quad (6.39)$$

□

In dealing with point (d) of Theorem 6.10 above, we used some unproven property of Gaussian measurements with pure seeds. We now clarify this issue by stating two lemmas which complete the above argument.

Lemma 6.11. *Let*

$$\begin{pmatrix} A & X \\ X^\dagger & B \end{pmatrix} \geq 0 \quad (6.40)$$

be a positive definite block matrix. Then

$$\|X\|_2^2 \leq \|A\|_2 \|B\|_2, \quad (6.41)$$

where $\|M\|_2 := \sqrt{\text{Tr} M^\dagger M}$ denotes the Hilbert-Schmidt norm.

Proof. One writes

$$\begin{aligned}
 \|X\|_2^2 &= \text{Tr } XX^\dagger \\
 &= \text{Tr } A A^{-1/2} X X^\dagger A^{-1/2} \\
 &\stackrel{(1)}{\leq} \|A\|_2 \|A^{-1/2} X X^\dagger A^{-1/2}\|_2 \\
 &= \|A\|_2 \sqrt{\text{Tr } A^{-1/2} X X^\dagger A^{-1} X X^\dagger A^{-1/2}} \\
 &= \|A\|_2 \sqrt{\text{Tr } (X^\dagger A^{-1} X)^2} \\
 &= \|A\|_2 \|X^\dagger A^{-1} X\|_2 \\
 &\stackrel{(2)}{\leq} \|A\|_2 \|B\|_2.
 \end{aligned}$$

The justification of the steps is as follows: (1) is just the Cauchy-Schwartz inequality for the Hilbert-Schmidt product; (2) follows from the inequality $X^\dagger A^{-1} X \leq B$, deduced from Lemma 5.1, and from the fact that the Hilbert-Schmidt norm is an increasing function on positive matrices. \square

Lemma 6.12. *Let $\{\psi_C^i\}_i$ be a measurement composed of multiples of rank-one projectors. Assume that the outcomes of this measurement when performed on a bipartite system BC in a state ρ_{BC} are always unitarily equivalent to a fixed density operator on the remaining system B , i.e.*

$$\text{Tr}_C [\mathbb{1}_B \otimes \psi_C^i \rho_{BC}] = p_i U_i^\dagger \tilde{\rho}_B U_i \quad \forall i. \quad (6.42)$$

Then

$$\text{Tr} [\tilde{\rho}_B^2] \geq \text{Tr} [\rho_{BC}^2]. \quad (6.43)$$

Proof. For the sake of brevity, in what follows we suppose that i is an index running over a finite alphabet, but the argument below extends straightforwardly to the more general case in which it belongs to a measurable space. Since the identity

$$\tilde{\rho}_B = \frac{1}{p_i} U_i \text{Tr}_C [\mathbb{1}_B \otimes \psi_C^i \rho_{BC}] U_i^\dagger, \quad (6.44)$$

is valid for all indices i , we obtain

$$\begin{aligned}
 \text{Tr} [\tilde{\rho}_B^2] &= \|\tilde{\rho}_B\|_2^2 \\
 &= \left(\sum_i p_i \|\tilde{\rho}_B\|_2 \right)^2 \\
 &= \left(\sum_i \left\| \text{Tr}_C [\mathbb{1}_B \otimes \psi_C^i \rho_{BC}] \right\|_2 \right)^2 \\
 &= \sum_{ij} \left\| \text{Tr}_C [\mathbb{1}_B \otimes \psi_C^i \rho_{BC}] \right\|_2 \left\| \text{Tr}_C [\mathbb{1}_B \otimes \psi_C^j \rho_{BC}] \right\|_2, \quad (6.45)
 \end{aligned}$$

where we omitted the subscripts C for the sake of brevity. Now, using the notation $\psi^i = |\psi^i\rangle\langle\psi^i|$ (with $|\psi^i\rangle$ not necessarily normalised) consider the map $\mathcal{N}_{C \rightarrow C'}$ from C to a new system C' as defined by

$$\mathcal{N}(X) := \left(\sum_i |i\rangle\langle\psi^i| \right) X \left(\sum_j |j\rangle\langle\psi^j| \right)^\dagger = \sum_{ij} |i\rangle\langle\psi^i| X |\psi^j\rangle\langle j|. \quad (6.46)$$

Obviously, \mathcal{N} is completely positive and trace-preserving. In particular, we deduce that $(I_B \otimes \mathcal{N}_{C \rightarrow C'}) (\rho_{BC}) \geq 0$ is a legitimate quantum state. This latter density matrix has blocks indexed by i, j and given by

$$\mathrm{Tr}_C [\mathbf{1}_B \otimes |\psi^j\rangle\langle\psi^i|_C \rho_{BC}].$$

Thanks to Lemma 6.11, we know that for all $i \neq j$,

$$\begin{aligned} & \left\| \mathrm{Tr}_C [\mathbf{1}_B \otimes |\psi^j\rangle\langle\psi^j|_C \rho_{BC}] \right\|_2^2 \\ & \leq \left\| \mathrm{Tr}_C [\mathbf{1}_B \otimes \psi_C^i \rho_{BC}] \right\|_2 \left\| \mathrm{Tr}_C [\mathbf{1}_B \otimes \psi_C^j \rho_{BC}] \right\|_2. \end{aligned}$$

This is also trivially true when $i = j$. Therefore,

$$\begin{aligned} & \sum_{ij} \left\| \mathrm{Tr}_C [\mathbf{1}_B \otimes \psi_C^i \rho_{BC}] \right\|_2 \left\| \mathrm{Tr}_C [\mathbf{1}_B \otimes \psi_C^j \rho_{BC}] \right\|_2 \\ & \geq \sum_{ij} \left\| \mathrm{Tr}_C [\mathbf{1}_B \otimes |\psi^j\rangle\langle\psi^i|_C \rho_{BC}] \right\|_2^2 \\ & = \sum_{ij} \mathrm{Tr}_{BC} \left[(\mathbf{1}_B \otimes \psi_C^i) \rho_{BC} (\mathbf{1}_B \otimes \psi_C^j) \rho_{BC} \right] \\ & = \mathrm{Tr} [\rho_{BC}^2], \end{aligned}$$

where we used the normalization condition $\sum_i \psi_C^i = \mathbf{1}_C$. Inserting the above inequality into (6.45) yields the claim (6.43). \square

Remark. By following the above argument, we see that all claims of Theorem 6.10 with the exception of (d) hold for arbitrary continuous variable states, not necessarily Gaussian. On the other hand, our current proof of the monotonicity of $\mathcal{G}(A)B$ under general, non-deterministic Gaussian maps on B relies on the specification to Gaussian states of AB . We leave it as an open problem whether a more general proof of part (d) could be obtained, valid even for non-Gaussian states.

6.3.4 Monogamy of Gaussian steerability

Our framework allows us to address the general problem of the monogamy of $\mathcal{G}(A)B$. For a state with QCM $V_{AB_1 \dots B_k}$, consider the following inequalities

$$\mathcal{G}(A)B_1 \dots B_k \geq \sum_{j=1}^k \mathcal{G}(A)B_j, \quad (6.47)$$

$$\mathcal{G}(B_1 \dots B_k)A \geq \sum_{j=1}^k \mathcal{G}(B_j)A. \quad (6.48)$$

In a very recent study [261], both inequalities were proven in the special case of a $(k + 1)$ -mode system with one single mode per party, i.e. $n_A = n_{B_j} = 1$ ($j = 1, \dots, k$). We now show that only one of these constraints holds in full generality.

Theorem 6.13 (Monogamy of Gaussian steerability). *(a) Inequality (6.47) holds for any multimode QCM $V_{AB_1 \dots B_k}$. (b) Inequality (6.48) holds for any multimode QCM $V_{AB_1 \dots B_k}$ such that either A comprises a single mode ($n_A = 1$), or $V_{AB_1 \dots B_k}$ is pure, but can be violated otherwise.*

Proof. (a) It suffices to prove the inequality $\mathcal{G}(A)BC \geq \mathcal{G}(A)B + \mathcal{G}(A)C$ for a tripartite QCM V_{ABC} , as (6.47) would follow by iteration. Observe that V_{AB}/V_A and V_{AC}/V_A form the diagonal blocks of the bipartite matrix V_{ABC}/V_A . Then, applying (6.23) yields

$$\begin{aligned} \mathcal{G}(A)BC)_V &= g_-(V_{ABC}/V_A) \\ &\geq g_-(V_{AB}/V_A) + g_-(V_{AC}/V_A) \\ &= \mathcal{G}(A)B)_V + \mathcal{G}(A)C)_V, \end{aligned}$$

which concludes the proof. (b) For the case $n_A = 1$ with n_{B_j} arbitrary, one exploits the fact that only one term $\mathcal{G}(B_j)A$ in the right-hand side of (6.48) can be nonzero, due to the impossibility of jointly steering a single mode by Gaussian measurements as implied by (6.9) (see the original papers [266, 234]), combined with the monotonicity of $\mathcal{G}(B_1 \dots B_k)A$ under partial traces on the steering party as implied by Theorem 6.10. Finally, the case when $V_{AB_1 \dots B_k}$ is pure follows from the forthcoming Corollary 6.15, namely by combining the last claim of Theorem 6.14 with (6.62):

$$\begin{aligned} \mathcal{G}(B_1 \dots B_k)A)_V &= E_{F,2}^G(B_1 \dots B_k : A)_V \\ &\geq \sum_{j=1}^k E_{F,2}^G(B_j : A)_V \\ &\geq \sum_{j=1}^k \mathcal{G}(B_j)A)_V, \end{aligned}$$

where the first equality holds specifically for pure states. \square

The Gaussian steerability is thus not monogamous with respect to a common steered party A when the latter is made of two or more modes, with violations of (6.48) existing already in a tripartite setting ($k = 2$) with $n_{B_1} = n_{B_2} = 1$ and $n_A = 2$; a counterexample is reported in Appendix G. What is truly monogamous is the log-determinant of the Schur complement, which only happens to be directly linked to the function g_- when $n_A = 1$.

6.4 Correlations hierarchy in Gaussian states

The previous section was concerned with the problem of quantifying steering in an appropriate way. For bipartite Gaussian states, we found that a suitable quantifier based directly on the covariance matrix does the job. In this section, we look at the more general problem of measuring other kinds of correlations (both classical and quantum) by looking at the QCM. In Subsection 6.4.1 we provide a brief introduction to an axiomatic framework to deal with the problem that was put forward in [18]. In the most favourable case, such framework requires a strong constraint to be met, namely that the quantum part of the correlations should not exceed *half* of the total. After having introduced Rényi-2 Gaussian quantifiers in Subsection 6.4.2, in the following Subsection 6.4.3 we proceed to show that bipartite Gaussian states do satisfy this strong inequality when their correlations are measured with said Rényi-2 measures. This observation is the content of Theorem 6.14, which constitutes the main result of the section. Finally, Subsection 6.4.4 presents some applications of this result, most importantly concerning the monogamy of the Rényi-2 Gaussian entanglement of formation (Corollary 6.15).

6.4.1 Correlations in bipartite quantum systems

Given a bipartite quantum state ρ_{AB} , can we draw a line between its classical and its quantum correlations? How to measure the two, and what requirements to impose on candidate quantifiers? These questions have been debated extensively in the literature [277, 278, 279, 18], and we do not attempt to give a full account of the discussion here. However, for the sake of the presentation, let us summarise briefly the point of view of [18].

The **quantum mutual information** is given by

$$I(A : B)_\rho := S(\rho_A) + S(\rho_B) - S(\rho_{AB}), \quad (6.49)$$

where $S(\rho) := -\text{Tr}[\rho \log \rho]$ is the von Neumann entropy given by (3.30).⁴ As it turns out, (6.49) is an excellent candidate for a measure of total correlations, because of its operational significance. In fact, it has been shown that the mutual information quantifies the maximal amount of information that can be sent securely using ρ_{AB} in a one-time-pad cryptographic scheme [280], and also the minimal amount of noise that is needed to erase the correlations of ρ_{AB} [279]. In view of all these considerations, we will take (6.49) as our measure of total correlations.

Now, assume that we can always divide the correlations in $I(A : B)$ between classical and quantum. That is, assume that we can come up with two functions $C(\rho)$ and $Q(\rho)$, which measure respectively the classical and quantum correlations, and such that $C + Q = I(A : B)$. As argued in [18], a reasonable

⁴This is nothing but a special case of the already mentioned conditional mutual information (3.29), which reduces to (6.49) for a decorrelated C .

assumption to make is that the inequality

$$Q(\rho) \leq C(\rho) \tag{6.50}$$

holds for all states ρ . Since $Q + C = I(A : B)$, (6.50) is equivalent to

$$Q(\rho) \leq \frac{1}{2}I(A : B)_\rho. \tag{6.51}$$

An intuitive argument in favour of taking (6.50) (or (6.51)) as a postulate goes as follows [18].

First, examine its status for pure states. On the one hand, the quantum correlations contained in a pure state $|\psi\rangle_{AB}$ should be measured by its entanglement, and virtually any measure of entanglement is commonly required to reduce to the **entanglement entropy** $S(\text{Tr}_B \psi_{AB})$ on pure states [110, §15.6]. Therefore, we can take as granted that $Q(\psi) = S(\text{Tr}_B \psi_{AB})$ holds on pure states. On the other hand, the classical correlations as measured for instance by the *classical mutual information*, i.e. by the maximal mutual information that is obtainable by local measurements, are also known to coincide with the entanglement entropy for pure states [281]. This leads us to the assumption that $Q(\psi) = C(\psi)$ holds for pure states.

Mixed states are naturally obtained as classical mixtures of pure states. Classical mixing can increase C but it should never increase the quantum correlations Q . Given that the two are equal on pure states, inequality (6.50) seems like a reasonable expectation for general mixed states. This matches also the intuition of the authors of [18, 279]. And in fact, many entanglement measures Q do satisfy (6.51), most notably the squashed entanglement (3.28) we already encountered in Chapter 3. However, some other, equally important entanglement monotones, such as the entanglement of formation, do not obey (6.51) [18]. Even worse, the entanglement of formation can even exceed the quantum mutual information itself, as observed in [282, 18].

We do not take a firm position on the status of (6.51) as a fundamental assumption. Indeed, we do not see any a priori clear reason why there should be a way to divide the total correlations into a classical and a quantum part. On the contrary, it could well be that they are inextricably linked and that any attempt of separating them will lead eventually, *in the most general case*, to a contradiction. Instead, here we regard (6.51) more as a *desirable* property that we can hope to find to hold in some special cases, i.e. for some restricted sets of states and for suitable measures of correlations. This is exactly what we will establish in the next subsection: *for Gaussian states, quantifiers based on the Rényi-2 entropy do satisfy the inequality (6.51)*, and can thus be regarded – from this point of view – as defining a desirable framework for the problem.

6.4.2 Rényi-2 Gaussian entanglement of formation

The aim of this subsection is to provide the reader with an introduction to Rényi-2 quantifiers for Gaussian states, extensively discussed in [260, 209].

We start by recalling the notion of **quantum Rényi- α entropy** of a state ρ , given by

$$S_\alpha(\rho) := \frac{1}{1-\alpha} \log \text{Tr} [\rho^\alpha]. \quad (6.52)$$

It is not hard to show that (6.52) reproduces the conventional von Neumann entropy in the limit $\alpha \rightarrow 1$. As expected, the Rényi- α entropy becomes a function of the covariance matrix alone for Gaussian states. Adopting the shorthand notation $S_\alpha(\rho^G(V, w)) = S_\alpha(V)$, we have

$$S_\alpha(V) = -\frac{1}{\alpha-1} \sum_{i=1}^n \log \frac{2^\alpha}{(\nu_i(V)+1)^\alpha - (\nu_i(V)-1)^\alpha} \quad (6.53)$$

for $\alpha > 1$, and

$$S_1(V) = \sum_{i=1}^n \left(\frac{\nu_i(V)+1}{2} \log \frac{\nu_i(V)+1}{2} - \frac{\nu_i(V)-1}{2} \log \frac{\nu_i(V)-1}{2} \right) \quad (6.54)$$

for the von Neumann case $\alpha = 1$. For details, see [209, Eq. (108) and (109)] or [210, Eq. (3.96)]. Here we are mostly interested in the case $\alpha = 2$, in which case we have

$$S_2(\rho) = \frac{1}{2} \log \det V \quad (6.55)$$

for an arbitrary Gaussian state ρ with QCM V .

Rényi entropies can be used to define a corresponding measure of entanglement. For a bipartite quantum state ρ_{AB} , the **Rényi- α entanglement of formation** is constructed as the convex hull of the Rényi- α entropy of entanglement defined on pure states [92], i.e.

$$E_{F,\alpha}(A : B)_\rho := \inf \sum_i p_i S_\alpha(\psi_A^i) \quad (6.56)$$

s.t. $\rho_{AB} = \sum_i p_i \psi_{AB}^i$,

where ψ_{AB}^i are density matrices of pure states, $\psi_A^i = \text{Tr}_B \psi_{AB}^i$ is the reduced state (marginal), and S_α is defined in (6.52).

For quantum Gaussian states, an upper bound to this quantity can be derived by restricting the decompositions appearing in the above infimum to be comprised of pure Gaussian states only. One obtains what is called **Rényi- α Gaussian entanglement of formation**, a monotone under Gaussian local operations and classical communication, that in terms of the QCM V_{AB} of ρ_{AB} is given by the simpler formula [263]

$$E_{F,\alpha}^G(A : B)_V = \inf S_\alpha(\gamma_A) \quad (6.57)$$

s.t. γ_{AB} pure QCM and $\gamma_{AB} \leq V_{AB}$,

where with a slight abuse of notation we denoted with $S_\alpha(W)$ the Rényi- α entropy of a Gaussian state with QCM W as given by (6.53), and γ_{AB} stands

for the QCM of a pure Gaussian state. Observe that $E_{F,\alpha}(A : B)$ and also $E_{F,\alpha}^G$ are symmetric under the exchange of A and B , since *the entropies of the two local shares of a pure state are the same*. Incidentally, it has been proven [264, 283] that for some two-mode Gaussian states the formula (6.57) reproduces exactly (6.56), i.e. Gaussian decompositions in (6.56) are globally optimal.

Remark. We wrote an infimum in (6.57), but with little work one can show that this is in fact a minimum, because the set over which we are optimising is compact, and the objective function is clearly uniformly continuous (remember that $\det \gamma_A \geq 1$).

The most commonly used $E_{F,\alpha}$ is the one corresponding to the von Neumann entropy, $\alpha = 1$. However, we already saw that in the Gaussian setting the choice $\alpha = 2$ is also natural, an intuitive reason being that it respects the quadratic nature of the states. More operationally motivated reasons will be discussed in the forthcoming Chapter 7. Thus, from now on we will focus on the case $\alpha = 2$. Under this assumption, thanks to (6.55) we see that (6.57) becomes

$$E_{F,2}^G(A : B)_V = \inf \frac{1}{2} \log \det \gamma_A \quad (6.58)$$

s.t. γ_{AB} pure QCM and $\gamma_{AB} \leq V_{AB}$.

We will find it convenient to rewrite the above equation in a slightly different form, that employs the **Rényi-2 mutual information**

$$I_M(A : B)_V := \frac{1}{2} \log \frac{\det V_A \det V_B}{\det V_{AB}}, \quad (6.59)$$

a quantity that will be the subject of more thorough investigation in Chapter 7. Using the readily verified fact that $I_M(A : B)_\gamma = \log \det \gamma_A = \log \det \gamma_B$ when γ_{AB} is the QCM of a pure state [209] (essentially because $\det \gamma_{AB} = 1$ and the two marginals have equal entropies), we obtain

$$E_{F,2}^G(A : B)_V = \inf \frac{1}{2} I_M(A : B)_\gamma \quad (6.60)$$

s.t. γ_{AB} pure QCM and $\gamma_{AB} \leq V_{AB}$.

Among other things, (6.60) has the advantage of being evidently faithful on Gaussian states [260]. In other words, it becomes zero iff the condition (5.44) in Lemma 5.4 is met, i.e. iff the Gaussian state with QCM V_{AB} is separable.

6.4.3 Gaussian entanglement and correlations hierarchy

In this subsection we prove that the inequality (6.51) holds for Gaussian states when Q is chosen to be the Rényi-2 Gaussian entanglement of formation of (6.58). In fact, in Appendix G we show that one could rather take Q equal to *any* Rényi- α Gaussian entanglement of formation as long as $\alpha \geq 2$. Our main result reads as follows.

Theorem 6.14 (Gaussian Rényi-2 correlations hierarchy). *Let V_{AB} be an arbitrary bipartite QCM. Then the Rényi-2 mutual information (6.59), the Rényi-2 Gaussian entanglement of formation (6.58), and the Gaussian steerability (6.8) satisfy*

$$\frac{1}{2}I_M(A : B)_V \geq E_{F,2}^G(A : B)_V \geq \mathcal{G}(A)B)_V. \quad (6.61)$$

If V_{AB} is pure, all the above three quantities coincide with the reduced Rényi-2 entropy $\frac{1}{2} \log \det V_A$.

Proof. First of all, let us prove the rightmost inequality. Consider the optimal pure QCM $\gamma_{AB} \leq V_{AB}$ that saturates the infimum (actually, the minimum) in (6.58). Then write

$$\begin{aligned} E_{F,2}^G(A : B)_V &= \frac{1}{2} \log \det \gamma_B \\ &\stackrel{(1)}{=} g_+(\gamma_B) \\ &\stackrel{(2)}{=} g_-(\gamma_B^{-1}) \\ &\stackrel{(3)}{=} g_-(\Omega_B^T \gamma_B^{-1} \Omega_B) \\ &\stackrel{(4)}{=} g_-(\gamma_{AB}/\gamma_A) \\ &= \mathcal{G}(A)B)_\gamma \\ &\stackrel{(5)}{\geq} \mathcal{G}(A)B)_V. \end{aligned}$$

These steps can be justified as follows: (1) follows straight from the definition of g_+ , (6.7), together with the fact that γ_B is a legitimate QCM and hence all its symplectic eigenvalues are no smaller than 1; (2) is an application of (6.16); (3) uses the symplectic invariance of g_\pm , (6.15); (4) follows from Lemma 6.3; and (5) is a consequence of the anti-monotonicity of $\mathcal{G}(A)B)_V$ in V , Theorem 6.10(a).

We now move on to the proof of the leftmost inequality in (6.61). The key step leverages the construction of a special pure state to use as an ansatz in (6.60). Consider $\gamma_V^\# = V \# (\Omega V^{-1} \Omega^T)$, which is a pure QCM by Lemma 5.10. Plugging $\gamma_V^\#$ into (6.60) is possible again thanks to Lemma 5.10(c), which guarantees that the inequality $\gamma_V^\# \leq V$ holds. Doing this we obtain

$$\begin{aligned} E_{F,2}^G(A : B)_V &\leq \frac{1}{2} \log \det (\gamma_V^\#)_A \\ &= \frac{1}{2} \log \det (\Pi_A (V_{AB} \# (\Omega_{AB} V_{AB}^{-1} \Omega_{AB}^T)) \Pi_A^T). \end{aligned}$$

where we denoted by Π_A the projector onto the A component. To proceed further, we employ the inequality (5.18) [242, Theorem 3], with $\Lambda(\cdot) \mapsto \Pi_A(\cdot)\Pi_A^T$,

$A \mapsto V_{AB}$ and $B \mapsto \Omega_{AB} V_{AB}^{-1} \Omega_{AB}^T$, finding

$$\begin{aligned} \Pi_A (V_{AB} \# (\Omega_{AB} V_{AB}^{-1} \Omega_{AB}^T)) \Pi_A^T &\leq V_A \# (\Pi_A (\Omega_{AB} V_{AB}^{-1} \Omega_{AB}^T) \Pi_A^T) \\ &= V_A \# (\Omega_A (V_{AB}/V_B)^{-1} \Omega_A^T), \end{aligned}$$

where for the last step we used the formula (5.6) for the inverse of a 2×2 block matrix. Inserting this operator inequality into the above upper bound for $E_{F,2}^G(A : B)_V$ we obtain

$$\begin{aligned} E_{F,2}^G(A : B)_V &\leq \frac{1}{2} \log \det (V_A \# (\Omega_A (V_{AB}/V_B)^{-1} \Omega_A^T)) \\ &= \frac{1}{4} \log \frac{\det V_A}{\det V_{AB}/V_B} \\ &= \frac{1}{4} \log \frac{\det V_A \det V_B}{\det V_{AB}} \\ &= \frac{1}{2} I_M(A : B)_V, \end{aligned}$$

and we are done. \square

Intuitively, this proves that the involved measures quantitatively capture the general hierarchy of correlations [284] in arbitrary Gaussian states [209]: the Gaussian steerability is generally smaller than the entanglement degree, which accounts for a portion of quantum correlations up to half the total ones.

Theorem 6.14 is a significant improvement over previous results. For instance, using [260, Eq. (14) and (17)] we deduce the weaker inequality $E_{F,2}^G(A : B) \leq I_M(A : B)$ (proven there only for two-mode Gaussian states). We will generalise the inequality (6.61) in the next chapter, where we will show that on the leftmost side one can put instead the Rényi-2 *conditional* mutual information computed on *any* tripartite Gaussian extension of the state (Theorem 7.11).

6.4.4 Applications

An outstanding consequence of Theorem 6.14 is that the Rényi-2 measure of entanglement can now be proven monogamous for *arbitrary* Gaussian states with any number of modes per party.

Corollary 6.15 (Monogamy of Gaussian Rényi-2 entanglement). *The Gaussian Rényi-2 entanglement of formation (6.58) is monogamous for any multipartite Gaussian state, i.e.*

$$E_{F,2}^G(A : B_1 \dots B_k)_V \geq \sum_{j=1}^k E_{F,2}^G(A : B_j)_V. \quad (6.62)$$

for all QCMs $V_{AB_1 \dots B_k}$.

Proof. It suffices again to prove that

$$E_{F,2}^G(A : BC)_V \geq E_{F,2}^G(A : B)_V + E_{F,2}^G(A : C)_V$$

holds for any tripartite QCM V_{ABC} . In order to do this, take the pure QCM $\gamma_{ABC} \leq V_{ABC}$ that saturates the infimum in the definition of $E_{F,2}^G(A : BC)$, and notice that

$$\begin{aligned} E_{F,2}^G(A : BC) &= \frac{1}{2} \log \det \gamma_A \\ &= \frac{1}{2} I_M(A : BC)_\gamma \\ &= \frac{1}{2} I_M(A : B)_\gamma + \frac{1}{2} I_M(A : C)_\gamma, \end{aligned}$$

where the last equality holds specifically for pure states, being ultimately a consequence of the fact that the two marginals of any pure state have equal Rényi entropies. Applying (6.61) to each of the two rightmost addends yields

$$\begin{aligned} E_{F,2}^G(A : BC)_V &\geq E_{F,2}^G(A : B)_\gamma + E_{F,2}^G(A : C)_\gamma \\ &\geq E_{F,2}^G(A : B)_V + E_{F,2}^G(A : C)_V, \end{aligned}$$

where the last step follows as $E_{F,2}^G$ is a non-increasing function of the QCM, which is clear already from the definition (6.58). \square

Corollary 6.15 yields the *most general* result to date regarding quantitative monogamy of continuous variable entanglement [209], as all previous proofs (for the Rényi-2 measure [260] or other quantifiers [285, 286]) were restricted to the special case of *one* mode per party. A monogamy inequality is a powerful tool in dealing with entanglement measures. For instance, when combined with monotonicity under local operations, it leads to the additivity of the measure under examination.

Corollary 6.16. *The Gaussian Rényi-2 entanglement of formation (6.58) is additive under tensor products (equivalently, direct sum of covariance matrices). In formula,*

$$E_{F,2}^G(A_1 A_2 : B_1 B_2)_{V_{A_1 B_1} \oplus W_{A_2 B_2}} = E_{F,2}^G(A_1 : B_1)_V + E_{F,2}^G(A_2 : B_2)_W. \quad (6.63)$$

Proof. Applying first (6.62) and then the monotonicity of $E_{F,2}^G$ under the operation of discarding some local subsystems, we obtain

$$\begin{aligned} &E_{F,2}^G(A_1 A_2 : B_1 B_2)_{V_{A_1 B_1} \oplus W_{A_2 B_2}} \\ &\geq E_{F,2}^G(A_1 A_2 : B_1)_{V_{A_1 B_1} \oplus W_{A_2}} + E_{F,2}^G(A_1 A_2 : B_2)_{V_{A_1} \oplus W_{A_2 B_2}} \\ &\geq E_{F,2}^G(A_1 : B_1)_V + E_{F,2}^G(A_2 : B_2)_W. \end{aligned}$$

The opposite inequality follows by inserting factorised ansatzes $\gamma_{A_1 B_1} \oplus \tau_{A_2 B_2}$ into (6.58). \square

As established in this section, the Rényi-2 Gaussian entanglement of formation alias Rényi-2 Gaussian squashed entanglement also emerges as a rare example of an additive entanglement monotone (within the Gaussian framework) which satisfies the general monogamy inequality (6.62). By contrast, the conventional (Rényi-1) entanglement of formation can not fundamentally be monogamous [287].

6.5 Conclusions

We have derived fundamental inequalities for the Schur complement of positive semidefinite matrices and explored their far-reaching applications to quantum information theory. This enabled us to recover seemingly unrelated findings from recent literature, like the strong subadditivity for log-determinant of covariance matrices [260, 266, 234] and basic properties of measures of continuous variable entanglement [260] and steering [17, 261], and to reach substantially beyond. In particular, we proved that the Gaussian steerability [17, 271] for Gaussian states is a convex monotone under Gaussian local operations and classical communication, i.e. it is a fully fledged steering measure [267] within the Gaussianity restriction; we further proved it is monogamous with respect to the steering party but not with respect to the steered party if the latter has more than one mode and the overall state is mixed. We also proved that the Gaussian Rényi-2 measure of entanglement [260] is monogamous for any Gaussian state with an arbitrary number of modes per party. This key result is a simple corollary of a general hierarchical relation here established for measures of correlations based on log-determinant of covariance matrices.

This work further reveals how pursuing *prima facie* technical advances in classical information theory and linear algebra can significantly impact on the identification of possibilities and limitations for quantum technologies, which had eluded a general quantitative analysis so far. It will be worth investigating adaptations of our results to the study of quantum correlations in discrete variable stabiliser states, useful resources for quantum computing [288] which share deep mathematical analogies with continuous variable Gaussian states [289, 265].

Chapter 7

From log-det inequalities to Gaussian entanglement via recoverability theory

7.1 Introduction

In the preceding Chapter 6 we stumbled upon a matrix inequality dubbed *strong subadditivity of log-det entropy* (log-det SSA), which is reported here as (7.6). We showed that this relation has a major impact on Gaussian quantum information, either directly or through suitable generalisations. In this chapter we develop further insights on the properties of the above inequality and its applications to classical and quantum information theory.

The material is structured as follows. The rest of this section is devoted to introducing the framework where our work belongs (Subsection 7.1.1) and to pointing the reader to our main contributions (Subsection 7.1.2). In Section 7.3 we derive various characterisations of the case of saturation of log-det SSA (7.6) with equality. Then, in Section 7.4 we turn to the case of near-saturation, which leads us on the one hand to the theory of recovery maps, and on the other hand to establishing simple and faithful lower bounds on the log-det conditional mutual information. Up to that point, all results hold for general positive definite matrices. After that, in Section 7.5 we turn our attention to quantum Gaussian states and their quantum covariance matrices. There, we introduce a measure of entanglement for quantum Gaussian states based on the log-det conditional mutual information and prove its faithfulness and additivity. Quite remarkably, we show that the measure coincides with the Rényi-2 Gaussian entanglement of formation introduced in [260], equipping the latter with an interesting operational interpretation in the context of recoverability. We conclude in Section 7.6 with a number of open questions.

7.1.1 Building bridges between matrix analysis, probability theory and Gaussian quantum information

The concept of Gaussian random variable, which is of central importance in this last part of the present thesis, has been long known to constitute a bridge between the two seemingly unrelated fields of matrix analysis and probability theory. The key of this correspondence is to construct, given a positive definite $n \times n$ matrix $A > 0$, an n -dimensional Gaussian random variable X with covariance matrix A . One can show that all the Rényi entropies of the random variable X are equal (up to a constant that depends only on n) to the log-det entropy we defined in Chapter 6, given by $\frac{1}{2} \ln \det A$. This provides us with a systematic way to turn information theoretical inequalities into determinantal inequalities for positive matrices, a technique which has been the subject of growing interest in the last decades (see e.g. the reviews given in [276, 290]). In fact, the strong subadditivity of log-determinant entropies (7.6) can be immediately obtained in this way.

Analogously to their classical counterparts, also quantum Gaussian states are completely described by their covariance matrix (up to an often irrelevant displacement operation). Moreover, the outcomes of Gaussian measurements performed on Gaussian states are (unsurprisingly) classical Gaussian random variables themselves, which allows us to connect also this framework to the one described above. In particular, when one looks at Shannon-entropic quantities constructed out of *measured* correlations of quantum Gaussian states, the relevant quantity is again the log-det entropy.

One could wonder whether there is a more mathematically direct way to obtain the log-det entropy starting from the quantum state itself. As already mentioned in the previous chapter, one such way is to look at Rényi-2 quantifiers. In fact, the Rényi-2 entropy, given in general by the formula $S_2(\rho) := -\ln \text{Tr}[\rho^2]$ (see (6.52)), reduces to $S_2(\rho) = \frac{1}{2} \log \det V$ (6.55) when evaluated on Gaussian states. This paves the way to direct operational interpretations for the Rényi-2 entropy in the Gaussian setting.

To summarise, when one looks at measured correlation quantifiers in the quantum Gaussian setting, the natural quantities that appears is the Shannon entropy of the measurement outcome, which in turn reduces to the log-det entropy of the (quantum) covariance matrix and thus coincides (up to additive constants) with the Rényi-2 entropy of the quantum state itself.

Note. All entropies that appear in this chapter are measured with the natural logarithm \ln instead of the binary logarithm \log . This is in agreement with the most common conventions in continuous variable information theory.

Note. Throughout this chapter, positive definite matrices $V > 0$ will be occasionally referred to as *covariance matrices*, a nomenclature that is very much in the spirit of highlighting the connections with probability theory. As usual, we instead call *quantum covariance matrix* (QCM) any matrix V of even size that satisfies the stronger condition $V + i\Omega \geq 0$, corresponding to the Heisenberg

principle for Gaussian states (5.29). Here, Ω is the standard symplectic form defined in (5.24).

7.1.2 Our contributions

This chapter is based on the homonymous paper [236]:

- L. Lami, C. Hirche, G. Adesso, and A. Winter. From log-determinant inequalities to Gaussian entanglement via recoverability theory. *IEEE Trans. Inf. Theory*, Preprint, 2017.

Our goal here is to conduct a deeper investigation of the log-det SSA (7.6), and to explore some of its far-reaching implications in quantum information theory. Our analysis rests crucially on the connection between Gaussian random variables and positive definite matrices we have outlined here, which allows us to use tools taken from matrix analysis [240] to explore properties of the log-det conditional mutual information (7.11), an approach we already followed in the previous Chapter 6.

More in detail, in Section 7.3 we will show how to find known and new necessary and sufficient conditions under which saturation with equality occurs (Theorem 7.3). Along the way, we develop new insights into the properties of information theoretic quantities based on the log-det entropy. Most notably, in Theorem 7.2 we show that: (a) log-det mutual information and *conditional* mutual information can be obtained from one another by taking the *inverse* of the covariance matrix; and (b) the log-det mutual information is convex on the geodesics defined by the *trace metric*, defined on the sets of positive definite matrices A by the formula $ds^2 = \text{Tr}[(A^{-1}dA)^2]$. This is surprising, since in general the mutual information is known to be neither convex nor concave in the probability distribution. Moreover, since this metric is also used to define the matrix geometric mean (see (5.19)), the above result establishes further connections between matrix analysis and information theory, in the very spirit of the present work.

Subsequently, in Section 7.4 we move on to the problem of improving the log-det SSA inequality, with special emphasis on the case of near saturation. In this context, we discuss the role of the classical transpose channel, also known as Petz recovery map, and find its action explicitly (Proposition 7.4). We then use this result to prove some extensions of the saturation theorem, in the form of faithful lower bounds on the log-det conditional mutual information. Exploiting the aforementioned connection with the theory of matrix means, we further establish a somewhat different lower bound – of a less information theoretical and more matrix analytic nature – from which the saturation conditions are easily readable (Theorem 7.7).

Section 7.5 is devoted to exploring some of the applications of log-det SSA in the field of Gaussian quantum information. As we said, for Gaussian states the log-det entropy is equivalent to the Rényi entropy of order 2. Our first result, Theorem 7.11, provides a strengthening of log-det SSA for quantum covariance matrices that involves the Rényi-2 Gaussian entanglement of formation, an

entanglement measure we already encountered in Chapter 6 (see (6.58)). We then employ this result to define a log-det entropy equivalent of the squashed entanglement (3.28), which is then shown to coincide with the Rényi-2 Gaussian entanglement of formation (Theorem 7.14). In particular, we prove that in the simpler Gaussian setting there is no need for extensions of unbounded size, which makes the measure efficiently computable. This is remarkable, since no attempt to establish a similar result for the standard squashed entanglement based on the von Neumann entropy has succeeded yet. Finally, simpler proofs of useful properties like monogamy, faithfulness, and additivity can be found thanks to the identity between these two Gaussian entanglement measures.

7.2 Mathematical preliminaries

The goal of this introductory section is to acquaint the reader with some of the tools we will use in the rest of the chapter. Namely, in Subsection 7.2.1 we discuss the basics of information theory with Gaussian random variables, explaining in detail the nature of the connection we mentioned in Subsection 7.1.1. Next, Subsection 7.2.2 is devoted to a more systematic introduction of log-det entropies, which we already encountered at the end of Chapter 6. Finally, in Subsection

7.2.1 Information theory with Gaussian random variables

As we already mentioned, the idea of using information theoretical reasoning to prove determinantal inequalities for positive definite matrices has received much attention recently [276, 290]. The key of the above correspondence is to associate, to each positive matrix¹ $A \in \mathcal{M}_n(\mathbb{R})$, an n -dimensional Gaussian random variable $X \in \mathbb{R}^n$ with mean 0 and variance (aka covariance matrix) $\text{Var } X = \mathbb{E} X X^T = A$. The density of X is given by

$$p_A(x) = \frac{e^{-\frac{1}{2}x^T A^{-1}x}}{\sqrt{(2\pi)^n \det A}}. \quad (7.1)$$

This has the nice feature that for two independent Gaussian random variables X and Y with mean 0 and covariance matrices A and B , respectively, the sum $A + B$ is the covariance matrix of $X + Y$.

Under the density (7.1), the differential entropy

$$h(X) := - \int d^n x p_A(x) \ln p_A(x) \quad (7.2)$$

of (7.1) takes the form

$$h(X) = \frac{1}{2} \ln \det A + \frac{n}{2} (\ln 2\pi + 1), \quad (7.3)$$

¹In this chapter we consider only real matrices since they are more relevant for the applications we are interested in, but all the results we find apply also to the hermitian case with minor modifications.

while the relative entropy $D(p_A \| p_B) := \int d^n x p_A(x) \ln \frac{p_A(x)}{p_B(x)}$ is given by

$$D(p_A \| p_B) = \frac{1}{2} \ln \frac{\det B}{\det A} + \frac{1}{2} \text{Tr}(B^{-1}A) - \frac{n}{2}. \quad (7.4)$$

Here and in the remainder of the chapter we denote by \ln the natural logarithm. The positivity of (7.4) as a function of the matrices A and B can be seen as an instance of Klein's inequality applied to the natural logarithm [291].

In this picture, general inequalities involving entropies can be turned into inequalities involving determinants thanks to (7.3) and (7.4). A prominent example of the usefulness of this approach is constituted by *strong subadditivity* (SSA), the basic 'Shannon-type' entropy inequality [292]. Consider a Gaussian distributed vector $X_{ABC} = (X_A, X_B, X_C)^T \in \mathbb{R}^{n_A+n_B+n_C}$ with covariance matrix V_{ABC} :

$$V_{ABC} = \begin{pmatrix} A & X & Y \\ X^T & B & Z \\ Y^T & Z^T & C \end{pmatrix} \geq 0. \quad (7.5)$$

The positivity of the conditional mutual information $I(X_A : X_B | X_C)$ then yields the already encountered strong subadditivity of log-det entropy (6.1). Since this inequality is our main subject of study in this chapter, we report it here:

$$\ln \det V_{AC} + \ln \det V_{BC} - \ln \det V_{ABC} - \ln \det V_C \geq 0. \quad (7.6)$$

Observe how the above combination of log-det entropies mimics that appearing in the celebrated SSA of the quantum von Neumann entropy S [293, 294, 159, 160], which is nowadays widely regarded as one of the cornerstones upon which quantum information theory is built [109]. In (7.6), the local reductions V_{AC} , V_{BC} and V_C are the principal submatrices of V_{ABC} corresponding to the components AC , BC and C , respectively:

$$V_{AC} = \begin{pmatrix} A & Y \\ Y^T & C \end{pmatrix}, \quad V_{BC} = \begin{pmatrix} B & Z \\ Z^T & C \end{pmatrix}, \quad V_C = C. \quad (7.7)$$

Let us observe that since (7.6) is balanced, the contribution of the inhomogeneous second terms of (7.3) cancel out.

Note. Throughout the present chapter, we denote with A , B and C both the three groups of components of the random vector X_{ABC} and – occasionally – the corresponding blocks in the covariance matrix (7.5). Since the former are always employed as subscripts, this should not lead to any ambiguity.

Inequality (7.6) was considered in [295] (see also [296, §4.5]), although it has been known long before under the name of Hadamard-Fisher inequality (see for instance [198, Exercise 14, p.485]).² From the point of view of matrix analysis, (7.6) lends itself to straightforward generalisations. In fact, inequalities of the same form have recently been investigated. In particular, the problem

²I thank Minghua Lin for bringing this to my attention.

of determining all the continuous functions $f : \mathbb{R}_+ \rightarrow \mathbb{R}$ such that for all block matrices $V_{ABC} \geq 0$,

$$\mathrm{Tr} f(V_{AC}) + \mathrm{Tr} f(V_{BC}) - \mathrm{Tr} f(V_{ABC}) - \mathrm{Tr} f(C) \geq 0, \quad (7.8)$$

was considered in full generality in [297], where a sufficient condition was found: (7.8) holds as soon as $-f'$ is matrix monotone. Later on, it was shown that this condition is also necessary [298]. By virtue of Löwner's theorems characterising matrix monotone functions [299, V], this yields an explicit characterisation of all the functions f obeying (7.8). Here we are mainly concerned with the particular choice $f(x) = \ln x$, that turns (7.8) into (7.6).

7.2.2 Log-det entropy

The differential Rényi- α entropy of a Gaussian random variable X with density $p_A(x)$, i.e. $H_\alpha(X) := \frac{1}{1-\alpha} \ln \int d^n x p_A(x)^\alpha$, is given by

$$h_\alpha(X) = \frac{1}{2} \ln \det A + \frac{n}{2} \left(\ln 2\pi + \frac{1}{\alpha-1} \ln \alpha \right),$$

showing that all the differential Rényi entropies of Gaussian random vectors are essentially equivalent to the differential Shannon entropy (7.3), up to a characteristic universal additive offset. In view of this and the above remarks, we are motivated, given a vector valued random variable X with covariance matrix V , to introduce the notation

$$M(X) := M(V) := \frac{1}{2} \ln \det V. \quad (7.9)$$

We will refer to the quantity defined in (7.9), already encountered in Subsection 6.4.2 as the **log-det entropy** of V . We borrow from Subsection 6.4.2 also the definition of *log-det mutual information*, defined for a bipartite covariance matrix $V_{AB} > 0$ by

$$I_M(A : B)_V := \frac{1}{2} \ln \frac{\det V_A \det V_B}{\det V_{AB}} = M(V_A) + M(V_B) - M(V_{AB}). \quad (7.10)$$

For tripartite covariance matrices $V_{ABC} > 0$ we can also construct a **log-det conditional mutual information**:

$$\begin{aligned} I_M(A : B|C)_V &:= \frac{1}{2} \ln \frac{\det V_{AC} \det V_{BC}}{\det V_C \det V_{ABC}} \\ &= M(V_{AC}) + M(V_{BC}) - M(V_{ABC}) - M(V_C). \end{aligned} \quad (7.11)$$

Every (balanced) entropic inequality thus yields a corresponding inequality for the log-determinants of positive block matrices [300]. Thanks to the work of Zhang and Yeung [301] and followers [302, 303], infinitely many independent such inequalities, so-called 'non-Shannon-type inequalities', are known by now.

The question of what are the precise constraints on the determinants of the 2^n principal submatrices of a positive matrix of size $n \times n$ has been raised much earlier, either directly in a matrix setting [304] or more recently in the guise of the balanced entropy inequalities of Gaussian random variables (both real valued or vector valued) [305, 306]. Remarkably, the latter papers show that while the entropy region of three Gaussian real random variables is convex but not a cone, the entropy region of three Gaussian random vectors is a convex cone and that the *linear* log-det inequalities for three Gaussian random variables (and equivalently Gaussian random vectors) are the same as the inequalities for the differential entropy of any three variables – which in turn coincide with the Shannon inequalities, cf. [292, 300]. It is conjectured that the same identity between Gaussian vector inequalities and general differential inequalities holds for any number parties.

Our concrete interest in (7.6) is partly motivated by its applications in quantum information theory with continuous variables [209], as first explored in [260, 265]. Every continuous variable quantum state ρ of n modes (Gaussian or non-Gaussian), subject to mild regularity conditions, has a corresponding $2n \times 2n$ quantum covariance matrix V of the phase space variables defined as in (5.28). Thus, quantities like log-det entropy and (conditional) mutual information can be associated with any sufficiently regular quantum state via its QCM. By construction, such entropic quantifiers captures the correlations encoded in the second moments of the state. These are exactly the correlations that one can access through Gaussian measurements, whose outputs are in fact Gaussian multivariate random variables (Subsection 5.2.3), as quantified by combinations of Shannon entropies.

As we saw in Subsection 6.4.2, for Gaussian states the log-det entropy reduces to the quantum Rényi-2 entropy, according to (6.55). We report that identity here since it is of central importance in the rest of the chapter:

$$S_2(\rho) := -\ln \text{Tr}[\rho^2] = \frac{1}{2} \ln \det V = M(V). \quad (7.12)$$

In particular, in the relevant case of tripartite quantum Gaussian states, the general inequality (7.6) for log-det entropy takes the form of a SSA inequality for the Rényi-2 entropy [260, 265, 234], holding in addition to the standard one for Rényi-1 entropy aka von Neumann entropy, which is valid for arbitrary (Gaussian or not) tripartite quantum states. Note that in general it is not advisable to form entropy expressions from Rényi entropies, since they do not obey any nontrivial constraints in a general multi-partite system [307]. In information theory, this is addressed by defining directly well-behaved notions of conditional Rényi entropy and Rényi mutual information [308]. Here, we evade those issues as we are restricting to Gaussian states. In fact, thanks to the correspondence discussed here, the following important result is readily established.

Lemma 7.1. *Rényi-2 entropies of Gaussian states satisfy all inequalities that are obeyed by Shannon entropies of Gaussian multivariate random variables.*

Moreover, every balanced inequality [300] obeyed by Rényi-2 entropies of Gaussian states can be obtained in this way.

Proof. The first statement is a consequence of the correspondence law (7.12). As for the second, it follows from the observation that every positive definite matrix $A > 0$ is a positive multiple of a QCM $A/\nu_{\min}(A)$, where $\nu_{\min}(A)$ is the minimum symplectic eigenvalue of A (5.43). If the inequality under examination is balanced, this scaling does not affect it. \square

Let us stress that the second claim of Lemma 7.1 does not hold for non-balanced inequalities. An example illustrating this fact is provided by the inequality (6.9), proven in [234], which can not hold for all positive definite V (that is, for all classical covariance matrices), as it can be easily seen by rescaling it via $V \mapsto kV$, for $k > 0$. Observe that the new matrix V becomes unphysical for sufficiently small k , as it violates the uncertainty principle (5.29).

7.3 SSA saturation and exact recovery

In this section we proceed to study the conditions under which (7.6) is saturated with equality. A necessary and sufficient condition was already found in [295] (for a comprehensive discussion, see [296]), but here we present new proofs as well as alternative formulations, which may provide new insights.

Let us start by fixing our notation concerning classical Gaussian channels, whose action can be described as follows. Denote the input random variable by X , and consider an independent Gaussian variable $Z \sim P_K$, where P_K is a normal distribution with covariance matrix K and zero mean. Then the output variable Y of the Gaussian channel N is given by $N(X) := Y := HX + Z$ for some matrix H of appropriate size. At the level of covariance matrices this translates to

$$N : V \mapsto V' = HVH^T + K, \quad (7.13)$$

where the only constraint to be obeyed is $K \geq 0$.

The following theorem gathers some notable facts concerning log-det conditional mutual information, and provides a neat example of how useful the interplay between matrix analysis and information theory with Gaussian random variables can be. In particular, the theory of Schur complements and matrix means we presented in Subsections 5.2.1 and 5.2.2, respectively, will play a relevant role in what follows. The results we prove now, in particular, are used extensively throughout the chapter, and some of them play an important role already in the proof of the main theorem of this section, Theorem 7.3.

Theorem 7.2. *For all positive, tripartite matrices $V = V_{ABC} > 0$, the following identities hold true:*

$$I_M(A : B|C)_V = I_M(A : B)_{V_{ABC}/V_C}, \quad (7.14)$$

$$I_M(A : B|C)_V = I_M(A : B)_{V^{-1}}. \quad (7.15)$$

Furthermore, for all pairs of positive definite matrices $V_{AB}, W_{AB} > 0$, the log-det mutual information is convex on the trace metric geodesic connecting them as in (5.19), i.e.

$$I_M(A : B)_{V\#_t W} \leq (1-t)I_M(A : B)_V + tI_M(A : B)_W \quad (7.16)$$

for all $t \in [0, 1]$.

Proof. Let us start by showing (7.14). Using repeatedly the determinant factorisation property (5.4), we find

$$\begin{aligned} I_M(A : B)_{V_{ABC}/V_C} &= \frac{1}{2} \ln \frac{\det(V_{AB}/V_C) \det(V_{BC}/V_C)}{\det(V_{ABC}/V_C)} \\ &= \frac{1}{2} \ln \frac{(\det V_{AB})(\det V_C)^{-1}(\det V_{BC})(\det V_C)^{-1}}{(\det V_{ABC})(\det V_C)^{-1}} \\ &= \frac{1}{2} \ln \frac{(\det V_{AB})(\det V_{BC})}{(\det V_{ABC})(\det V_C)} \\ &= I_M(A : B|C)_V. \end{aligned}$$

We now move to (7.15). The block inverse formula (5.6) give us

$$\begin{aligned} (V^{-1})_{AB} &= (V_{ABC}/V_C)^{-1}, \\ (V^{-1})_A &= (V_{ABC}/V_{BC})^{-1}, \\ (V^{-1})_B &= (V_{ABC}/V_{AC})^{-1}. \end{aligned}$$

Putting all together we find

$$\begin{aligned} I_M(A : B)_{V^{-1}} &= \frac{1}{2} \ln \frac{\det(V^{-1})_A \det(V^{-1})_B}{\det(V^{-1})_{AB}} \\ &= \frac{1}{2} \ln \frac{\det(V_{ABC}/V_{BC})^{-1} \det(V_{ABC}/V_{AC})^{-1}}{\det(V_{ABC}/V_C)^{-1}} \\ &= \frac{1}{2} \ln \frac{\det(V_{ABC}/V_C)}{\det(V_{ABC}/V_{BC}) \det(V_{ABC}/V_{AC})} \\ &= \frac{1}{2} \ln \frac{(\det V_{ABC})(\det V_C)^{-1}}{(\det V_{ABC})(\det V_{BC})^{-1}(\det V_{ABC})(\det V_{AC})^{-1}} \\ &= \frac{1}{2} \ln \frac{\det V_{AC} \det V_{BC}}{\det V_{ABC} \det V_C} \\ &= I_M(A : B|C)_V, \end{aligned}$$

which is what we wanted to show.

Finally, let us consider (7.16). We start by applying the monotonicity of the weighted geometric mean under positive maps (5.20), to the positive map

$\Phi(\cdot) := \Pi_A(\cdot)\Pi_A^T$, where Π_A is the projector onto the A components. This yields

$$(V\#_tW)_A = \Pi_A(V\#_tW)\Pi_A^T \leq V_A\#_tW_A.$$

Taking the determinant of both sides of this latter inequality and using for the right hand side the explicit formula (5.21) we obtain

$$\det(V\#_tW)_A \leq \det(V_A\#_tW_A) = (\det V_A)^{1-t}(\det W_A)^t.$$

Together with the analogous inequality for the B system, this gives

$$\begin{aligned} I_M(A : B)_{V\#_tW} &= \frac{1}{2} \ln \frac{(\det(V\#_tW)_A) (\det(V\#_tW)_B)}{\det(V\#_tW)_{AB}} \\ &\leq \frac{1}{2} \ln \frac{(\det V_A)^{1-t}(\det W_A)^t (\det V_B)^{1-t}(\det W_B)^t}{(\det V_{AB})^{1-t}(\det W_{AB})^t} \\ &= (1-t)I_M(A : B)_V + tI_M(A : B)_W, \end{aligned}$$

concluding the proof. □

Remark. Inequality (7.16) is especially notable because in general the log-det mutual information is not convex over the set of positive matrices. However, it is convex when restricted to geodesics in the trace metric, as we have just shown. Moreover, we note in passing that an analogous inequality to (7.16) does not seem to hold for the log-det conditional mutual information.

We are now ready to provide necessary and sufficient conditions for the saturation of the log-det SSA inequality (7.6).

Theorem 7.3. *For an arbitrary $V_{ABC} > 0$ written in block form as in (7.5), the following are equivalent:*

- (a) $I_M(A : B|C)_V = 0$, i.e. (7.6) is saturated;
- (b) $V_{ABC}/V_{BC} = V_{AC}/V_C$, i.e. (6.10) is saturated;
- (c) $(V^{-1})_{AB} = (V^{-1})_A \oplus (V^{-1})_B$;
- (d) $X = YC^{-1}Z^T$ (see [295] or [296, Theorem 4.49]);
- (e) there is a classical Gaussian channel $N_{C \rightarrow BC}$ such that $(I_A \oplus N_{C \rightarrow BC})(V_{AC}) = V_{ABC}$.

Proof. We divide the proof as follows.

- (a) \Leftrightarrow (b). Saturation of (7.6) and (6.10) are equivalent concepts, since it is very easy to verify that if $M \geq N > 0$ then $M = N$ if and only if $\det M = \det N$.

(a) \Leftrightarrow (c). It is well-known that $W_{AB} > 0$ satisfies $\det W_{AB} = \det W_A \det W_B$ iff its off-diagonal block is zero, i.e. iff $W_{AB} = W_A \oplus W_B$. For instance, this can be easily seen as a consequence of (5.4). Thanks to Theorem 7.2, identity (7.15), applying this observation with $W = V^{-1}$ yields the claim.

(b) \Rightarrow (d). This is known in linear algebra [295], but for the sake of completeness we provide a different proof that fits more with the spirit of the present work. Namely, we see that the variational representation of Schur complements (5.10) guarantees that (6.10) is saturated if and only if

$$\begin{aligned} V_{ABC} - (V_{AC}/V_C) \oplus 0_{BC} &= \begin{pmatrix} A - V_{AC}/V_C & X & Y \\ X^T & B & Z \\ Y^T & Z^T & C \end{pmatrix} \\ &= \begin{pmatrix} YC^{-1}Y^T & X & Y \\ X^T & B & Z \\ Y^T & Z^T & C \end{pmatrix} \\ &\geq 0. \end{aligned} \quad (7.17)$$

A necessary condition for (7.17) to hold is obtained by taking suitable matrix elements:

$$\begin{aligned} 0 &\leq \begin{pmatrix} v \\ w \\ -C^{-1}Y^T v \end{pmatrix}^T \begin{pmatrix} YC^{-1}Y^T & X & Y \\ X^T & B & Z \\ Y^T & Z^T & C \end{pmatrix} \begin{pmatrix} v \\ w \\ -C^{-1}Y^T v \end{pmatrix} \\ &= 2v^T(X - YC^{-1}Z^T)w + w^T Bw. \end{aligned}$$

This can only be true for all v and w if $X = YC^{-1}Z^T$. Moreover, this latter condition (together with the positivity of V_{ABC}) is enough to guarantee that (7.17) is satisfied. Indeed, we can write

$$\begin{aligned} \begin{pmatrix} YC^{-1}Y^T & YC^{-1}Z^T & Y \\ ZC^{-1}Y^T & B & Z \\ Y^T & Z^T & C \end{pmatrix} &= \begin{pmatrix} 0 & & \\ & B - ZC^{-1}Z^T & \\ & & 0 \end{pmatrix} + \begin{pmatrix} YC^{-\frac{1}{2}} \\ ZC^{-\frac{1}{2}} \\ C^{\frac{1}{2}} \end{pmatrix} \begin{pmatrix} YC^{-\frac{1}{2}} \\ ZC^{-\frac{1}{2}} \\ C^{\frac{1}{2}} \end{pmatrix}^T \\ &\geq 0, \end{aligned}$$

where $B - ZC^{-1}Z^T \geq 0$ follows from $\begin{pmatrix} B & Z \\ Z^T & C \end{pmatrix} \geq 0$.

(d) \Rightarrow (e). If in (7.13) we define

$$\begin{aligned} H &= H_R := \begin{pmatrix} \mathbf{1} & 0 \\ 0 & ZC^{-1} \\ 0 & \mathbf{1} \end{pmatrix}, \\ K &= K_R := \begin{pmatrix} 0 & & \\ & B - ZC^{-1}Z^T & \\ & & 0 \end{pmatrix}, \end{aligned} \quad (7.18)$$

we obtain straightforwardly

$$\begin{aligned} (I_A \oplus N_{C \rightarrow BC})(V_{AC}) &= H_R \begin{pmatrix} A & X \\ X^T & C \end{pmatrix} H_R^T + K_R \\ &= \begin{pmatrix} A & X & Y \\ X^T & B & Z \\ Y^T & Z^T & C \end{pmatrix} \\ &= V_{ABC}, \end{aligned}$$

provided that $X = YC^{-1}Z^T$. We will see in the next section that this map is nothing but a specialisation to the Gaussian case of a general construction known as ‘transpose channel’ or ‘Petz recovery map’.

(e) \Rightarrow (b). Since in the previous chapter we have shown that classical Gaussian channels acting on C never decrease the Schur complement V_{AC}/V_C (Theorem 6.2), it is clear that the equality in (6.10) is a necessary condition for the existence of a Gaussian recovery map $N_{C \rightarrow BC}$.

□

7.4 Improvements of log-det SSA

This section is devoted to presenting some improvements of the inequality (7.6). In Subsection 7.4.1 we employ techniques inspired from the recently developed *recoverability theory* to find an equivalent expression for $I_M(A : B|C)_M$. We leverage standard information theoretical results to obtain faithful lower bounds for this expression, leading to our first strengthening of the inequality. In Subsection 7.4.2 we take instead a matrix analytic point of view and to improve the saturation conditions of Theorem 7.3 more directly.

7.4.1 Gaussian recoverability

We start by discussing the role of some well-known remainder terms for inequalities of the form (7.6). These terms have been introduced recently in the context of sufficient statistics [309] and its approximate variants [164], or so-called ‘recoverability’. In [165], a form involving recovery maps was proposed for such a term in the fully quantum case (i.e., considering the SSA for von Neumann entropy) based on the fidelity of recovery, and subsequently strengthened to a bound involving the measured relative entropy [310]; in both cases the given bounds turn out to be operationally meaningful quantities [311]. The much simpler classical reasoning (with a better bound) was presented in [164]. We will translate these results into the Gaussian setting in order to find an explicit expression for a remainder term to be added to (7.6).

For classical probability distributions p and q over a discrete alphabet, in [164] the following inequality was shown, which improves on the monotonicity of the relative entropy under channels:

$$D(p||q) - D(Np||Nq) \geq D(p||RNp), \tag{7.19}$$

where $N = (N_{ji})$ is any stochastic map (channel) and the action of the transpose channel $R = R_{q,N}$ on an input distribution r is uniquely defined via the requirement that $N_{ji}q_i = R_{ij}(Nq)_j$ for all i and j . Explicitly,

$$(R_{q,N} r)_i := \sum_j \frac{q_i N_{ji}}{(Nq)_j} r_j. \quad (7.20)$$

Observe that $R_{q,N}$ is a bona fide channel, since

$$\sum_i (R_{q,N})_{ij} = \sum_i \frac{q_i N_{ji}}{(Nq)_j} = \frac{(Nq)_j}{(Nq)_j} = 1.$$

For obvious reasons, we will call the right hand side of (7.19) the **relative entropy of recovery**. The proof of (7.19) is a simple application of the concavity of the logarithm, and we reproduce it here for the benefit of the reader.

$$\begin{aligned} D(p \| R_{q,N} Np) &= \sum_i p_i \left(\ln p_i - \ln (R_{q,N} Np)_i \right) \\ &= \sum_i p_i \left(\ln p_i - \ln \sum_j \frac{q_i N_{ji}}{(Nq)_j} (Np)_j \right) \end{aligned} \quad (7.21)$$

$$\begin{aligned} &\leq \sum_i p_i \left(\ln p_i - \sum_j N_{ji} \ln \frac{q_i}{(Nq)_j} (Np)_j \right) \\ &= D(p \| q) - D(Np \| Nq). \end{aligned} \quad (7.22)$$

Although we wrote out the proof only for random variables taking values in a discrete alphabet, all of the above expressions make perfect sense also in more general cases, e.g. when i and j are multivariate real variables. If N is a classical Gaussian channel acting as in (7.13), it can easily be verified that the ‘transition probabilities’ $N(x, y)$ satisfying

$$(Np)(x) = \int dy N(x, y)p(y) \quad (7.23)$$

take the form

$$N(x, y) = \frac{e^{-\frac{1}{2}(x-Hy)^T K^{-1}(x-Hy)}}{\sqrt{(2\pi)^n \det K}}. \quad (7.24)$$

Following again [164], we observe that if the output of the random channel N is a deterministic function of the input, then (7.19) is always saturated with equality. This can be seen by noticing that in that case for all i there is only one index j such that $N_{ji} \neq 0$ (and so $N_{ji} = 1$). Therefore, the step from (7.21) to (7.22) is an equality. This remark is useful for a very important special case. Consider a triple of random variables XYZ distributed according to $p(xyz)$, a second probability distribution $q(xyz) = p(x)p(yz)$, and the channel N consisting of discarding Y . Obviously, in this case the output is

a deterministic function of the input. It is easily seen that the reconstructed global probability distribution $\tilde{p} = R_{q,N}Np$ is

$$\tilde{p}(xyz) = p(xz)p(y|z). \quad (7.25)$$

Then the saturation of (7.19) allows us to write

$$I(X : Y|Z) = D(p||q) - D(Np||Nq) = D(p||\tilde{p}). \quad (7.26)$$

From now on, we will consider the case in which N is a classical Gaussian channel transforming covariance matrices according to the rule (7.13). As can be easily verified, if also q is a multivariate Gaussian distribution, then $R_{q,N}$ becomes a classical Gaussian channel as well. We compute its action in the case we are mainly interested in, that is, when the left-hand side of (7.19) corresponds to the difference of the two sides of (7.6), and verify that it coincides with the recovery map introduced in Section 7.3 (via the general action (7.13) with the substitutions (7.18)).

Remark. Before we present our result, a quick side note. While here we are interested in the *classical* Gaussian Petz recovery map, a very recent study of mine [312] has investigated the quantum Petz map for Gaussian states, showing it to be a Gaussian channel. The results and techniques in [312] are however of somewhat different nature than those presented here, since they involve quantum states instead of classical random variables.

Proposition 7.4. *Let q be a tripartite Gaussian probability density with zero mean and covariance matrix*

$$V_A \oplus V_{BC} = \begin{pmatrix} A & 0 & 0 \\ 0 & B & Z \\ 0 & Z^T & C \end{pmatrix},$$

and let the channel N correspond to the action of discarding the B components, i.e. $H = \Pi_{AC} = \begin{pmatrix} \mathbf{1} & 0 & 0 \\ 0 & 0 & \mathbf{1} \end{pmatrix}$ and $K = 0$ in (7.13). Then, the action $C \rightarrow BC$ of the Petz recovery map (7.20) on Gaussian variables with zero mean can be written at the level of covariance matrices as (7.13), where H_R and K_R are given by (7.18).

Proof. The Petz recovery map (7.20) is a composition of three operations: first the pointwise division by a Gaussian distribution, then the transpose of a deterministic channel, and eventually another pointwise Gaussian multiplication. It should be obvious from (7.1) that a pointwise multiplication by a Gaussian distribution with covariance matrix A is a Gaussian (non-deterministic) channel that leaves the mean vector invariant and acts on covariance matrices as $V \mapsto V' = (V^{-1} + A^{-1})^{-1}$. Furthermore, it can be proven that the transpose N^T of the channel N in (7.13) sends Gaussian variables with zero mean to other Gaussian variables with zero mean, while on the inverses of the covariance matrices it acts as

$$N^T : V^{-1} \mapsto (V')^{-1} = H^T(V + K)^{-1}H. \quad (7.27)$$

A way to prove the above equation is by using (7.24) to compute directly the action of N^T on a Gaussian input distribution.

After the preceding discussion, it should be clear that under our hypotheses the action of the Petz recovery map can be written as

$$\sigma_{AC} \mapsto \sigma'_{ABC} = \left(V_A^{-1} \oplus V_{BC}^{-1} + (\sigma_{AC}^{-1} - V_A^{-1} \oplus V_C^{-1}) \oplus 0_B \right)^{-1}. \quad (7.28)$$

The Woodbury matrix identity (see [313], or [232, Eq. (6.0.10)]),

$$(S + UTV)^{-1} = S^{-1} - S^{-1}U(VS^{-1}U + T^{-1})^{-1}VS^{-1}, \quad (7.29)$$

can be used to bring (7.28) into the canonical form (7.13):

$$\begin{aligned} \sigma'_{ABC} &= (V_A^{-1} \oplus V_{BC}^{-1} + (\sigma_{AC}^{-1} - V_A^{-1} \oplus V_C^{-1}) \oplus 0_B)^{-1} \\ &= (V_A^{-1} \oplus V_{BC}^{-1} + \Pi_{AC}^T(\sigma_{AC}^{-1} - V_A^{-1} \oplus V_C^{-1})\Pi_{AC})^{-1} \\ &= V_A \oplus V_{BC} - (V_A \oplus V_{BC})\Pi_{AC}^T \\ &\quad \cdot \left((\sigma_{AC}^{-1} - V_A^{-1} \oplus V_C^{-1})^{-1} + \Pi_{AC}(V_A \oplus V_{BC})\Pi_{AC}^T \right)^{-1} \Pi_{AC}(V_A \oplus V_{BC}) \\ &= V_A \oplus V_{BC} - (V_A \oplus V_{BC})\Pi_{AC}^T \\ &\quad \cdot \left(-V_A \oplus V_C - (V_A \oplus V_C)(\sigma_{AC} - V_A \oplus V_C)^{-1}(V_A \oplus V_C) + V_A \oplus V_C \right)^{-1} \\ &\quad \cdot \Pi_{AC}(V_A \oplus V_{BC}) \\ &= V_A \oplus V_{BC} \\ &\quad + (V_A \oplus V_{BC})\Pi_{AC}^T(V_A^{-1} \oplus V_C^{-1})(\sigma_{AC} - V_A \oplus V_C)(V_A^{-1} \oplus V_C^{-1}) \\ &\quad \cdot \Pi_{AC}(V_A \oplus V_{BC}) \\ &= H_R \sigma_{AC} H_R^T + K_R, \end{aligned}$$

where $H_R = (V_A \oplus V_{BC})\Pi_{AC}^T(V_A^{-1} \oplus V_C^{-1})$ and K_R are defined according to (7.18). This concludes the proof. \square

We are ready to employ the classical theory of recoverability in order to find the expression of the relative entropy of recovery in the Gaussian case.

Proposition 7.5. *For all tripartite covariance matrices $V_{ABC} > 0$ written in block form as in (7.5), we have*

$$I_M(A : B|C)_V = \frac{1}{2} \ln \frac{\det V_{AC} \det V_{BC}}{\det V_{ABC} \det V_C} = D(V_{ABC} \| \tilde{V}_{ABC}), \quad (7.30)$$

where

$$\tilde{V}_{ABC} := \begin{pmatrix} A & YC^{-1}Z^T & Y \\ ZC^{-1}Y^T & B & Z \\ Y^T & Z^T & C \end{pmatrix} \quad (7.31)$$

and the relative entropy function $D(\|\cdot)$ is given by (7.4).

Proof. This is just an instance of (7.26) applied to the continuous Gaussian variable (X_A, X_B, X_C) . \square

The identity (7.30) is useful in deducing new constraints that will be much less obvious coming from a purely matrix analysis perspective. For instance, it is well known that $D(p\|q) \geq -\ln \mathcal{F}^2(p, q)$ (see e.g. [314, 315]), where the fidelity is given by $\mathcal{F}(p, q) = \sum_i \sqrt{p_i q_i}$ in the discrete case. In case of Gaussian variables with the same mean, it holds

$$\mathcal{F}^2(p_A, p_B) = \frac{\det(A!B)}{\sqrt{\det A \det B}}, \quad (7.32)$$

where $(A!B) := 2(A^{-1} + B^{-1})^{-1}$ is the *harmonic mean* of A and B . Inserting this standard lower bound into (7.30) we obtain

$$\frac{\det V_{AC} \det V_{BC}}{\det V_{ABC} \det V_C} \geq \frac{\det V_{ABC} \det \tilde{V}_{ABC}}{\left(\det(V_{ABC}! \tilde{V}_{ABC})\right)^2}, \quad (7.33)$$

leading to

$$I_M(A : B|C)_V \geq \frac{1}{2} \ln \frac{\det V_{ABC} \det \tilde{V}_{ABC}}{\left(\det(V_{ABC}! \tilde{V}_{ABC})\right)^2}. \quad (7.34)$$

Using furthermore

$$\begin{aligned} \det \tilde{V}_{ABC} &= \det \tilde{V}_{BC} \det(\tilde{V}_{ABC}/\tilde{V}_{BC}) \\ &= \det V_{BC} \det(\tilde{V}_{AC}/\tilde{V}_C) \\ &= \det V_{BC} \det(V_{AC}/V_C), \end{aligned}$$

we also arrive at the inequality

$$\det V_{ABC} \leq \det(V_{ABC}! \tilde{V}_{ABC}). \quad (7.35)$$

To illustrate the power of this relation, we note that inserting the harmonic-geometric mean inequality for matrices (5.22)

$$A!B \leq A\#B$$

yields again SSA (7.6) in the form $\det \tilde{V}_{ABC} \geq \det V_{ABC}$.

7.4.2 A lower bound on $I_M(A : B|C)_V$

Throughout this subsection, we explore some alternative ways of strengthening Theorem 7.3, i.e. of finding suitable lower bounds on the log-det conditional mutual information $I_M(A : B|C)_V$. The expression we are seeking should have two main features: (i) it should be easily computable in terms of the blocks of V_{ABC} ; and (ii) the explicit saturation condition in Theorem 7.3(d) should

be easily readable from it. This latter requirement can be accommodated, for example, if the lower bound involves some kind of distance between the off-diagonal block X and its ‘saturation value’ $YC^{-1}Z^T$. We start with a preliminary result.

Proposition 7.6. *For all matrices*

$$V_{AB} = \begin{pmatrix} A & X \\ X^T & B \end{pmatrix} \geq 0,$$

we have

$$I_M(A : B)_V \geq \frac{1}{2} \text{Tr} [A^{-1}XB^{-1}X^T] = \frac{1}{2} \|A^{-1/2}XB^{-1/2}\|_2^2. \quad (7.36)$$

Proof. Using, in this order, the standard factorisation of the determinant in terms of the Schur complement, the identity $\ln \det V = \text{Tr} \ln V$ (valid when $V > 0$), and the inequality $\ln(\mathbf{1} + \Delta) \leq \Delta$ (for Hermitian $\Delta > -\mathbf{1}$), we find

$$\begin{aligned} I_M(A : B)_V &= \frac{1}{2} \ln \frac{\det V_A \det V_B}{\det V_{AB}} \\ &= -\frac{1}{2} \ln \det V_A^{-1/2} (V_{AB}/V_B) V_A^{-1/2} \\ &= -\frac{1}{2} \ln \det (\mathbf{1} - A^{-1/2}XB^{-1}X^T A^{-1/2}) \\ &= -\frac{1}{2} \text{Tr} \ln (\mathbf{1} - A^{-1/2}XB^{-1}X^T A^{-1/2}) \\ &\geq \frac{1}{2} \text{Tr} A^{-1/2}XB^{-1}X^T A^{-1/2} \\ &= \frac{1}{2} \text{Tr} A^{-1}XB^{-1}X^T \\ &= \frac{1}{2} \|A^{-1/2}XB^{-1/2}\|_2^2, \end{aligned}$$

and we are done. \square

Theorem 7.7. *For all $V_{ABC} > 0$ written in block form as in (7.5), we have the following chain of inequalities:*

$$\begin{aligned} I_M(A : B|C)_V &\geq \frac{1}{2} \text{Tr} \left[(V_{AC}/V_C)^{-1} (X - YC^{-1}Z^T) (V_{BC}/V_C)^{-1} (X - YC^{-1}Z^T)^T \right] \\ &\geq \frac{1}{2} \text{Tr} \left[A^{-1} (X - YC^{-1}Z^T) B^{-1} (X - YC^{-1}Z^T)^T \right] \\ &= \frac{1}{2} \left\| A^{-1/2} (X - YC^{-1}Z^T) B^{-1/2} \right\|_2^2. \end{aligned}$$

Proof. We want to use the identity (7.15) to lower bound $I_M(A : B|C)_V$ via Proposition 7.6. In order to do so, we need to write out the A - B off-diagonal block of the inverse $(V_{ABC})^{-1}$. Introducing the projectors onto the A and B components, denoted by Π_A and Π_B respectively, this amounts to saying that we are seeking an explicit expression for $\Pi_A(V_{ABC})^{-1}\Pi_B^T$. Remember that for an arbitrary bipartite block matrix W_{12} the block-inversion formula (5.6) gives

$$\Pi_1(W_{12})^{-1}\Pi_1^T = (W_{12}/W_2)^{-1}, \quad (7.37)$$

$$\Pi_1(W_{12})^{-1}\Pi_2^T = -W_1^{-1}(\Pi_1 W_{12} \Pi_2^T)(W_{12}/W_1)^{-1}. \quad (7.38)$$

This allows us to write

$$\begin{aligned} & \Pi_A(V_{ABC})^{-1}\Pi_B^T \\ &= \Pi_A \Pi_{AB} (V_{ABC})^{-1} \Pi_{AB}^T \Pi_B^T \\ &= \Pi_A (V_{ABC}/V_C)^{-1} \Pi_B^T \\ &= -(V_{AC}/V_C)^{-1} (\Pi_A V_{ABC}/V_C \Pi_B^T) ((V_{ABC}/V_C)/(V_{AC}/V_C))^{-1} \\ &= -(V_{AC}/V_C)^{-1} (X - YC^{-1}Z^T)(V_{ABC}/V_{AC})^{-1}. \end{aligned}$$

By exchanging A and B in this latter expression and subsequently taking the transpose, we arrive also at

$$\Pi_A(V_{ABC})^{-1}\Pi_B^T = -(V_{ABC}/V_{BC})^{-1} (X - YC^{-1}Z^T) \cdot (V_{BC}/V_C)^{-1}.$$

Now we are ready to invoke Proposition 7.6 to write

$$\begin{aligned} & I_M(A : B|C)_V \\ &= I_M(A : B)_{V^{-1}} \\ &\geq \frac{1}{2} \text{Tr} \left[(V^{-1})_A^{-1} (\Pi_A V^{-1} \Pi_B^T) (V^{-1})_B^{-1} (\Pi_B^T V^{-1} \Pi_A) \right] \\ &= \frac{1}{2} \text{Tr} \left[(V_{ABC}/V_{BC}) ((V_{ABC}/V_{BC})^{-1} (X - YC^{-1}Z^T) (V_{BC}/V_C)^{-1}) \right. \\ &\quad \left. \cdot (V_{ABC}/V_{AC}) ((V_{AC}/V_C)^{-1} (X - YC^{-1}Z^T) (V_{ABC}/V_{AC})^{-1})^T \right] \\ &= \frac{1}{2} \text{Tr} \left[(V_{AC}/V_C)^{-1} (X - YC^{-1}Z^T) (V_{BC}/V_C)^{-1} (X - YC^{-1}Z^T)^T \right]. \end{aligned}$$

Since on the one hand $V_{AC}/V_C \leq V_A = A$, and on the other hand the expression $\text{Tr} RKS K^T$ is clearly monotonic in $R, S \geq 0$, we finally obtain

$$\begin{aligned} I_M(A : B|C)_V &\geq \frac{1}{2} \text{Tr} \left[A^{-1} (X - YC^{-1}Z^T) B^{-1} (X - YC^{-1}Z^T)^T \right] \\ &= \frac{1}{2} \|A^{-1/2} (X - YC^{-1}Z^T) B^{-1/2}\|_2^2. \end{aligned}$$

and we are done. □

It can easily be seen that the above result satisfies the requirements stated in the beginning of the section. In fact, the above lower bound for $I_M(A : B|C)_V$ is (i) easily computable in terms of the blocks of V_{ABC} , and (ii) zero iff the condition in Theorem 7.3(d) is met.

7.5 SSA for quantum covariance matrices and Rényi-2 Gaussian squashed entanglement

In this final section we show how to apply results on log-det conditional mutual information to infer properties of Gaussian states in quantum optics. First, in Subsection 7.5.1 we have a closer look at purifications of Gaussian states, stating some elementary results that complement those presented in Subsection 5.2.3. The, Subsection 7.5.2 presents a lower bound on the log-det conditional mutual information that holds specifically for quantum covariance matrices (Theorem 7.11). Using this result, in the following Subsection 7.5.3 we define a Rényi-2 Gaussian version of the squashed entanglement we already encountered in Subsection 3.3.3. Our main result there is a proof of the remarkable fact that in the Gaussian setting this quantity coincides with the Rényi-2 Gaussian entanglement of formation (Theorem 7.14), which in particular provides an efficient way to compute it.

7.5.1 More about Gaussian states

We start by proving some results on Gaussian purifications that will be instrumental in the rest of the section. Remember that the commutation relations of the canonical operators x_i, p_j are encoded in the $2n \times 2n$ matrix Ω defined in (5.24). In the rest of the chapter, the antisymmetric, non-degenerate quadratic form identified by Ω will be called **standard symplectic product**, and the linear space \mathbb{R}^{2n} endowed with this product will be referred to as a **symplectic space**. In what follows, the symplectic space associated with a quantum optical system A will be denoted with Σ_A .³ For an introduction to symplectic geometry, we refer the reader to the excellent monograph [275].

When the system under examination is made of several parties (each comprising a certain number of modes), the global QCM will have a block structure as in (7.5). The symplectic form in this case is simply given by the direct sum of the local symplectic forms (5.32). This can be rephrased by saying that the symplectic space associated with the system AB is the direct sum of the symplectic spaces associated with A and B , in formula $\Sigma_{AB} = \Sigma_A \oplus \Sigma_B$ [275, Eq. (1.4)]. Conversely, discarding a subsystem corresponds to performing an orthogonal projection of the QCM onto the corresponding symplectic subspace [275, §1.2.1], in formula $V_A = \Pi_A V_{AB} \Pi_A^T$.

Pure Gaussian states enjoy many useful properties that we will exploit multiple times throughout this section. To explore them, it is essential to

³Not to be confused with the matrix that changes the signs of the momenta (5.41), denoted with the same symbol.

make a clever use of the complementarity between the two pictures, at the Hilbert space level on the one hand and at the QCM level on the other hand. Let us illustrate this point by presenting three lemmas.

Lemma 7.8. *Let V_{AB} be a QCM of bipartite system AB . Denote by $V_A = \Pi_A V_{AB} \Pi_A^T$ the reduced QCM corresponding to the subsystem A , and analogously for V_B . If V_A is pure, then $V_{AB} = V_A \oplus V_B$.*

Proof. The statement becomes obvious at the Hilbert space level. In fact, the reduced state on A of a bipartite state ρ_{AB} is given by $\rho_A = \text{Tr}_B \rho_{AB}$, where Tr_B denotes partial trace [109]. Evaluating the ranks of both sides of this equation shows that if ρ_A is pure then the global state must be factorised, i.e. the QCM must have vanishing off-diagonal blocks. \square

Extending the system as to include auxiliary degrees of freedom is a standard technique in quantum information, popularly referred to as going to the ‘Church of the larger Hilbert space’⁴. Such a technique can be most notably employed in order to *purify* the system under examination, as detailed in the following lemma [220].

Lemma 7.9. *For all QCMs V_A pertaining to a system A there exists an extension AE of A and a pure QCM γ_{AE} such that $\Pi_A \gamma_{AE} \Pi_A^T = V_A$, where Π_A is the projector onto the symplectic subspace $\Sigma_A \subset \Sigma_{AE}$.*

Proof. See [220, §III.D]. \square

Finally, let us present here another useful observation.

Lemma 7.10. *For all QCMs $V_A \geq i\Omega_A$ of a system A , there is a decomposition $\Sigma_A = \Sigma_{A_1} \oplus \Sigma_{A_2}$ of the global symplectic space into a direct sum of two symplectic subspaces such that*

$$V_A = V_{A_1} \oplus \eta_{A_2}, \tag{7.39}$$

where $V_{A_1} > i\Omega_{A_1}$ and η_{A_2} is a pure QCM. Furthermore, for every purification γ_{AE} of V_A (see Lemma 7.9) there is a symplectic decomposition of E as $\Sigma_E = \Sigma_{E_1} \oplus \Sigma_{E_2}$ such that: (a) $\gamma_{AE} = \gamma_{A_1 E_1} \oplus \eta_{A_2} \oplus \tau_{E_2}$, with η_{A_2}, τ_{E_2} pure QCMs; (b) $n_{A_1} = n_{E_1}$; and (c) $\gamma_{E_1} > i\Omega_{E_1}$.

Proof. The first claim is a direct consequence of Williamson’s decomposition (5.43). The subspace Σ_{A_2} corresponds to those symplectic eigenvalues of V_A that are equal to 1.

Now, let us prove the second claim. Consider an arbitrary pure QCM γ_{AE} that satisfies $\gamma_A = V_A = V_{A_1} \oplus \eta_{A_2}$. Since in particular $\gamma_{A_2} = \eta_{A_2}$, we can apply Lemma 7.8 and conclude that $\gamma_{AE} = \gamma_{A_1 E} \oplus \eta_{A_2}$. Claim (a) then tells us that $\gamma_E = \gamma_{E_1} \oplus \tau_{E_2}$, with $\gamma_{E_1} > i\Omega_{E_1}$ and τ_{E_2} pure. Again, Lemma 7.8

⁴This expression was coined by John A. Smolin, see also [316].

yields $\gamma_{AE} = \gamma_{A_1 E_1} \oplus \eta_{A_2} \oplus \tau_{E_2}$, corresponding to statement (b). Hence, we have only to show that $n_{A_1} = n_{E_1}$. In order to show this, let us write

$$\gamma_{A_1 E_1} = \begin{pmatrix} V_{A_1} & L \\ L^T & \gamma_{E_1} \end{pmatrix}.$$

We can invoke Lemma 6.3 to deduce the identity $V_{A_1} - L\gamma_{E_1}^{-1}L^T = \Omega V_{A_1}^{-1}\Omega^T$, that is, $L\gamma_{E_1}^{-1}L^T = V_{A_1} - \Omega V_{A_1}^{-1}\Omega^T$. Since the right hand side has maximum rank $2n_{A_1}$ thanks to the strict inequality $V_{A_1} > i\Omega$ (last claim of Lemma 5.10), we conclude that $2n_{E_1} \geq \text{rk}(L\gamma_{E_1}^{-1}L^T) = 2n_{A_1}$, and hence $n_{E_1} \geq n_{A_1}$. But the same reasoning can be applied with A_1 and E_1 exchanged, thus giving $n_{A_1} \geq n_{E_1}$, which concludes the proof. \square

7.5.2 SSA for quantum covariance matrices

We are now ready to apply our results to strengthen the SSA inequality (7.6) in the quantum case. This subsection is thus devoted to finding a sensible lower bound on the log-det conditional mutual information for all QCMs. This bound will be given by the *Rényi-2 Gaussian entanglement of formation*, already encountered in Subsection 6.4.2 (see (6.58) and (6.60)). We report its expression here:

$$E_{F,2}^G(A : B)_V = \inf \frac{1}{2} \log \det \gamma_A \quad (7.40)$$

s.t. γ_{AB} pure QCM and $\gamma_{AB} \leq V_{AB}$.

As we saw, the above equation can be rephrased in an equivalent form using the identity between local entropies of a pure state, which in our case reads $M(\gamma_A) = M(\gamma_B) = \frac{1}{2}I_M(A : B)_\gamma$ for all pure QCMs γ_{AB} . We obtain

$$E_{F,2}^G(A : B)_V = \inf \frac{1}{2} I_M(A : B)_\gamma \quad (7.41)$$

s.t. γ_{AB} pure QCM and $\gamma_{AB} \leq V_{AB}$.

Remember that in the preceding Chapter 6 we used (7.41), together with the notable inequality $E_{F,2}^G(A : B)_V \leq \frac{1}{2}I_M(A : B)_V$ (6.61), to show that the Rényi-2 Gaussian entanglement of formation is monogamous (6.62). We are now in position to apply some of the tools we have been developing so far to prove a generalisation of said inequality (6.61) that is of interest to us since it constitutes also a strengthening of (7.6).

Theorem 7.11. *For all tripartite QCMs $V_{ABC} \geq i\Omega_{ABC}$, it holds that*

$$\frac{1}{2}I_M(A : B|C)_V \geq E_{F,2}^G(A : B)_V. \quad (7.42)$$

Proof. We employ a similar trick to the one used in the proof of Theorem 6.14: for any QCM V_{ABC} , using the notation of (5.51), define

$$\gamma_{AB} := \gamma_{V_{ABC}/V_C}^\#.$$

By Lemma 5.10, we see that γ_{AB} is a pure QCM. Now we proceed to show that $\gamma_{AB} \leq V_{AB}$. On the one hand, the very definition of Schur complement implies that $V_{ABC}/V_C \leq V_{AB}$, while on the other hand (6.14) gives us

$$V_{ABC}/V_C \geq \Omega_{AB}^T V_{AB}^{-1} \Omega_{AB},$$

which upon matrix inversion becomes

$$\Omega_{AB}(V_{ABC}/V_C)^{-1} \Omega_{AB}^T \leq V_{AB}.$$

Since the geometric mean is well-known to be monotonic (see Subsection 5.2.2 as well as [242]), we obtain immediately

$$\begin{aligned} \gamma_{AB} &= (V_{ABC}/V_C) \# (\Omega_{AB}(V_{ABC}/V_C) \Omega_{AB}^T) \\ &\leq V_{AB} \# V_{AB} \\ &= V_{AB}. \end{aligned}$$

This shows that γ_{AB} can be used as an ansatz in (7.41). We can then write

$$\begin{aligned} E_{F,2}^G(A : B)_V &\leq \frac{1}{2} I_M(A : B)_\gamma \\ &= \frac{1}{2} I_M(A : B)_{(V_{ABC}/V_C) \# (\Omega_{AB}(V_{ABC}/V_C)^{-1} \Omega_{AB}^T)} \\ &\stackrel{(1)}{\leq} \frac{1}{4} I_M(A : B)_{V_{ABC}/V_C} + \frac{1}{4} I_M(A : B)_{\Omega_{AB}(V_{ABC}/V_C)^{-1} \Omega_{AB}^T} \\ &\stackrel{(2)}{=} \frac{1}{4} I_M(A : B)_{V_{ABC}/V_C} + \frac{1}{4} I_M(A : B)_{(V_{ABC}/V_C)^{-1}} \\ &\stackrel{(3)}{=} \frac{1}{4} I_M(A : B|C)_V + \frac{1}{4} I_M(A : B|C)_V \\ &= \frac{1}{2} I_M(A : B|C)_V, \end{aligned}$$

where we employed, in order: (1) the convexity of log-det mutual information on the trace metric geodesics (7.16), (2) the obvious fact that since $\Omega_{AB} = \Omega_A \oplus \Omega_B$ is a local symplectic operation, the equality $I_M(A : B)_{\Omega W \Omega^T} = I_M(A : B)_W$ holds true; and (3) the identity (7.14) for the first term and (7.15) followed again by (7.14) for the second. \square

7.5.3 Rényi-2 Gaussian squashed entanglement

As we mentioned in Subsection 3.3.3, in finite-dimensional quantum systems the positivity of conditional mutual information allows to construct a powerful entanglement measure called *squashed entanglement* (3.28) (see also the original paper [11]). Let us report the definition here for convenience. For a bipartite state ρ_{AB} , one constructs

$$E_{\text{sq}}(A : B)_\rho := \inf_{\rho_{ABC}} \frac{1}{2} I(A : B|C)_\rho,$$

where the infimum ranges over all possible ancillary quantum systems C and over all the possible states ρ_{ABC} that have ρ_{AB} as their marginal. We are now in position to discuss a similar quantity tailored to Gaussian states. First, we can restrict the infimum by considering only Gaussian extensions, which corresponds to the step leading from (6.56) to (6.57). Secondly, as it was done to arrive at (6.58) (here reported as (6.58)), we can substitute von Neumann entropies with Rényi-2 entropies. The result is

$$E_{\text{sq},2}^G(A : B)_V := \inf_{V_{ABC}} \frac{1}{2} I_M(A : B|C)_V, \quad (7.43)$$

where the infimum is on all extended QCMs V_{ABC} satisfying the condition $\Pi_{AB} V_{ABC} \Pi_{AB}^T = V_{AB}$ on the AB marginal (and (5.29)). We name the quantity in (7.43) **Rényi-2 Gaussian squashed entanglement**, stressing that it is a quantifier specifically tailored to Gaussian states and different from the Rényi squashed entanglement defined in [317] for general states, where an alternative expression for the conditional Rényi- α mutual information is adopted instead.

Despite the complicated appearance of the expression (7.43), it turns out that *the Rényi-2 Gaussian squashed entanglement coincides with the Rényi-2 Gaussian entanglement of formation for all bipartite QCMs*. This unexpected fact shows once more that Rényi-2 quantifiers are particularly well behaved when employed to analyse Gaussian states, while at the same time it provides us with a novel, alternative expression of $E_{F,2}^G$ that can be used to understand its basic properties in a different, and sometimes more intuitive, way. Before stating the main result of this subsection, we need some preliminary results.

Lemma 7.12. *Let γ_{AB} be a pure QCM of a bipartite system AB such that $n_A = n_B = n$ and $\gamma_A > i\Omega_A$. Then*

$$(\gamma_{AB} + i\Omega_{AB}) / (\gamma_A + i\Omega_A) = 0_B.$$

Proof. From Williamson's decomposition (5.43), we see that whenever γ_{AB} is pure one has $\text{rk}(\gamma_{AB} + i\Omega_{AB}) = n_A + n_B = 2n$ (i.e. half the maximum). Since already $\text{rk}(\gamma_A + i\Omega_A) = 2n$, the additivity of ranks under Schur complements (5.5) tells us that $\text{rk}((\gamma_{AB} + i\Omega_{AB}) / (\gamma_A + i\Omega_A)) = 0$, concluding the proof. \square

Proposition 7.13. *Let V_{AB} be a QCM of a bipartite system, and let γ_{ABC} be a fixed purification of V_{AB} (see Lemma 7.9). Then, for all pure QCMs $\tau_{AB} \leq V_{AB}$ there exists a one-parameter family of pure QCMs $\sigma_C(t)$ (where $0 < t \leq 1$) on C such that*

$$\gamma'_{AB}(t) := (\gamma_{ABC} + 0_{AB} \oplus \sigma_C(t)) / (\gamma_C + \sigma_C(t)) \quad (7.44)$$

is a pure QCM for all $t > 0$, and $\lim_{t \rightarrow 0^+} \gamma'_{AB}(t) = \tau_{AB}$. Equivalently, there is a sequence of Gaussian measurements on C , identified by pure seeds $\sigma_C(t)$, such that the QCM of the post-measurement state on AB (5.39) is pure and tends to τ_{AB} .

Proof. Let us start by applying Lemma 7.10 to decompose the symplectic space of AB as $\Sigma_{AB} = \Sigma_R \oplus \Sigma_S$ in such a way that $V_{AB} = V_R \oplus \eta_S$, where $V_R > i\Omega_R$ and η_S is a pure QCM. According to Lemma 7.10, the purification γ_{ABC} can be taken to be of the form $\gamma_{ABC} = \gamma_{RC_1} \oplus \eta_S \oplus \delta_{C_2}$, with $\gamma_{C_1} > i\Omega_{C_1}$, $n_{C_1} = n_R$, and δ_{C_2} pure. If $\tau \leq V$ is a pure QCM, a projection onto Σ_S reveals that $\tau_S = \Pi_S \tau \Pi_S^T \leq \eta_S$. Since τ_S must be a legitimate QCM, and pure states are minimal within the set of QCMs, we deduce that $\tau_S = \eta_S$. Then, an application of Lemma 7.8 allows us to conclude that $\tau = \tau_R \oplus \eta_S$, and accordingly $\tau_R \leq V_R$.

We claim that for all pure $\tau_R < V_R$ there is a pure QCM σ_{C_1} such that

$$(\gamma_{RC_1} + 0_R \oplus \sigma_{C_1}) / (\gamma_{C_1} + \sigma_{C_1}) = \tau_R. \quad (7.45)$$

Constructing the extension $\sigma_C := \sigma_{C_1} \oplus \tilde{\sigma}_{C_2}$, where $\tilde{\sigma}_{C_2}$ is an arbitrary pure QCM, we see that (7.45) can be rewritten as

$$(\gamma_{ABC} + 0_{AB} \oplus \sigma_C) / (\gamma_C + \sigma_C) = \tau_R \oplus \eta_S. \quad (7.46)$$

In fact, adding the ancillary system C_2 does not produce any effect on the Schur complement, since there are no off-diagonal block linking C_2 with any other subsystem. Analogously, the S component of the AB system can be brought out of the Schur complement because it is in direct sum with the rest.

In light of (7.46), we know that once (7.45) has been established, in (7.44) we can achieve all QCMs γ' that can be written as $\tau_R \oplus \eta_S$, with $\tau_R < V_R$. It is not difficult to see that this would allow us to conclude. Before proving (7.45), let us see why. The main point here is that every pure QCM $\tau_R \leq V_R$ can be thought of as the limit of a sequence of pure QCMs $\tau_R(t) < V_R$. An explicit formula for such a sequence reads $\tau_R(t) = \tau_R \#_t \gamma_{V_R}^\#$, where $\gamma_{V_R}^\#$ is the pure QCM defined in (5.51), and $\#_t$ denotes the weighted geometric mean (5.19). Observe that: (1) $\tau_R(t)$ is a QCM since it is known that the set of QCMs is closed under weighted geometric mean [252, Corollary 8]; (2) $\tau_R(t)$ is in fact a pure QCM, because according to (5.21) its determinant satisfies $\det \tau_R(t) = (\det \tau_R)^{1-t} (\det \gamma_{V_R}^\#)^t = 1$; (3) $\lim_{t \rightarrow 0^+} \tau_R(t) = \tau_R$ as can be seen easily from (5.19); and (4) $\tau_R(t) < V_R$ for all $t > 0$. This latter fact can be justified as follows. Since $V_R > i\Omega_R$, from the last claim of Lemma 5.10 we deduce $\gamma_{V_R}^\# < V_R$. Taking into account that $\tau_R \leq V_R$, the claim follows from the strict monotonicity of the weighted geometric mean, in turn an easy consequence of (5.19).

Now, let us prove (7.45). We start by writing

$$\gamma_{RC_1} = \begin{pmatrix} V_R & L \\ L^T & \gamma_{C_1} \end{pmatrix},$$

where $V_R > i\Omega_R$, $\gamma_{C_1} > i\Omega_{C_1}$, and the off-diagonal block L is square. As a matter of fact, more is true, namely that L is also invertible. The simplest way to see this involves two ingredients: (1) the identity $\Omega V_R^{-1} \Omega^T = \gamma_{RC_1} / \gamma_{C_1} = V_R - L \gamma_{C_1}^{-1} L^T$, easily seen to be a special case of [235, Eq. (8)]; and (2) the fact

that $V_R > \Omega V_R^{-1} \Omega^T$ because of Lemma 5.10. Combining these two ingredients we see that

$$V_R > \Omega V_R^{-1} \Omega^T = V_R - L \gamma_{C_1}^{-1} L^T,$$

which implies $L \gamma_{C_1}^{-1} L^T > 0$ and in turn the invertibility of L . Now, for a pure QCM $\tau_R < V_R$, take $\sigma_{C_1} = L^T (V_R - \tau_R)^{-1} L - \gamma_{C_1}$. On the one hand,

$$(\gamma_{RC_1} + 0_R \oplus \sigma_{C_1}) / (\gamma_{C_1} + \sigma_{C_1}) = V_R - L (\gamma_{C_1} + \sigma_{C_1})^{-1} L^T = \tau_R$$

by construction. On the other hand, write

$$\begin{aligned} \sigma_{C_1} - i\Omega_{C_1} &= L^T (V_R - \tau_R)^{-1} L - (\gamma_{C_1} + i\Omega_{C_1}) \\ &= L^T (V_R - \tau_R)^{-1} L - L^T (V_R + i\Omega_R)^{-1} L \\ &= L^T ((V_R - \tau_R)^{-1} - (V_R + i\Omega_R)^{-1}) L \\ &= L^T (V_R - \tau_R)^{-1} ((V_R + i\Omega_R) - (V_R - \tau_R)) (V_R + i\Omega_R)^{-1} L \\ &= L^T (V_R - \tau_R)^{-1} (\tau_R + i\Omega_R) (V_R + i\Omega_R)^{-1} L, \end{aligned}$$

where we employed Lemma 7.12 in the form $\gamma_{C_1} + i\Omega_{C_1} = L^T (V_R + i\Omega_R)^{-1} L$ and performed some elementary algebraic manipulations. Now, from the third line of the above calculation it is clear that $\sigma_{C_1} - i\Omega_{C_1} \geq 0$, since from $V_R - i\Omega_R \geq V_R - \tau_R > 0$ we immediately deduce $(V_R - \tau_R)^{-1} \geq (V_R + i\Omega_R)^{-1}$. This shows that σ_{C_1} is a valid QCM. Moreover, observe that

$$\begin{aligned} \text{rk}(\sigma_{C_1} - i\Omega_{C_1}) &= \text{rk}(L^T (V_R - \tau_R)^{-1} (\tau_R + i\Omega_R) (V_R + i\Omega_R)^{-1} L) \\ &= \text{rk}(\tau_R + i\Omega_R) \\ &= n_R \\ &= n_{C_1}, \end{aligned}$$

which tells us that σ_{C_1} is also a pure QCM. \square

Now, we are ready to state the conclusive result of the present chapter.

Theorem 7.14. *For all bipartite QCMs $V_{AB} \geq i\Omega_{AB}$, the Rényi-2 Gaussian squashed entanglement coincides with the Rényi-2 Gaussian entanglement of formation, i.e.*

$$E_{\text{sq},2}^G(A : B)_V = E_{F,2}^G(A : B)_V. \quad (7.47)$$

Proof. The inequality $E_{\text{sq},2}^G(A : B)_V \geq E_{F,2}^G(A : B)_V$ is an easy consequence of (7.42) together with (7.43). To show the converse, we employ the expression (7.41) for the Rényi-2 Gaussian entanglement of formation. Consider an arbitrary purification γ_{ABC} of V_{AB} , and pick a pure state $\tau_{AB} \leq V_{AB}$. By construction, we have $\gamma_{AB} = V_{AB}$. Now, thanks to Proposition 7.13 one can

construct a sequence of measurements identified by $\sigma_C(t)$ such that (7.45) holds. Then, we have

$$\begin{aligned} \frac{1}{2}I_M(A : B)_\tau &= \frac{1}{2}I_M(A : B)_{\lim_{t \rightarrow 0^+} (\gamma_{ABC} + 0_{AB} \oplus \sigma_C(t)) / (\gamma_C + \sigma_C(t))} \\ &\stackrel{(1)}{=} \lim_{t \rightarrow 0^+} \frac{1}{2}I_M(A : B)_{(\gamma_{ABC} + 0_{AB} \oplus \sigma_C(t)) / (\gamma_C + \sigma_C(t))} \\ &\stackrel{(2)}{=} \lim_{t \rightarrow 0^+} \frac{1}{2}I_M(A : B|C)_{\gamma_{ABC} + 0_{AB} \oplus \sigma_C(t)} \\ &\stackrel{(3)}{\geq} E_{\text{sq},2}^G(A : B)_V, \end{aligned}$$

where we used, in order: (1) the continuity of the log-det mutual information; (2) the identity (7.14); and (3) the fact that the QCMs $\gamma_{ABC} + 0_{AB} \oplus \sigma_C(t)$ constitute valid extensions of V_{AB} and hence legitimate ansatzes in (7.43). \square

Remark. A by-product of the above proof of Theorem 7.14 is that in (7.43) we can restrict ourselves to systems of bounded size $n_C \leq n_{AB} = n_A + n_B$. Moreover, up to limits the extension can be taken of the form $\gamma_{ABC} + 0_{AB} \oplus \sigma_C$, where γ_{ABC} is a fixed purification of V_{AB} and σ_C is a pure QCM.

This surprising identity between two seemingly very different entanglement measures, even though tailored to Gaussian states, is remarkable. On the one hand, it provides an interesting operational interpretation for the Rényi-2 Gaussian entanglement of formation in terms of log-det conditional mutual information, via the recoverability framework. On the other hand, it simplifies the notoriously difficult evaluation of the squashed entanglement, in this case restricted to Gaussian extensions and log-det entropy, because it recasts it as an optimisation of the form (7.40), which thus involves matrices of bounded instead of unbounded size (more precisely, of the same size as the mixed QCM whose entanglement is being computed). In general, Theorem 7.14 allows us to export useful properties between the two frameworks it connects. For instance, it follows from the identity (7.47) that the Rényi-2 Gaussian squashed entanglement is faithful on Gaussian states and a monotone under Gaussian local operations and classical communication; in contrast, proving the property of faithfulness for the standard squashed entanglement was a very difficult step to perform [162]. On the other hand, the arguments establishing many basic properties of the standard squashed entanglement can be imported from [11] and applied to (7.43), providing new proofs of the same properties for the Rényi-2 Gaussian entanglement of formation. Let us give an example of how effective is the interplay between the two frameworks by providing an alternative, one-line proof of Corollary 6.15, i.e. of the inequality

$$E_{F,2}^G(A : B_1 \dots B_k)_V \geq \sum_{j=1}^k E_{F,2}^G(A : B_j)_V.$$

Alternative proof of Corollary 6.15. As usual, it suffices to show the above inequality in the case of three parties:

$$E_{F,2}^G(A : BC) \geq E_{F,2}^G(A : B) + E_{F,2}^G(A : C).$$

Thanks to Theorem 7.14, we can prove this relation for the Rényi-2 Gaussian squashed entanglement instead. We use basically the same argument as in [11, Proposition 4]. Namely, call V_{ABC} the QCM of the system ABC . Then for all extensions V_{ABCE} of V_{ABC} one has

$$\begin{aligned} I_M(A : BC|E)_V &= I_M(A : B|E)_V + I_M(A : C|BE) \\ &\geq 2 E_{\text{sq},2}^G(A : B)_V + 2 E_{\text{sq},2}^G(A : C)_V, \end{aligned}$$

where we applied the chain rule for the conditional mutual information together with the obvious facts that V_{ABE} is a valid extension of V_{AB} and V_{ABCE} a valid extension of V_{AC} . \square

This proof is substantially identical to the one provided in [11] to show the monogamy of the standard squashed entanglement. Then, in a way, the identity (7.47) explains why the Rényi-2 Gaussian entanglement of formation is monogamous and additive, something that looked a bit like an accident in Chapter 6. Now, we can say that it is not an accident, and on the contrary follows from the fact that the Rényi-2 Gaussian entanglement of formation coincides with its corresponding squashed entanglement, and the latter is monogamous and additive almost by construction.

7.6 Conclusions

In this chapter, we analysed the SSA inequality for the log-det entropy from a matrix analysis viewpoint and explored some of its applications. We first derived new necessary and sufficient conditions for saturation of said inequality. In the context of classical recoverability, we then provided an explicit form for the Gaussian Petz recovery map and further obtained a strengthening of SSA by constructing a faithful lower bound to a log-det entropy based conditional mutual information. We finally specialised to quantum Gaussian states, for which the log-det entropy reduces to the Rényi-2 entropy, and defined a corresponding Gaussian version of the squashed entanglement measure. Surprisingly, we showed that the latter measure coincides with the Rényi-2 entanglement of formation defined via a Gaussian convex roof construction [260]. In turn, this allows us to build a bridge connecting the two frameworks, that can be used to establish new properties of a measure by looking at the other, or to provide simpler and more instructive proofs of known properties. This work, together with the preceding Chapter 6, casts further light on the connections between matrix analysis (in particular determinantal inequalities) and information theory in both classical and quantum settings.

Despite the progress, many of the quantifiers we talked about in this chapter still present some thorny open questions. A major problem concerns the entanglement of formation of Gaussian states, since to the best of our knowledge it is still unknown whether Gaussian decompositions always attain the global infimum in the convex roof optimisation [?]. Analogously, one could wonder whether for Gaussian states also the optimisation that defines the standard

squashed entanglement [11] can be restricted to Gaussian extensions. This would considerably simplify the computation of the squashed entanglement on states very relevant for applications in quantum optics. Finally, it could be interesting to establish whether the equivalence between the Rényi-2 Gaussian squashed entanglement defined here and the Rényi-2 Gaussian entanglement of formation defined in [260] further extends to a third measure of entanglement, namely the recently introduced Gaussian intrinsic entanglement [318]. This latter measure is operationally relevant, since it constitutes – by construction – an upper bound on the secret key rate that is achievable by the application of local Gaussian measurements.

In future work we will focus on these problem, thus following the guiding star of quantum information science, operational relevance.

Appendix A

Computation rules for polars

This appendix is dedicated to presenting a proof of the computation rules (1.10), (1.11), (1.12) and (1.13) we made use of in the main text. We will mainly adapt material from [42, IV §1.5] and [85]. In the following, let E be a Banach space with Banach dual E^* . For the definition of the polars M°, N° of two subsets $M \subseteq E$ and $N \subseteq E^*$, see (1.8) and (1.9). Some observations are in order:

- (a) for an arbitrary family $\{M_i\}_i$ of subsets either of E or of E^* , one has $(\bigcup_i M_i)^\circ = \bigcap_i M_i^\circ$;
- (b) polar sets are always convex, and moreover norm-closed (if in E) or weak*-closed (if in E^*);
- (c) for $M \subseteq E$ and $N \subseteq E^*$, one has

$$M^\circ = (\text{co}(M))^\circ = (\text{cl}(M))^\circ = (\text{cl}_w(M))^\circ, \quad (\text{A.1})$$

$$N^\circ = (\text{co}(N))^\circ = (\text{cl}(N))^\circ = (\text{cl}_{w^*}(N))^\circ; \quad (\text{A.2})$$

- (d) if $C \subseteq E$ is a cone, then for all $M \subseteq E$ one has $(M \pm C)^\circ = M^\circ \cap (\mp C^*)$.

The following lemma contains as special cases the two computation rules for double polar, (1.10) and (1.11).

Lemma A.1. *For $M \subseteq E$ and $N \subseteq E^*$, one has*

$$M^{\circ\circ} = \text{cl}(\text{co}(M \cup \{0\})) \quad (\text{A.3})$$

$$N^{\circ\circ} = \text{cl}_{w^*}(\text{co}(M \cup \{0\})). \quad (\text{A.4})$$

Proof. Let us start by proving (A.3). Since $\{0\}, M \subseteq M^{\circ\circ}$ and the latter set is convex and closed, one has $\text{cl}(\text{co}(M \cup \{0\})) \subseteq M^{\circ\circ}$. Conversely, consider $x \notin \text{cl}(\text{co}(M \cup \{0\}))$. We have to show that $x \notin M^{\circ\circ}$, i.e. that there is $\varphi \in M^\circ$ such that $\langle \varphi, x \rangle > 1$. Applying Corollary 1.8 (a) with $M = \{x\}$ and $N = \text{cl}(\text{co}(M \cup \{0\}))$, we can find $0 < \epsilon < 1$ and a continuous functional

$\varphi' \in E^*$ such that $\langle \varphi', y \rangle \geq -1 + \epsilon$ for all $y \in \text{cl}(\text{co}(M \cup \{0\}))$ and $\langle \varphi', x \rangle \leq -1$. By restricting y to belong to M , we see immediately that $\varphi := -\frac{\varphi'}{1-\epsilon} \in M^\circ$ but $\langle \varphi, x \rangle \geq \frac{1}{1-\epsilon} > 1$, yielding the claim.

The proof of (A.4) proceeds in an analogous fashion. On the one hand, it is relatively easy to verify that $\text{cl}_{w^*}(\text{co}(N \cup \{0\})) \subseteq N^{\circ\circ}$. On the other hand, if $\varphi \notin \text{cl}_{w^*}(\text{co}(N \cup \{0\}))$ by definition there is $y \in E$ such that $(\varphi + y^{-1}([-1, 1])) \cap \text{co}(N \cup \{0\}) = \emptyset$, where we are seeing y as a functional on E^* . This is the same as saying that $|\langle \varphi - \psi, y \rangle| > 1$ for all $\psi \in \text{co}(N \cup \{0\})$. Since the set of values taken by $\langle \varphi - \psi, y \rangle$ for different choices of ψ is clearly convex, we can assume that $\langle \varphi - \psi, y \rangle > 1$ holds for all $\psi \in \text{co}(N \cup \{0\})$. Picking $\psi = 0$ we find $\langle \varphi, y \rangle > 1$, while restricting to $\psi \in N$ and defining $x := \frac{y}{\langle \varphi, y \rangle - 1}$ yields $x \in N^\circ$ after a quick computation. Since $\langle \varphi, x \rangle = \frac{\langle \varphi, y \rangle}{\langle \varphi, y \rangle - 1} > 1$, we see that $\varphi \notin N^{\circ\circ}$, implying that $N^{\circ\circ} \subseteq \text{cl}_{w^*}(\text{co}(N \cup \{0\}))$ and concluding the proof. \square

Remark. The reader might have noticed some sort of asymmetry between (A.3), where the norm closure appears, and (A.4), where we have instead the weak* closure. This asymmetry in fact artificial, in the sense that $\text{cl}(N) = \text{cl}_w(N)$ for all convex subsets $N \subseteq E$ thanks to Corollary 1.12, and thus in (A.4) we can replace the norm closure with the weak closure, if needed.

Lemma A.2. *Let $M, N \subseteq E$ be convex, closed, and such that $0 \in M \cap N$. Then*

$$(M \cap N)^\circ = \text{cl}_{w^*}(\text{co}(M^\circ \cup N^\circ)). \quad (\text{A.5})$$

Proof. We start by observing that

$$\begin{aligned} (\text{cl}_{w^*}(\text{co}(M^\circ \cup N^\circ)))^\circ &\stackrel{(1)}{=} (M^\circ \cup N^\circ)^\circ \\ &\stackrel{(2)}{=} M^{\circ\circ} \cap N^{\circ\circ} \\ &\stackrel{(3)}{=} M \cap N. \end{aligned}$$

The justification of the above derivation is as follows: (1) is an application of identities (A.2); (2) is deduced from observation (a) above; and (3) follows from Lemma A.1. Now, employing again Lemma A.1 we find

$$\text{cl}_{w^*}(\text{co}(M^\circ \cup N^\circ)) = (\text{cl}_{w^*}(\text{co}(M^\circ \cup N^\circ)))^{\circ\circ} = (M \cap N)^\circ,$$

which concludes the proof. \square

The above rule for computing polars of intersections is so useful, that it makes sense to extend it further. In particular, we are interested in the following strengthening.

Lemma A.3. [85, Lemma 1] *Let $M, N \subseteq E$ be convex and such that $0 \in M \cap N$.*

(a) *If $\text{cl}(M \cap N) = \text{cl}(M) \cap \text{cl}(N)$ then*

$$(M \cap N)^\circ = \text{cl}_{w^*}(\text{co}(M^\circ \cup N^\circ)).$$

(b) *The identity $\text{cl}(M \cap N) = \text{cl}(M) \cap \text{cl}(N)$ holds if M, N are closed or if $0 \in \text{int}(M) \cap \text{int}(N)$.*

Proof. We start with claim (a). Using (A.1) together with Lemma A.2, we find

$$\begin{aligned} (M \cap N)^\circ &= (\text{cl}(M \cap N))^\circ \\ &= (\text{cl}(M) \cap \text{cl}(N))^\circ \\ &= \text{cl}_{w^*}(\text{co}(\text{cl}(M)^\circ \cup \text{cl}(N)^\circ)) \\ &= \text{cl}_{w^*}(\text{co}(M^\circ \cup N^\circ)). \end{aligned}$$

Now, let us come to claim (b). If M, N are closed then so is $M \cap N$, and hence trivially $\text{cl}(M \cap N) = M \cap N = \text{cl}(M) \cap \text{cl}(N)$. In general, it is easy to verify that $\text{cl}(M \cap N) \subseteq \text{cl}(M) \cap \text{cl}(N)$ holds always. Thus, it remains to show that the equality holds when $0 \in \text{int}(M) \cap \text{int}(N)$. Take $\delta > 0$ such that $\|y\| \leq \delta \Rightarrow y \in M \cap N$. Consider $x \in \text{cl}(M) \cap \text{cl}(N)$, and let us show that $tx \in M \cap N$ for all $0 \leq t < 1$, so that $x = \lim_{t \rightarrow 1} tx \in \text{cl}(M \cap N)$. One can pick $x' \in M$ such that $\|x - x'\| \leq \frac{(1-t)\delta}{t}$, which allows us to write $tx = tx' + (1-t)y$, with $y := \frac{t}{1-t}(x - x')$ satisfying $\|y\| \leq \delta$ and thus $y \in M$. Since we have written tx as a convex combination of two points $x', y \in M$, and M is convex, we deduce that $tx \in M$. Analogously, one can show that $tx \in N$, so that $tx \in M \cap N$ for all $t \in [0, 1)$ and finally $x \in \text{cl}(M \cap N)$. \square

Now, we are ready to prove (1.12) and (1.13) from the main text.

Corollary A.4. *Let E be an ordered Banach space with closed positive cone E_+ , and let the Banach dual E^* be ordered by the cone E_+^* of positive continuous functionals. Denote by B and B° the unit balls of E and E^* , respectively. Then $[B]^\circ =]B^\circ[$ and $[B^\circ]^\circ = \text{cl}]B[$.*

Proof. Let us start with the first claim. We have:

$$\begin{aligned} [B]^\circ &= ((B + E_+) \cap (B - E_+))^\circ \\ &\stackrel{(1)}{=} \text{cl}_{w^*}(\text{co}((B + E_+)^\circ \cup (B - E_+)^\circ)) \\ &\stackrel{(2)}{=} \text{cl}_{w^*}(\text{co}((B^\circ \cap -E_+^*) \cup (B^\circ \cap E_+^*))) \\ &\stackrel{(3)}{=} \text{co}((B^\circ \cap -E_+^*) \cup (B^\circ \cap E_+^*)) \\ &=]B^\circ[. \end{aligned}$$

The justification of these steps is as follows: (1) follows from Lemma A.3, which is applicable since the sets $B \pm E_+$ are clearly convex and have the origin as an interior point; (2) see observation (d) above; (3) since $(B^\circ \cap -E_+^*) \cup (B^\circ \cap E_+^*)$ is a weak*-closed subset of the unit ball B° , which is weak*-compact by the Banach-Alaoglu theorem (Theorem 1.14), it is itself weak*-compact; therefore, it is not difficult to realise that $\text{co}((B^\circ - E_+^*) \cup (B^\circ + E_+^*))$ is already weak*-closed.

The second claim can be shown in an analogous fashion, but we have first to convince ourselves that $B^\circ + E_+^*$ is weak*-closed. This can be done once more thanks to the Banach-Alaoglu theorem (Theorem 1.14). In fact, if $(\lambda_\alpha + \mu_\alpha)_\alpha \subseteq B^\circ + E_+^*$ is such that $w^*\text{-lim}_\alpha (\lambda_\alpha + \mu_\alpha) = \varphi \in E^*$, the weak*-compactness of B° ensures that we can extract a subnet $(\lambda_\beta)_\beta$ of $(\lambda_\alpha)_\alpha \subseteq B^\circ$ that converges to some $\lambda \in B^\circ$ in the weak* topology. It is then easy to realise that on this subnet $w^*\text{-lim}_\beta \mu_\beta = \varphi - \lambda$ necessarily. Moreover, since E_+^* is weak*-closed, we get $\varphi - \lambda \in E_+^*$ and in turn $\varphi = \lambda + (\varphi - \lambda) \in B^\circ + E_+^*$. Naturally, the same reasoning also shows that $B^\circ - E_+^*$ is weak*-closed. Now we can employ Lemma A.2 and conclude that

$$\begin{aligned}
[B^\circ]^\circ &= ((B^\circ + E_+^*) \cap (B^\circ - E_+^*))^\circ \\
&= \text{cl}_w (\text{co}((B^\circ + E_+^*)^\circ \cup (B^\circ - E_+^*)^\circ)) \\
&= \text{cl}_w (\text{co}((B - E_+) \cup (B + E_+))) \\
&= \text{cl}_w]B[\\
&= \text{cl}]B[,
\end{aligned}$$

where in the last step we used once more the fact that norm closure and weak closure coincide for convex sets. \square

Appendix B

Complements on the proof of Ludwig's embedding theorem

Throughout this appendix, we will show directly that any two Banach spaces E_1, E_2 satisfying the hypotheses of Theorem 1.43 must be connected by an isometric order isomorphism, in the sense discussed at the beginning of Subsection 1.7.9. Let us start with a preliminary lemma that will allow us to characterise isometric order isomorphisms between base norm spaces in a computationally convenient way.

Lemma B.1. *Let E_1, E_2 be two base norm spaces with corresponding bases K_1, K_2 , and let $\Phi : E_1 \rightarrow E_2$ be an injective linear map. Then Φ is an isometric order isomorphism iff $\Phi(K_1) = K_2$.*

Proof. On the one hand, if Φ is an isometric order isomorphism, then by applying Lemmas 1.33 and 1.37 we see that

$$\begin{aligned} K_2 &= \{y \in E_{2+} : \|y\| = 1\} \\ &= \{\Phi(x) : x \in E_{1+}, \|\Phi(x)\| = 1\} \\ &= \{\Phi(x) : x \in E_{1+}, \|x\| = 1\} \\ &= \Phi(\{x \in E_{1+} : \|x\| = 1\}) \\ &= \Phi(K_1). \end{aligned}$$

On the other hand, let us assume that $\Phi(K_1) = K_2$ and that Φ is injective. Then $x \in \lambda K_1 \Leftrightarrow \Phi(x) \in \lambda K_2$, for all $x \in E_1$ and $\lambda \geq 0$. Applying this observation repeatedly, we see that for all $x, y \in E_1$ one has

$$\begin{aligned} x \geq 0 &\iff \exists \lambda \geq 0 : x \in \lambda K_1 \\ &\iff \exists \lambda \geq 0 : \Phi(x) \in \lambda K_2 \\ &\iff \Phi(x) \geq 0, \end{aligned}$$

hence Φ is an order isomorphism. Verifying that Φ is also an isometry is again straightforward using (1.21):

$$\begin{aligned} \|x\| &= \inf \{|\alpha| + |\beta| : x = \alpha a - \beta b, \alpha, \beta \in \mathbb{R}, a, b \in K_1\} \\ &= \inf \{|\alpha| + |\beta| : \Phi(x) = \alpha\Phi(a) - \beta\Phi(b), \alpha, \beta \in \mathbb{R}, \Phi(a), \Phi(b) \in K_2\} \\ &= \inf \{|\alpha| + |\beta| : \Phi(x) = \alpha\alpha' - \beta\beta', \alpha, \beta \in \mathbb{R}, \alpha', \beta' \in K_2\} \\ &= \|\Phi(x)\|. \end{aligned}$$

The last thing to check is that Φ is also surjective. This follows immediately from the fact that K_2 spans the positive cone E_{2+} , which is in turn spanning since E_2 is a base norm space. \square

Let us recall here also the definition of **adjoint** of a linear map between Banach spaces. For details, we refer the reader to [77, §4.10]. The adjoint $\Phi^* : E_2^* \rightarrow E_1^*$ of a map $\Phi : E_1 \rightarrow E_2$ between two Banach spaces E_1, E_2 is constructed as the unique linear operator such that

$$\langle \varphi_2, \Phi(x_1) \rangle = \langle \Phi^*(\varphi_2), x_1 \rangle \quad \forall \varphi_2 \in E_2^*, x_1 \in E_1. \quad (\text{B.1})$$

It is fairly easy to verify that the following holds.

Lemma B.2. *If a linear map $\Phi : E_1 \rightarrow E_2$ between two ordered Banach spaces is an isometric order isomorphism, then the same is true for $\Phi^{-1} : E_2 \rightarrow E_1$ and $\Phi^* : E_2^* \rightarrow E_1^*$.*

We are going to need two elementary results in linear algebra, which we state now without proof.

Lemma B.3. *Let V, W be real vector spaces, and let $K \subseteq V$ be a convex set, whose affine hull we denote by $\text{aff}(K) = \{\sum_{i=1}^n \lambda_i x_i : x_i \in K, \sum_{i=1}^n \lambda_i = 1\}$. Then every convex-linear application $f : K \rightarrow W$ admits a unique extension to an affine map $f : \text{aff}(K) \rightarrow W$.*

Lemma B.4. *Let $A \subseteq V$ an affine subspace of a real vector space V . If $0 \notin A$, then every affine application $f : A \rightarrow W$ (W another real vector space) can be extended to a linear map $f : V \rightarrow W$. If $\text{span}(A) = V$, this extension is unique.*

Another trivial yet useful observation is a slight generalisation of the dual formula (1.4) for the norm of a Banach space.

Lemma B.5. *Let E be a Banach space with dual E^* . If W is weak*-dense in E^* , then*

$$\|x\| = \sup_{\varphi \in W, \|\varphi\|_* \leq 1} |\langle \varphi, x \rangle|$$

for all $x \in E$.

Let us come now to the main result of this appendix, i.e. the direct proof of the uniqueness claim of Theorem 1.43.

Proposition B.6. *Let $\Omega \subseteq E_1$, $\Lambda \subseteq E_1^*$ convex subsets of a Banach space E_1 and of its dual E_1^* . Let E_1 be a base norm space with base $\text{cl}(\Omega)$, and assume $\text{span}(\Lambda)$ to be dense in E_1^* in the weak*-topology. Let $\phi : \Omega \rightarrow E_2$ and $\psi : \Lambda \rightarrow E_2^*$ be embeddings of Ω and Λ into another dual pair of Banach spaces E_2 and E_2^* such that:*

(i) $\langle \psi(\lambda), \phi(\omega) \rangle = \langle \lambda, \omega \rangle$ for all $\omega \in \Omega$ and $\lambda \in \Lambda$;

(ii) E_2 is a base norm space with base $\text{cl}(\phi(\Omega))$;

(iii) $\text{span}(\psi(\Lambda))$ is dense in E_2^* with the weak*-topology.

Then ϕ and ψ can be extended to isometric order isomorphisms $\Phi : E_1 \rightarrow E_2$ and $\Psi : E_1^* \rightarrow E_2^*$ such that $\Psi = (\Phi^*)^{-1}$

Proof.

- First of all, $\psi : \Lambda \rightarrow E_2^*$ is linear, in the sense that whenever one considers $\lambda, \eta \in \Lambda$ and $\alpha, \beta \in \mathbb{R}$ such that also $\alpha\lambda + \beta\eta \in \Lambda$, one has $\psi(\alpha\lambda + \beta\eta) = \alpha\psi(\lambda) + \beta\psi(\eta)$. This can be verified as follows.

$$\begin{aligned}
& \|\psi(\alpha\lambda + \beta\eta) - \alpha\psi(\lambda) - \beta\psi(\eta)\|_* \\
& \stackrel{(1)}{=} \sup_{\omega' \in \text{cl}(\phi(\Omega))} |\langle \psi(\alpha\lambda + \beta\eta) - \alpha\psi(\lambda) - \beta\psi(\eta), \omega' \rangle| \\
& \stackrel{(2)}{=} \sup_{\omega' \in \phi(\Omega)} |\langle \psi(\alpha\lambda + \beta\eta) - \alpha\psi(\lambda) - \beta\psi(\eta), \omega' \rangle| \\
& = \sup_{\omega \in \Omega} |\langle \psi(\alpha\lambda + \beta\eta) - \alpha\psi(\lambda) - \beta\psi(\eta), \phi(\omega) \rangle| \\
& = \sup_{\omega \in \Omega} |\langle \psi(\alpha\lambda + \beta\eta), \phi(\omega) \rangle - \alpha \langle \psi(\lambda), \phi(\omega) \rangle - \beta \langle \psi(\eta), \phi(\omega) \rangle| \\
& \stackrel{(3)}{=} \sup_{\omega \in \Omega} |\langle \alpha\lambda + \beta\eta, \omega \rangle - \alpha \langle \lambda, \omega \rangle - \beta \langle \eta, \omega \rangle| \\
& = 0.
\end{aligned}$$

The justification of these steps is the following: (1) we applied Corollary 1.39; (2) we restricted the supremum from $\text{cl}(\phi(\Omega))$ to $\phi(\Omega)$, which is possible because the function we are taking the supremum of is clearly continuous; (3) we used hypothesis (i) to drop ϕ and ψ inside the scalar product.

- Thanks to the aforementioned observations, it is clear that ψ can be extended uniquely to a linear map $\psi : \text{span}(\Lambda) \rightarrow E_2^*$. Few manipulations similar to those we carried out above show also that ψ is norm-preserving.

In fact, employing hypothesis (ii) together with Corollary 1.39 yields

$$\begin{aligned}
\|\psi(y)\|_* &= \sup_{\omega' \in \text{cl}(\phi(\Omega))} |\langle \psi(y), \omega' \rangle| \\
&= \sup_{\omega' \in \phi(\Omega)} |\langle \psi(y), \omega' \rangle| \\
&= \sup_{\omega \in \Omega} |\langle \psi(y), \phi(\omega) \rangle| \\
&= \sup_{\omega \in \Omega} |\langle y, \omega \rangle| \\
&= \|y\|_*
\end{aligned}$$

for all $y \in \text{span}(\Lambda)$. For the second last step we made use of the equality in hypothesis (i) extended from $\Lambda \times \Omega$ to $\text{span}(\Lambda) \times \Omega$.

- $\phi : \Omega \rightarrow E_2$ is convex-linear. In fact, for all $\omega_1, \omega_2 \in \Omega$, $\lambda \in \Lambda$, and $p \in [0, 1]$, we have

$$\begin{aligned}
\langle \phi(p\omega_1 + (1-p)\omega_2), \psi(\lambda) \rangle &= \langle p\omega_1 + (1-p)\omega_2, \lambda \rangle \\
&= p \langle \omega_1, \lambda \rangle + (1-p) \langle \omega_2, \lambda \rangle \\
&= p \langle \phi(\omega_1), \psi(\lambda) \rangle + (1-p) \langle \phi(\omega_2), \psi(\lambda) \rangle \\
&= \langle p\phi(\omega_1) + (1-p)\phi(\omega_2), \psi(\lambda) \rangle.
\end{aligned}$$

Since $\text{span}(\psi(\Lambda))$ is weak*-dense in E_2 , the above equality implies that

$$\phi(p\omega_1 + (1-p)\omega_2) = p\phi(\omega_1) + (1-p)\phi(\omega_2),$$

showing that $\phi : \Omega \rightarrow E_2$ is indeed convex-linear, as claimed.

- Now, we can use the above observation together with Lemma B.3 to extend ϕ to an affine application defined on the whole $\text{aff}(\Omega)$. Since $\text{aff}(\Omega) \subseteq \text{aff}(\text{cl}(\Omega))$ and the latter (being the linear span of a base of a base norm space) does not contain the origin, we can immediately apply Lemma B.4 and extend further ϕ to a linear application $\phi : \text{span}(\Omega) \rightarrow E_2$. Let us show that also ϕ preserves the norm, i.e. that $\|\phi(x)\| = \|x\|$ for all $x \in E_1$.

$$\begin{aligned}
\|\phi(x)\| &= \sup \{ |\langle y', \phi(x) \rangle| : y' \in E_2, \|y'\|_* \leq 1 \} \\
&\stackrel{(1)}{=} \sup \{ |\langle y', \phi(x) \rangle| : y' \in \text{span}(\psi(\Lambda)), \|y'\|_* \leq 1 \} \\
&= \sup \{ |\langle \psi(y), \phi(x) \rangle| : y \in \text{span}(\Lambda), \|\psi(y)\|_* \leq 1 \} \\
&\stackrel{(2)}{=} \sup \{ |\langle y, x \rangle| : y \in \text{span}(\Lambda), \|\psi(y)\|_* \leq 1 \} \\
&\stackrel{(3)}{=} \sup \{ |\langle y, x \rangle| : y \in \text{span}(\Lambda), \|y\|_* \leq 1 \} \\
&\stackrel{(4)}{=} \|x\|.
\end{aligned}$$

In the above derivation, we used: (1) Lemma B.5; (2) hypothesis (i), extended from $\Omega \times \Lambda$ to $\text{span}(\Omega) \times \text{span}(\Lambda)$; (3) the already shown fact that ψ preserves the norm; (4) again Lemma B.5.

- Since $\phi : \text{span}(\Omega) \rightarrow E_2$ is bounded, there is a continuous (linear) extension $\Phi : \text{cl}(\text{span}(\Omega)) \rightarrow E_2$. Now, it is easy to verify that $\text{cl}(\text{span } M) \supseteq \text{span}(\text{cl } \Omega)$ for all subsets M of any Banach space. Applying this with $M = \Omega$ yields $\text{cl}(\text{span}(\Omega)) \supseteq \text{span}(\text{cl } \Omega) = E_1$, where the last equality holds because $\text{cl } \Omega$ is a base of the positive cone of a base norm space, and as such is spanning. Hence $\text{cl}(\text{span}(\Omega)) = E_1$, and we have constructed an extension $\Phi : E_1 \rightarrow E_2$. Few useful properties of ϕ that are retained by Φ are as follows.

- Hypothesis (i) can be extended to $\langle \psi(y), \Phi(x) \rangle = \langle y, x \rangle$ for all $x \in E_1$ and $y \in \text{span}(\Lambda)$.
- Performing a continuous linear extension preserves the norm, hence the identity $\|\Phi(x)\| = \|x\|$ is obeyed by all $x \in E_1$. This implies that Φ is an isometry, and in particular that it is injective.
- Consequently, $\Phi(\text{cl}(\Omega)) = \text{cl}(\phi(\Omega))$ holds. To show this, we first take $\omega \in \text{cl}(\Omega)$ and construct a sequence $(\omega_n)_n \subset \Omega$ such that $\lim_n \|\omega - \omega_n\| = 0$. Then, we verify that

$$\lim_n \|\Phi(\omega) - \Phi(\omega_n)\| = \lim_n \|\Phi(\omega - \omega_n)\| = \lim_n \|\omega - \omega_n\| = 0,$$

which implies that $\Phi(\omega) \in \text{cl}(\phi(\Omega))$ because $\Phi(\omega_n) = \phi(\omega_n) \in \phi(\Omega)$. Since $\omega \in \text{cl}(\Omega)$ was generic, we obtain $\Phi(\text{cl}(\Omega)) \subseteq \text{cl}(\phi(\Omega))$.

Conversely, pick $\omega' \in \text{cl}(\phi(\Omega))$, and let $(\omega_n)_n \subset \Omega$ be a sequence such that $\lim_n \|\omega' - \phi(\omega_n)\| = 0$. Since ϕ is an isometry and $(\phi(\omega_n))_n$ is a Cauchy sequence, the same is true for $(\omega_n)_n$. Hence, the sequence $(\omega_n)_n$ converges in norm to some $\omega \in \text{cl}(\Omega)$, which in turn satisfies $\omega' = \Phi(\omega) \in \Phi(\text{cl}(\Omega))$ because Φ is a bounded (equivalently, continuous) map. Since $\omega' \in \text{cl}(\phi(\Omega))$ was generic, we have shown that $\text{cl}(\phi(\Omega)) \subseteq \Phi(\text{cl}(\Omega))$ and thus completed the proof of the claim.

Since $\text{cl}(\Omega)$ and $\text{cl}(\phi(\Omega))$ are the bases of the base norm spaces E_1 and E_2 , from Lemma B.1 we immediately deduce that Φ is an isometric order isomorphism.

- We are now in position to extend also the action of ψ from $\text{span}(\Lambda)$ to the whole E_1^* . To do this, we can exploit the already proven surjectivity of Φ . Namely, for a generic $y \in E_1^*$, we can simply *define* the action of the transformed functional $\Psi(y) \in E_2^*$ to be

$$\langle \Psi(y), \Phi(x) \rangle := \langle y, x \rangle .$$

Observe that Ψ is clearly an extension of ψ , since for $y \in \text{span}(\Lambda)$ by the known properties of Φ one has $\langle \psi(y), \Phi(x) \rangle = \langle y, x \rangle = \langle \Psi(y), \Phi(x) \rangle$ for all $x \in E_1$, which implies $\langle \Psi(y) - \psi(y), \Phi(x) \rangle = 0$ and in turn $\Psi(y) = \psi(y)$, since $\Phi(x)$ runs over the whole space E_2 . Clearly, since $\langle y, x \rangle = \langle \Psi(y), \Phi(x) \rangle = \langle y, (\Psi^* \Phi)(x) \rangle$ holds for all $x \in E_1$ and $y \in E_1^*$,

we have $\Psi^*\Phi = \text{id}$ and therefore $\Psi = (\Phi^*)^{-1}$, as claimed. In particular, Lemma B.2 tells us that Ψ is an isometric isomorphism because so is Φ . This concludes the proof.

□

Appendix C

Tensor product rule for GPTs

This appendix is giving an elementary proof of the tensor product rule for finite dimensional GPTs (Lemma 2.7). Let us first restate it.

Lemma. *Let the finite-dimensional GPTs A, B and AB satisfy Axioms 4, 5, 6. Then there are isomorphisms $\mathcal{F} : V_A \otimes V_B \rightarrow V_{AB}$ and $\mathcal{F}_* : V_A^* \otimes V_B^* \rightarrow V_{AB}^*$ such that:*

- (a) $\mathcal{F}(\omega_A \otimes \tau_B) = \mathcal{J}(\omega_A, \tau_B)$ for all states $\omega_A \in \Omega_A, \tau_B \in \Omega_B$;
- (b) $\mathcal{F}_*(e_A \otimes f_B) = \mathcal{J}_*(e_A, f_B)$ for all effects $e_A \in [0, u_A], f_B \in [0, u_B]$;
- (c) $\mathcal{F}_*(u_A \otimes u_B) = u_{AB}$; and
- (d) $\mathcal{F}_*^{-1} = \mathcal{F}^*$, where \mathcal{F}^* is the adjoint of \mathcal{F} .

Proof. We first show that the composition map $\mathcal{J} : \Omega_A \times \Omega_B \rightarrow \Omega_{AB} \subseteq V_{AB}$ can be extended to a bilinear application acting on the whole $V_A \times V_B$. Since \mathcal{J} is convex-bilinear, for all fixed $\omega_A \in \Omega_A$ the resulting map $\mathcal{J}(\omega_A, \cdot) : \Omega_B \rightarrow V_{AB}$ is convex-linear. We can then apply Lemmas B.3 and B.4 to extend it to a linear application $\mathcal{J}(\omega_A, \cdot) : V_B \rightarrow V_{AB}$. This way we have sketched of obtaining a linear application starting from a state $\omega_A \in \Omega_A$ amounts to a map $\Omega_A \rightarrow \mathcal{L}(V_B, V_{AB})$. This latter map is easily seen to be convex-linear, and therefore (again by Lemmas B.3 and B.4) can be extended to a linear application $V_A \rightarrow \mathcal{L}(V_B, V_{AB})$. A little thought shows that one can equivalently think of such an application as a bilinear map $V_A \times V_B \rightarrow V_{AB}$, which is immediately seen to extend \mathcal{J} and therefore will be denoted again with \mathcal{J} .

Now, the universal property of the tensor product allows us to ‘lift’ \mathcal{J} to a linear application $\mathcal{F} : V_A \otimes V_B \rightarrow V_{AB}$, which satisfies $\mathcal{F}(x_A \otimes x_B) = \mathcal{J}(x_A, x_B)$ for all $x_A \in V_A, x_B \in V_B$, meeting requirement (a). Clearly, the same reasoning can be applied to the effects, and one obtains instead a linear map $\mathcal{F}_* : V_A^* \otimes V_B^* \rightarrow V_{AB}^*$ that meets (b). The factorisation rule for the probabilities now reads

$$\langle \mathcal{F}_*(\varphi_A \otimes \varphi_B), \mathcal{F}(x_A \otimes x_B) \rangle = \langle \varphi_A, x_A \rangle \langle \varphi_B, x_B \rangle .$$

The above identity implies that $\mathcal{F}, \mathcal{F}_*$ are injective. Let us show this latter claim for \mathcal{F} , as the reasoning for \mathcal{F}_* is totally analogous. If one has $\mathcal{F}(x_{AB}) = 0$ with $x_{AB} = \sum_i x_A^{(i)} \otimes x_B^{(i)} \in V_A \otimes V_B$, one can write

$$\begin{aligned} \langle \varphi_A \otimes \varphi_B, x_{AB} \rangle &= \sum_i \langle \varphi_A, x_A^{(i)} \rangle \langle \varphi_B, x_B^{(i)} \rangle \\ &= \sum_i \langle \mathcal{F}_*(\varphi_A \otimes \varphi_B), \mathcal{F}(x_A^{(i)} \otimes x_B^{(i)}) \rangle \\ &= \langle \mathcal{F}_*(\varphi_A \otimes \varphi_B), \mathcal{F}(x_{AB}) \rangle \\ &= 0. \end{aligned}$$

Since product functionals generate $V_A^* \otimes V_B^*$, this is possible only if $x_{AB} = 0$.

Next, let us show that $\mathcal{F}, \mathcal{F}_*$ are in fact surjective, so that they become linear isomorphisms. To this end, we first use Axiom 6 at the level of effects to see that $\mathcal{F}_*(u_A \otimes u_B) = u_{AB}$, thus verifying (c). Since both sides of the equation are legitimate effects, we just need to check that they produce the same statistics on all local states:

$$\begin{aligned} &\langle \mathcal{F}_*(u_A \otimes u_B) - u_{AB}, \mathcal{F}(\omega_A \otimes \omega_B) \rangle \\ &= \langle \mathcal{F}_*(u_A \otimes u_B), \mathcal{F}(\omega_A \otimes \omega_B) \rangle - \langle u_{AB}, \mathcal{F}(\omega_A \otimes \omega_B) \rangle \\ &= \langle u_A, \omega_A \rangle \langle u_B, \omega_B \rangle - \langle u_{AB}, \mathcal{F}(\omega_A \otimes \omega_B) \rangle \\ &= 1 - 1 \\ &= 0. \end{aligned}$$

Then, we proceed to show that the image $\mathcal{F}_*(V_A^* \otimes V_B^*)$ of \mathcal{F}_* has trivial annihilator, i.e. is such that

$$\langle \varphi_{AB}, x_{AB} \rangle = 0 \quad \forall \varphi_{AB} \in \mathcal{F}_*(V_A^* \otimes V_B^*) \quad \implies \quad x_{AB} = 0,$$

for all $x_{AB} \in V_{AB}$. In fact, if this were not the case, one could pick a state ω_{AB} that belongs to the relative interior of Ω_{AB} (equivalently, to the interior of the positive cone C_{AB}), a vector $x_{AB} \in V_{AB}$ enjoying the above property, and construct $\omega'_{AB} := \omega_{AB} + \epsilon x_{AB}$. This ω'_{AB} is normalised, because

$$\langle u_{AB}, x_{AB} \rangle = \langle \mathcal{F}_*(u_A \otimes u_B), x_{AB} \rangle = 0$$

and hence $\langle u_{AB}, \omega'_{AB} \rangle = 1$, and moreover is a legitimate state for sufficiently small ϵ , as ω_{AB} was chosen to be in the interior of the positive cone. It is also easy to see that ω'_{AB} and ω_{AB} have the same local statistics, as x_{AB} annihilates all local effects $\mathcal{F}_*(e_A \otimes e_B)$. This shows that the annihilator of $\mathcal{F}_*(V_A^* \otimes V_B^*)$ is the trivial subspace, and since the only subspace of the dual with trivial annihilator is the whole space, we conclude that $\mathcal{F}_*(V_A^* \otimes V_B^*) = V_{AB}^*$ and \mathcal{F}_* is surjective, thus an isomorphism. A simple dimension count then implies that also \mathcal{F} is an isomorphism, as

$$\dim V_{AB} = \dim V_{AB}^* = \dim(V_A^* \otimes V_B^*) = \dim(V_A^*) \dim(V_B^*) = \dim V_A \dim V_B.$$

Finally, Axiom 6 together with the definition of adjoint (B.1) shows that

$$\begin{aligned} \langle \varphi_A, x_A \rangle \langle \varphi_B, x_B \rangle &= \langle \mathcal{F}_*(\varphi_A \otimes \varphi_B), \mathcal{F}(x_A \otimes x_B) \rangle \\ &= \langle (\mathcal{F}^* \mathcal{F}_*)(\varphi_A \otimes \varphi_B), x_A \otimes x_B \rangle, \end{aligned}$$

implying that $\mathcal{F}^* \mathcal{F}_*(\varphi_A \otimes \varphi_B) = \varphi_A \otimes \varphi_B$ for all φ_A, φ_B (because product vectors span $V_A \otimes V_B$) and in turn that $\mathcal{F}^* \mathcal{F}_* = \text{id}$, because product functionals span $V_A^* \otimes V_B^*$. This shows that also condition (d) is met and thus concludes the proof. \square

Remark. It is not enough to assume that the local tomography principle holds just at the level of states. In fact, consider $V_{AB} = (V_A \otimes V_B) \oplus \mathbb{R}$, with $u_{AB} = (u_A \otimes u_B, 1)$. With this definition, states in V_{AB} will have the form $\omega_{AB} = (kx_{AB}, 1 - k)$, where $x_{AB} \in V_A \otimes V_B$ is such that $\langle u_A \otimes u_B, x_{AB} \rangle = 1$, and $k \in \mathbb{R}$ is for now generic. The injection map $\mathcal{F} : \Omega_A \times \Omega_{AB} \rightarrow \Omega_{AB}$ acts as $\mathcal{F}(\omega_A, \omega_B) = (\omega_A \otimes \omega_B, 0)$, thus it is clearly convex-bilinear, and analogously for J_* . The rule of factorisation of probabilities is satisfied, and moreover every state is uniquely determined by the statistics of local effects, since from the numbers

$$\langle (e_A \otimes e_B, 0), \omega_{AB} \rangle = \langle (e_A \otimes e_B, 0), (kx_{AB}, 1 - k) \rangle = k \langle e_A \otimes e_B, x_{AB} \rangle$$

one can reconstruct k (by taking $e_A = u_A, e_B = u_B$) and consequently also x_{AB} , as product effects generate $V_A^* \otimes V_B^*$.

Appendix D

Upper bound on the inseparability ratio $S(2)$

The problem of determining the largest c satisfying Theorem 2.40 has been left open. Here we show at least that such an optimal c must be *strictly* less than 2, by providing the explicit upper bound $\sqrt{3}$.

Theorem D.1. *The real number $S(2)$ defined in (2.93) satisfies $S(2) \leq \sqrt{3}$. In other words, any c satisfying Theorem 2.40 must not exceed $\sqrt{3}$.*

Proof. We consider a set of norms N_t on \mathbb{R}^2 parametrised by the real number $t \in [0, 1]$. We take as the unit ball of N_t the octagon

$$B_{N_t} := \text{conv} \left\{ (\pm 1, 0), (0, \pm 1), \frac{1+\varepsilon}{2} (1, \pm 1), \frac{1+\varepsilon}{2} (\pm 1, 1) \right\}. \quad (\text{D.1})$$

In formulae, one can see that

$$N_t(x, y) = |x| + |y| - \frac{2t}{1+t} \min\{|x|, |y|\}. \quad (\text{D.2})$$

The unit ball of the dual norm is another octagon:

$$B_{N_{t^*}} := \text{co} \left\{ \left(\pm 1, \frac{1-t}{1+t} \right), \left(\frac{1-t}{1+t}, \pm 1 \right), \left(\pm 1, -\frac{1-t}{1+t} \right), \left(-\frac{1-t}{1+t}, \pm 1 \right) \right\}. \quad (\text{D.3})$$

In the following, we will denote the injective and projective norms on $\mathbb{R}^2 \otimes \mathbb{R}^2 \simeq \mathcal{M}_2(\mathbb{R})$ corresponding to N_t by $\|\cdot\|_{\varepsilon(t)}$ and $\|\cdot\|_{\pi(t)}$. From the comparison of the unit balls it follows easily that

$$N_t \leq N_0 \leq (1+t)^2 N_t. \quad (\text{D.4})$$

For

$$Z = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathcal{M}_2(\mathbb{R}), \quad (\text{D.5})$$

it can be easily seen that

$$\|Z\|_{\pi(0)} = |Z|_1 = |a| + |b| + |c| + |d|. \quad (\text{D.6})$$

A way to see this is by computing the dual norm $\|H\|_{\pi(0)^*} = \|H\|_{*\varepsilon(0)}$, which is the injective norm induced by $|\cdot|_\infty$ and therefore coincides with the maximum entry of the matrix H . The injective norm is slightly less transparent, but it can be seen that

$$\|Z\|_{\varepsilon(0)} = \begin{cases} |Z|_1 & \text{if } abcd \geq 0, \\ |Z|_1 - 2 \min\{|a|, |b|, |c|, |d|\} & \text{if } abcd < 0. \end{cases} \quad (\text{D.7})$$

For what follows it will be useful to define

$$p(Z) = \begin{cases} 0 & \text{if } abcd \geq 0, \\ 4 \frac{\min\{|a|, |b|, |c|, |d|\}}{|a|+|b|+|c|+|d|} & \text{if } abcd < 0. \end{cases} \quad (\text{D.8})$$

Thanks to this definition, we can write

$$r_0(Z) = \frac{\|Z\|_{\pi(0)}}{\|Z\|_{\varepsilon(0)}} = \frac{1}{1 - p(Z)/2} \quad (\text{D.9})$$

for $t = 0$. Incidentally, it is now clear that in this case there are matrices Z such that $r_0(Z) = 2$, an example of which is the Hadamard matrix $\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$. We have to perturb the model to establish tighter upper bounds on $r_t(Z)$. In general, the comparison with (D.4) yields

$$r_t(Z) \leq (1+t)^2 r_0(Z) = \frac{(1+t)^2}{1 - p(Z)/2} \quad (\text{D.10})$$

for all $t \in [0, 1]$. This is our first bound on the projective-injective ratio. Unfortunately, this upper bound is again useless when Z is sufficiently close to the Hadamard matrix we discussed above (or to some equivalent version). Therefore, our strategy is to come up with another bound specifically targeted on this matrix. Clearly, such a bound will work only for $t > 0$.

Consider a matrix Z as in (D.5). Up to operations that leave the norm N_t invariant, we can assume $a, b, c \geq 0$ and $|d| \leq a, b, c$. Consider first the case $d < 0$. Then (D.8) becomes $p(Z) = 4|d|/|Z|_1$. Taking as an ansatz in the supremum appearing in (2.1) two times the first functional in the list (D.3),

we see that

$$\begin{aligned}
\|Z\|_{\varepsilon(t)} &\geq \left(1 - \frac{1-t}{1+t}\right) \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} 1 \\ \frac{1-t}{1+t} \end{pmatrix} \\
&= a + b + c + d - \frac{2t}{1+t} (b + c + 2d) + \frac{4t^2}{(1+t)^2} d \\
&\geq a + b + c + d - \frac{2t}{1+t} (a + b + c + 3d) + \frac{4t^2}{(1+t)^2} d \\
&= |a| + |b| + |c| + |d| - 2|d| - \frac{2t}{1+t} (|a| + |b| + |c| + |d| - 4|d|) \\
&\quad - \frac{4t^2}{(1+t)^2} |d|.
\end{aligned}$$

Dividing by $|Z|_1$ yields

$$\frac{\|Z\|_{\varepsilon(t)}}{|Z|_1} \geq 1 - \frac{p(Z)}{2} - \frac{2t}{1+t} (1 - p(Z)) - \frac{t^2}{(1+t)^2} p(Z). \quad (\text{D.11})$$

A posteriori, it is easily seen that (D.11) is a fortiori true when $d \geq 0$ and hence $p(Z) = 0$. In fact, in this case one sees that

$$\begin{aligned}
\frac{\|Z\|_{\varepsilon(t)}}{|Z|_1} &= \frac{a + b + c - \frac{2t}{1+t}(b + c) + \left(\frac{1-t}{1+t}\right)^2 d}{|Z|_1} \\
&= 1 - \frac{2t}{1+t} + \frac{d}{|Z|_1} \left(-1 + \frac{4t}{1+t} + \left(\frac{1-t}{1+t}\right)^2 \right) \\
&= 1 - \frac{2t}{1+t} + \frac{4t^2}{(1+t)^2} \frac{d}{|Z|_1} \\
&\geq 1 - \frac{2t}{1+t}.
\end{aligned}$$

Now, in order to find an upper bound on $\|Z\|_{\pi(t)}$, we employ the definition (2.2) together with the expression (D.2) for the norm N_t . Write

$$\begin{aligned}
\|Z\|_{\pi(t)} &= \|(1, 0) \otimes (a, b) + (0, 1) \otimes (c, d)\|_{\pi(t)} \\
&\leq N_t(1, 0)N_t(a, b) + N_t(0, 1)N_t(c, d) \\
&= N_t(a, b) + N_t(c, d) \\
&= |a| + |b| + |c| + |d| - \frac{2t}{1+t} (\min\{|a|, |b|\} + \min\{|c|, |d|\}) \\
&\leq |Z|_1 - \frac{4t}{1+t} \min\{|a|, |b|, |c|, |d|\}.
\end{aligned}$$

Finally, using also (D.8) we obtain

$$\frac{\|Z\|_{\pi(t)}}{|Z|_1} \leq 1 - \frac{t}{1+t} p(Z). \quad (\text{D.12})$$

Putting together (D.11) and (D.12), we deduce

$$\begin{aligned} r_t(Z) &\leq \frac{1 - \frac{t}{1+t} p(Z)}{1 - \frac{p(Z)}{2} - \frac{2t}{1+t} (1 - p(Z)) - \frac{t^2}{(1+t)^2} p(Z)} \\ &= \frac{1 + t(1 - p(Z))}{1 - t - \frac{1-2t-t^2}{2(1+t)} p(Z)}. \end{aligned} \quad (\text{D.13})$$

This together with (D.10) yields

$$r_t(Z) \leq \min \left\{ \frac{(1+t)^2}{1 - \frac{p(Z)}{2}}, \frac{1 + t(1 - p(Z))}{1 - t - \frac{1-2t-t^2}{2(1+t)} p(Z)} \right\}. \quad (\text{D.14})$$

By definition of $S(2)$ we see that

$$S(2) \leq \min_{0 \leq t \leq 1} \max_{0 \leq p \leq 1} \min \left\{ \frac{(1+t)^2}{1 - \frac{p}{2}}, \frac{1 + t(1 - p)}{1 - t - \frac{1-2t-t^2}{2(1+t)} p} \right\} = \sqrt{3}, \quad (\text{D.15})$$

where the last step follows from some elementary considerations that we do not report. The t achieving the above minimum is $2 - \sqrt{3}$. \square

Remark. The above upper bound $\sqrt{3}$ is actually achieved by this model when $t = 2 - \sqrt{3}$. Even more is true, that is, for all $t \in [0, 2 - \sqrt{3}]$ the function on the right-hand side of (D.14) is maximal for $p(Z) = 1$, and the subsequent upper bound

$$r_t(Z) \leq \frac{2(1+t)}{1 + 2t - t^2} \quad (\text{D.16})$$

is tight and achieved by $Z = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$. In fact, on the one hand it is easy to convince ourselves that

$$\begin{aligned} \left\| \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \right\|_{\varepsilon(t)} &= \left(1 \quad \frac{1-t}{1+t} \right) \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} 1 \\ \frac{1-t}{1+t} \end{pmatrix} \\ &= \frac{2(1 + 2t - t^2)}{(1+t)^2}. \end{aligned}$$

On the other hand,

$$\begin{aligned} \left\| \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \right\|_{\pi(t)} &= \left\| \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \right\|_{*\varepsilon(t)} \\ &\geq \left(\frac{1+t}{2} \quad \frac{1+t}{2} \right) \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} \\ &= 1 + t, \end{aligned}$$

and this is actually an equality. Putting all together we see that

$$r_t \left(\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \right) = \frac{2(1+t)}{1 + 2t - t^2},$$

which in turn for $t \in [0, 2 - \sqrt{3}]$ corresponds to the maximum of the right-hand side of (D.14) over $0 \leq p(Z) \leq 1$ and reduces to $\sqrt{3}$ when $t = 2 - \sqrt{3}$.

Appendix E

Quantum data hiding with Werner states

Throughout this appendix, we briefly review the well-known techniques used to find lower bounds on data hiding ratios via Werner states, and apply them to prove (3.23) and (3.27). More precisely, we will show that those described in (3.23) and (3.27) are the optimal data hiding pairs within the Werner class, defined as the real span of symmetric and antisymmetric projector (3.22). Formally, this amounts to show that

$$\max_{(\alpha, \beta) \neq (0, 0)} \frac{\|\alpha \rho_S + \beta \rho_A\|_1}{\|\alpha \rho_S + \beta \rho_A\|_{\text{SEP}}} = n, \quad (\text{E.1})$$

$$\max_{(\alpha, \beta) \neq (0, 0)} \frac{\|\alpha \rho_S + \beta \rho_A\|_W}{\|\alpha \rho_S + \beta \rho_A\|_{\text{SEP}}} = 2n - 1. \quad (\text{E.2})$$

In other words, both maxima are achieved when $\alpha = n+1$, $\beta = -(n+1)$, which by the way makes $\alpha \rho_S + \beta \rho_A \propto F$. As a side remark, let us notice here that the above separability norms could be easily replaced with LOCC norms, since the extremal measurements in the Werner class are well-known to be LOCC [156, Propositions 1 and 3].

Consider a binary measurement $(E, \mathbb{1} - E)$, which we will choose later to be either standard, or a witness, or else an LOCC. Thanks to the fact that to estimate distinguishability norms we are going to compute quantities of the form $\text{Tr} EZ$ where Z is in the Werner class and thus invariant under the ‘twirling’ operation $\mathcal{T}(\cdot) := \int dU U \otimes U(\cdot)U^\dagger \otimes U^\dagger$, a very standard trick going back to the first paper on data hiding [9] allows us to conclude that we can suppose also the measurement elements $E, \mathbb{1} - E$ to belong to the Werner class. For $E = a\mathbb{1} + bF$, it is well-known that the measurement $(E, \mathbb{1} - E)$ is:

- a standard quantum mechanical measurement iff $0 \leq E \leq \mathbb{1}$, i.e. iff $0 \leq a \pm b \leq 1$;
- separable iff $E, \mathbb{1} - E$ are both separable operators (2.34), iff $0 \leq a - b \leq 1$ and $0 \leq a + nb \leq 1$;

- a W -theory measurement iff $E, \mathbb{1} - E$ are both witnesses (2.35), iff $0 \leq a \leq 1$ and $0 \leq a + b \leq 1$.

While the first two points are well-known results [130], the latter deserves a quick comment. An operator E is a witness iff $\langle \alpha\beta | E | \alpha\beta \rangle \geq 0$ for all states $|\alpha\rangle, |\beta\rangle \in \mathbb{C}^n$. We can thus use the properties of the flip operator to conclude that $E = a\mathbb{1} + bF$ is a witness iff $a + pb \geq 0$ for all $p \in [0, 1]$, which together with the analogous condition for $\mathbb{1} - E$ reproduced the above constraints. Optimising over all measurements in the Werner class we conclude that

$$\|\alpha\rho_S + \beta\rho_A\|_1 = |\alpha| + |\beta|, \quad (\text{E.3})$$

$$\|\alpha\rho_S + \beta\rho_A\|_{\text{SEP}} = \frac{2}{n+1} |\alpha| + \left| \frac{n-1}{n+1} \alpha + \beta \right|, \quad (\text{E.4})$$

$$\|\alpha\rho_S + \beta\rho_A\|_W = |\alpha - \beta| + 2|\beta|. \quad (\text{E.5})$$

Computing the maxima in (E.1) and showing that they are achieved for $\alpha = n + 1, \beta = -(n + 1)$ is now an elementary exercise.

Appendix F

Comparing $\|\cdot\|_W$ and $\|\cdot\|_2$

Here, we argue that the methods in [151] can not be applied directly to determine the data hiding ratio in W -theory against any locally constrained set of measurements. As detailed in Subsection 3.3.2, to this purpose it is enough to show the following.

Lemma F.1. *For all $\delta > 0$, there exists a sequence of operators Z_n acting on $\mathbb{C}^n \otimes \mathbb{C}^n$, such that $\|Z_n\|_2 = 1$ but $\|Z_n\|_W \geq \Omega(n^{3/2-\delta})$, where $\|\cdot\|_W$ is defined in (2.37).*

Proof. From the results of [282] it is known that a random orthogonal projector Π onto a subspace of $\mathbb{C}^n \otimes \mathbb{C}^n$ of dimension $k \gg n$ will satisfy

$$\langle \alpha\beta | \Pi | \alpha\beta \rangle \leq \frac{2k}{n^2} \quad \forall |\alpha\rangle, |\beta\rangle \in \mathbb{C}^n : \langle \alpha | \alpha \rangle = \langle \beta | \beta \rangle = 1 \quad (\text{F.1})$$

with high probability as n tends to infinity. Fixing $k = n^{1+2\delta}$ and picking one such Π_n for all $n \in \mathbb{N}$, we can construct $Z_n := n^{-1/2-\delta} \Pi_n$. By definition, $\|Z_n\|_2 = 1$ for all $n \in \mathbb{N}$. Moreover, defining $H := \mathbb{1} - \frac{n^2}{k} \Pi_n$, by (F.1) we have $-1 \leq \langle \alpha\beta | H | \alpha\beta \rangle \leq 1$ for all normalised $|\alpha\rangle, |\beta\rangle$, i.e. H belongs to the unit ball of the dual base norm of the GPT $\text{QM}_{\min}^{\otimes} \text{QM}_n$. Hence,

$$\|Z_n\|_W \geq \text{Tr } Z_n H = \frac{n^2 - k}{\sqrt{k}} = \Omega(n^{3/2-\delta}),$$

which yields the claim. \square

Appendix G

Additional remarks on some results of Chapter 6

This appendix complements Chapter 6. Here we complete the proof of Theorem 6.13(b) and extend Theorem 6.14 to include the case of the underlying entropy being a Rényi- α entropy with $\alpha \geq 2$.

G.1 Gaussian steerability is not monogamous with respect to the steering party

We present a counterexample showing that already in the simplest case $k = 2$, $n_A = 2$, $n_{B_1} = n_{B_2} = 1$, there exist mixed states violating inequality (6.48). This completes the proof of Theorem 6.13(b). A QCM that does the job, as found by numerical search, is as follows:

$$V_{AB_1B_2} = \begin{pmatrix} 1.2 & -0.3 & 0.4 & -2.7 & 1.8 & -1.9 & 0.4 & -0.1 \\ -0.3 & 0.9 & -1.2 & 0.4 & -1.2 & 0.5 & -0.4 & 0.1 \\ 0.4 & -1.2 & 4.5 & 1.6 & -1.4 & 1.8 & -0.1 & -0.3 \\ -2.7 & 0.4 & 1.6 & 12. & -9.5 & 10.1 & -1.4 & -0.3 \\ 1.8 & -1.2 & -1.4 & -9.5 & 11.9 & -11.5 & 1.6 & 0.8 \\ -1.9 & 0.5 & 1.8 & 10.1 & -11.5 & 11.9 & -1. & -1.4 \\ 0.4 & -0.4 & -0.1 & -1.4 & 1.6 & -1. & 2.4 & -2. \\ -0.1 & 0.1 & -0.3 & -0.3 & 0.8 & -1.4 & -2. & 2.8 \end{pmatrix}. \quad (\text{G.1})$$

Here, the first four rows and columns pertain to A , the fifth and sixth to B_1 , the last two to B_2 . It can be easily verified that the minimum symplectic eigenvalue of the above matrix with respect to the symplectic form $\Omega_A \oplus \Omega_{B_1} \oplus \Omega_{B_2}$ is $\nu_{\min}(V_{B_1B_2A}) = 1.01359$, so that $V_{B_1B_2A}$ is a legitimate QCM (obeying (5.29)). However,

$$\mathcal{G}(B_1B_2 \rangle A)_V - \mathcal{G}(B_1 \rangle A)_V - \mathcal{G}(B_2 \rangle A)_V = -0.816863 < 0, \quad (\text{G.2})$$

violating (6.48) as claimed.

G.2 Rényi extension of Theorem 6.14

This last part of the present appendix is devoted to discussing possible Rényi generalisations of (6.61). In fact, by looking at the proof Theorem 6.14, one could wonder what makes the Rényi-2 entropy special in this context. It turns out that for all Rényi- α entropies with $\alpha \geq 1$ (included the von Neumann one) we can always provide the upper bound

$$2 E_{F,\alpha}^G(A : B)_V \leq S_\alpha(V_A \# (\Omega_A(V_{AB}/V_B)^{-1} \Omega_A^T)), \quad (\text{G.3})$$

where the function $S_\alpha(V)$, defined in (6.53), gives the Rényi- α entropy of a Gaussian state with QCM V . However, the crucial inequality

$$S_\alpha(M \# N) \leq \frac{1}{2} S_\alpha(M) + \frac{1}{2} S_\alpha(N), \quad (\text{G.4})$$

that we used in the proof of Theorem 6.14 to simplify further the right-hand side of (G.3), breaks down for $\alpha < 2$. In particular, it can be violated for $\alpha = 1$. On the contrary, $I_\alpha(A : B) \geq 2 E_{F,\alpha}^G(A : B)$ is always true as long as $\alpha \geq 2$, as the next Lemma clarifies.

Lemma G.1. *Fix an integer $n \geq 1$. The inequality $S_\alpha(M \# N) \leq \frac{1}{2} S_\alpha(M) + \frac{1}{2} S_\alpha(N)$ holds for all $2n \times 2n$ real matrices $M, N > 0$ if and only if $\alpha \geq 2$.*

Proof. We claim that inequality (G.4) is equivalent to the convexity of the function

$$f_\alpha(x) := -\frac{1}{\alpha-1} \log \frac{2^\alpha}{(e^x+1)^\alpha - (e^x-1)^\alpha} \quad (\text{G.5})$$

defined on \mathbb{R}_+ , where conformally to (6.54) one defines

$$f_1(x) := \frac{e^x+1}{2} \log \frac{e^x+1}{2} - \frac{e^x-1}{2} \log \frac{e^x-1}{2}. \quad (\text{G.6})$$

In fact, on the one hand choosing $M = e^x \mathbf{1}$ and $N = e^y \mathbf{1}$ yields

$$S_\alpha(M \# N) = n f_\alpha\left(\frac{x+y}{2}\right),$$

$$\frac{1}{2} S_\alpha(M) + \frac{1}{2} S_\alpha(N) = \frac{n}{2} f_\alpha(x) + \frac{n}{2} f_\alpha(y),$$

so that f_α is necessarily convex when (G.4) holds. On the other hand, suppose that f_α is convex. From [252, Theorem 3] we learn that $\log \hat{\nu}(M \# N) \prec \frac{1}{2} \log \hat{\nu}(M) + \frac{1}{2} \log \hat{\nu}(N)$, where $\hat{\nu}(M) \in \mathbb{R}_+^{2n}$ is obtained by listing the symplectic eigenvalues of M each repeated twice and sorting the entries of the resulting vector in descending order, the logarithm of vectors is intended entrywise, and the symbol \prec denotes *majorization* [299, II]. What the above relation tells us is that the symplectic spectrum of the geometric mean is in a precise sense more disordered than the geometric mean of the two spectra. It is elementary to verify that whenever $x \prec y$ and f is convex, $\sum_{i=1}^n f(x_i) \leq \sum_{i=1}^n f(y_i)$ holds

true [299, Corollary II.3.4]. Choosing as f the function in (G.5) and observing that f_α is always monotonically increasing, we obtain

$$\begin{aligned} S_\alpha(M\#N) &= \sum_{i=1}^n f_\alpha(\log \nu_i(M\#N)) \\ &= \frac{1}{2} \sum_{i=1}^{2n} f_\alpha(\log \hat{\nu}_i(M\#N)) \\ &\leq \frac{1}{2} \sum_{i=1}^{2n} f_\alpha\left(\frac{1}{2} \log \hat{\nu}_i(M) + \frac{1}{2} \log \hat{\nu}_i(N)\right) \\ &\leq \frac{1}{2} \sum_{i=1}^{2n} \frac{1}{2} (f_\alpha(\log \hat{\nu}_i(M)) + f_\alpha(\log \hat{\nu}_i(N))) \\ &= \frac{1}{2} S_\alpha(M) + \frac{1}{2} S_\alpha(N). \end{aligned}$$

Now, the main claim will follow once we show that f_α defined via (G.5) is convex if and only if $\alpha \geq 2$. We can restrict our analysis to the case $\alpha > 1$ since the function in (G.6) is elementarily seen to be non-convex (actually, concave). Some tedious algebra leads us to the following expression for the second derivative of that function:

$$\begin{aligned} f''_\alpha(x) &= \frac{\alpha}{\alpha - 1} \frac{\cosh^\alpha(x/2) \sinh^\alpha(x/2)}{\sinh^2 x (\cosh^\alpha(x/2) - \sinh^\alpha(x/2))} \\ &\quad \cdot (\cosh^\alpha(x/2) \sinh^{2-\alpha}(x/2) - \sinh^\alpha(x/2) \cosh^{2-\alpha}(x/2) - \alpha + 1). \end{aligned} \tag{G.7}$$

Since everything else in the above expression is positive, we have only to prove that $\cosh^\alpha(x/2) \sinh^{2-\alpha}(x/2) - \sinh^\alpha(x/2) \cosh^{2-\alpha}(x/2) \geq \alpha - 1$ for all $x \geq 0$ if and only if $\alpha \geq 2$. That $\alpha \geq 2$ is necessary can be seen by taking the limit $x \rightarrow 0^+$. Conversely, if $\alpha = 2 + \delta$ with $\delta \geq 0$ one gets

$$\begin{aligned} &\cosh^\alpha(x/2) \sinh^{2-\alpha}(x/2) - \sinh^\alpha(x/2) \cosh^{2-\alpha}(x/2) \\ &= \cosh^2(x/2) \tanh^{-\delta}(x/2) - \sinh^2(x/2) \tanh^\delta(x/2) \\ &= \frac{1}{1-t^2} t^{-\delta} - \frac{t^2}{1-t^2} t^\delta \\ &=: \varphi_t(\delta), \end{aligned}$$

where we defined $t := \tanh(x/2)$. It is not difficult to see that the function $\varphi_t(\delta)$ is convex in δ since

$$\varphi''_t(\delta) = \frac{t^{-\delta} (1 - t^{2\delta+2}) \log^2(t)}{1 - t^2} \geq 0.$$

From this fact we deduce that

$$\varphi_t(\delta) \geq \varphi_t(0) + \delta \varphi'_t(0) = 1 + \delta = \alpha - 1,$$

implying that (G.7) is positive and hence that f_α defined by (G.5) is convex for all $\alpha \geq 2$, as claimed. \square

Bibliography

- [1] Dante Alighieri. *La Divina Commedia*.
- [2] A. Einstein, B. Podolsky, and N. Rosen. Can quantum-mechanical description of physical reality be considered complete? *Phys. Rev.*, 47:777–780, 1935.
- [3] J.S. Bell. On the Einstein-Podolsky-Rosen paradox. *Physics*, 1(3):195–200, 1964.
- [4] N. Bohr. Can quantum-mechanical description of physical reality be considered complete? *Phys. Rev.*, 48:696–702, 1935.
- [5] E. Schrödinger. Discussion of probability relations between separated systems. In *Math. Proc. Cambridge Philos. Soc.*, volume 31, pages 555–563. Cambridge University Press, 1935.
- [6] A. Aspect, J. Dalibard, and G. Roger. Experimental test of Bell’s inequalities using time-varying analyzers. *Phys. Rev. Lett.*, 49:1804–1807, 1982.
- [7] G. Ludwig. *An Axiomatic Basis for Quantum Mechanics: Derivation of Hilbert space structure*, volume 1. Springer-Verlag, 1985.
- [8] I. Namioka and R.R. Phelps. Tensor products of compact convex sets. *Pacific J. Math.*, 31(2):469–480, 1969.
- [9] B.M. Terhal, D.P. DiVincenzo, and D.W. Leung. Hiding bits in Bell states. *Phys. Rev. Lett.*, 86:5807–5810, 2001.
- [10] D.P. DiVincenzo, D.W. Leung, and B.M. Terhal. Quantum data hiding. *IEEE Trans. Inf. Theory*, 48(3):580–598, 2002.
- [11] M. Christandl and A. Winter. Squashed entanglement: An additive entanglement measure. *J. Math. Phys.*, 45(3):829–840, 2004.
- [12] S.N. Filippov and M. Ziman. Bipartite entanglement-annihilating maps: Necessary and sufficient conditions. *Phys. Rev. A*, 88(3):032316, 2013.
- [13] Sergey N Filippov. Ppt-inducing, distillation-prohibiting, and entanglement-binding quantum channels. *J. Russ. Laser Res.*, 35(5):484–491, 2014.
- [14] R. Simon. Peres-Horodecki separability criterion for continuous variable

- systems. *Phys. Rev. Lett.*, 84:2726–2729, 2000.
- [15] R.F. Werner and M.M. Wolf. Bound entangled Gaussian states. *Phys. Rev. Lett.*, 86:3658–3661, 2001.
- [16] M.M. Wolf, J. Eisert, and M.B. Plenio. Entangling power of passive optical elements. *Phys. Rev. Lett.*, 90:047904, 2003.
- [17] I. Kogias, A.R. Lee, S. Ragy, and G. Adesso. Quantification of Gaussian quantum steering. *Phys. Rev. Lett.*, 114:060403, 2015.
- [18] N. Li and S. Luo. Total versus quantum correlations in quantum states. *Phys. Rev. A*, 76:032327, 2007.
- [19] G. Mackey. *Mathematical Foundations of Quantum Mechanics*. Benjamin, 1963.
- [20] G. Ludwig. Versuch einer axiomatischen Grundlegung der Quantenmechanik und allgemeinerer physikalischer Theorien. *Z. Phys.*, 181(3):233–260, 1964.
- [21] G. Ludwig. Attempt of an axiomatic foundation of quantum mechanics and more general theories II. *Commun. Math. Phys.*, 4(5):331–348, 1967.
- [22] G. Ludwig. Attempt of an axiomatic foundation of quantum mechanics and more general theories III. *Commun. Math. Phys.*, 9(1):1–12, 1968.
- [23] G. Dähn. Attempt of an axiomatic foundation of quantum mechanics and more general theories IV. *Commun. Math. Phys.*, 9(3):192–211, 1968.
- [24] P. Stolz. Attempt of an axiomatic foundation of quantum mechanics and more general theories V. *Commun. Math. Phys.*, 11(4):303–313, 1969.
- [25] P. Stolz. Attempt of an axiomatic foundation of quantum mechanics and more general theories VI. *Commun. Math. Phys.*, 23(2):117–126, 1971.
- [26] G. Ludwig. An improved formulation of some theorems and axioms in the axiomatic foundation of the Hilbert space structure of quantum mechanics. *Comm. Math. Phys.*, 26(1):78–86, 1972.
- [27] G. Ludwig. *Deutung des Begriffs “physikalische Theorie” und axiomatische Grundlegung der Hilbertraumstruktur der Quantenmechanik durch Hauptsätze des Messens*. Springer Berlin Heidelberg, 1970.
- [28] E.B. Davies and J.T. Lewis. An operational approach to quantum probability. *Comm. Math. Phys.*, 17(3):239–260, 1970.
- [29] C.M. Edwards. The operational approach to algebraic quantum theory I. *Commun. Math. Phys.*, 16(3):207–230, 1970.
- [30] C.M. Edwards and M.A. Gerzon. Monotone convergence in partially ordered vector spaces. *Ann. Inst. H. Poincaré Phys. Théor.*, 12(4):323–328, 1970.
- [31] C.M. Edwards. Sets of simple observables in the operational approach to quantum theory. *Ann. Inst. H. Poincaré Phys. Théor.*, 15(1):1–14, 1971.

- [32] S. Gudder. Convex structures and operational quantum mechanics. *Comm. Math. Phys.*, 29(3):249–264, 1973.
- [33] C.M. Edwards. Classes of operations in quantum theory. *Comm. Math. Phys.*, 20(1):26–56, 1971.
- [34] C.M. Edwards. The theory of pure operations. *Comm. Math. Phys.*, 24(4):260–288, 1972.
- [35] B. Mielnik. Theory of filters. *Comm. Math. Phys.*, 15(1):1–46, 1969.
- [36] G. Dähn. The algebra generated by physical filters. *Comm. Math. Phys.*, 28(2):109–122, 1972.
- [37] G. Dähn. Symmetry of the physical probability function implies modularity of the lattice of decision effects. *Comm. Math. Phys.*, 28(2):123–132, 1972.
- [38] B. Mielnik. Generalized quantum mechanics. *Comm. Math. Phys.*, 37(3):221–256, 1974.
- [39] M. Kläy, C. Randall, and D. Foulis. Tensor products and probability weights. *International Journal of Theoretical Physics*, 26(3):199–219, 1987.
- [40] A. Wilce. Tensor products in generalized measure theory. *Int. J. Theor. Phys.*, 1992.
- [41] H. Barnum, J. Barrett, M. Leifer, and A. Wilce. Teleportation in general probabilistic theories. In *Proc. Sympos. Appl. Math.*, volume 71, pages 25–48, 2012.
- [42] H.H. Schaefer. *Topological Vector Spaces*. Springer-Verlag New York, 1971.
- [43] A.L. Peressini. *Ordered topological vector spaces*. Harper & Row New York, 1967.
- [44] G.J.O. Jameson. *Ordered linear spaces*. Number 141 in Lecture notes in mathematics. Springer-Verlag, 1970.
- [45] A.J. Ellis. The duality of partially ordered normed linear spaces. *J. London Math. Soc.*, 1(1):730–744, 1964.
- [46] A.J. Ellis. Linear operators in partially ordered normed vector spaces. *J. London Math. Soc.*, 1(1):323–332, 1966.
- [47] D.A. Edwards. On the homeomorphic affine embedding of a locally compact cone into a Banach dual space endowed with the vague topology. *Proc. London Math. Soc.*, 3(3):399–414, 1964.
- [48] S. Popescu and D. Rohrlich. Quantum nonlocality as an axiom. *Found. Phys.*, 24(3):379–385, 1994.
- [49] J. Barrett, N. Linden, S. Massar, S. Pironio, S. Popescu, and D. Roberts. Nonlocal correlations as an information-theoretic resource. *Phys. Rev. A*,

- 71:022101, 2005.
- [50] J. Barrett. Information processing in generalized probabilistic theories. *Phys. Rev. A*, 75(3):032304, 2007.
 - [51] P. Janotta, C. Gogolin, J. Barrett, and N. Brunner. Limits on nonlocal correlations from the structure of the local state space. *New J. Phys.*, 13(6):063024, 2011.
 - [52] W. van Dam. Implausible consequences of superstrong nonlocality. *Nat. Comput.*, 12(1):9–12, 2013.
 - [53] G. Brassard, H. Buhrman, N. Linden, A.A. Méthot, A. Tapp, and F. Unger. Limit on nonlocality in any world in which communication complexity is not trivial. *Phys. Rev. Lett.*, 96:250401, 2006.
 - [54] N. Linden, S. Popescu, A.J. Short, and A. Winter. Quantum nonlocality and beyond: Limits from nonlocal computation. *Phys. Rev. Lett.*, 99:180502, 2007.
 - [55] C.M. Lee and J. Barrett. Computation in generalised probabilistic theories. *New J. Phys.*, 17(8):083001, 2015.
 - [56] C. Pfister and S. Wehner. An information-theoretic principle implies that any discrete physical theory is classical. *Nat. Commun.*, 4:1851, 2013.
 - [57] C. Pfister. One simple postulate implies that every polytopic state space is classical. *Preprint arXiv:1203.5622*.
 - [58] G. Chiribella, G.M. D’Ariano, and P. Perinotti. Probabilistic theories with purification. *Phys. Rev. A*, 81:062348, 2010.
 - [59] H. Barnum, J. Barrett, M. Leifer, and A. Wilce. Generalized no-broadcasting theorem. *Phys. Rev. Lett.*, 99(24):240501, 2007.
 - [60] H. Barnum and A. Wilce. Information processing in convex operational theories. *Electron. Notes Theor. Comput. Sci.*, 270(1):3–15, 2011.
 - [61] H. Barnum, C.P. Gaebler, and A. Wilce. Ensemble steering, weak self-duality, and the structure of probabilistic theories. *Preprint arXiv:0912.5532*, 2009.
 - [62] H. Barnum, J. Barrett, L.O. Clark, M. Leifer, R. Spekkens, N. Stepanik, A. Wilce, and R. Wilke. Entropy and information causality in general probabilistic theories. *New J. Phys.*, 12(3):033024, 2010.
 - [63] A. Jenčová. Non-classical features in general probabilistic theories. *Preprint arXiv:1705.08008*, 2017.
 - [64] G. Chiribella and C.M. Scandolo. Entanglement and thermodynamics in general probabilistic theories. *New J. Phys.*, 17(10):103027, 2015.
 - [65] G. Ludwig. *Foundations of quantum mechanics*, volume 1. Springer-Verlag, 1983.
 - [66] J. Gunson. On the algebraic structure of quantum mechanics. *Comm.*

- Math. Phys.*, 6(4):262–285, 1967.
- [67] L. Hardy. Quantum theory from five reasonable axioms. *Preprint arXiv:quant-ph/0101012*, 2001.
- [68] R. Schack. Quantum theory from four of Hardy’s axioms. *Found. Phys.*, 2003.
- [69] L. Masanes and M.P. Müller. A derivation of quantum theory from physical requirements. *New J. Phys.*, 13(6):063001, 2011.
- [70] L. Masanes, M.P. Müller, R. Augusiak, and D. Pérez-García. Existence of an information unit as a postulate of quantum theory. *Proc. Natl. Acad. Sci. USA*, 110(41):16373–16377, 2013.
- [71] G. Chiribella, G.M. D’Ariano, and P. Perinotti. Informational derivation of quantum theory. *Phys. Rev. A*, 84:012311, 2011.
- [72] G.M. D’Ariano. On the missing axiom of quantum mechanics. *AIP Conf. Proc.*, 810(1):114–130, 2006.
- [73] G.M. D’Ariano. How to derive the Hilbert space formulation of quantum mechanics from purely operational axioms. *AIP Conf. Proc.*, 844(1):101–128, 2006.
- [74] A. Wilce. Four and a half axioms for finite dimensional quantum mechanics. *Preprint arXiv:0912.5530*, 2009.
- [75] H. Barnum, M.P. Müller, and C. Ududec. Higher-order interference and single-system postulates characterizing quantum theory. *New J. Phys.*, 16(12):123029, 2014.
- [76] A. Hartkämper and H. Neumann. *Foundations of Quantum Mechanics and Ordered Linear Spaces: Advanced Study Institute held in Marburg 1973*. Springer Berlin Heidelberg, 1974.
- [77] W. Rudin. *Functional Analysis*. McGraw-Hill, 1991.
- [78] J.L. Kelley. *General Topology*. Springer New York, 1975.
- [79] R.E. Megginson. *An introduction to Banach space theory*. Number 183 in Graduate Texts in Mathematics. Springer Science & Business Media, 2012.
- [80] N. Bourbaki. *Topologie Générale: Chapitres 1 à 4*. Springer Berlin Heidelberg, 2007.
- [81] J.L. Kelley and I. Namioka. *Linear topological spaces*. Springer-Verlag, 1976.
- [82] G.F. Simmons. *Introduction to Topology and Modern Analysis*. McGraw-Hill, 1963.
- [83] T. Tao. 245B, Notes 9: The Baire category theorem and its Banach space consequences. <https://terrytao.wordpress.com>, 2009.
- [84] E.M.V. Sereno and R.V. Mendoza. A nonstandard proof of the Banach-

- Steinhaus theorem. *Rend. Istit. Mat. Univ. Trieste*, 40:185–188, 2009.
- [85] K.F. Ng. On a computation rule for polars. *Math. Scand.*, 26(1):14–16, 1970.
- [86] J.W Tukey. Some notes on the separation of convex sets. *Port. Math.*, 3(2):95–102, 1942.
- [87] V.L. Klee Jr. Convex sets in linear spaces. *Duke Math. J.*, 18(2):443–466, 1951.
- [88] J. Grosberg and M.G. Krein. Sur la decomposition des fonctionnelles en composantes positives. *Doklady Akad. Nauk S.S.S.R. (N.S.)*, 25:723–726, 1939.
- [89] M. Krein. Sur la decomposition minimale d’une fonctionnelle lineaire en composantes positives. *Doklady Akad. Nauk S.S.S.R. (N.S.)*, 28:18–22, 1940.
- [90] V.L. Klee. Extremal structure of convex sets. *Arch. Math.*, 8(3):234–240, 1957.
- [91] R.F. Werner. Quantum states with Einstein-Podolsky-Rosen correlations admitting a hidden-variable model. *Phys. Rev. A*, 40:4277–4281, 1989.
- [92] R. Horodecki, P. Horodecki, M. Horodecki, and K. Horodecki. Quantum entanglement. *Rev. Mod. Phys.*, 81:865–942, 2009.
- [93] H. Barnum. Quantum information processing, operational quantum logic, convexity, and the foundations of physics. *Stud. Hist. Philos. Sci. B*, 34(3):343–379, 2003.
- [94] S.L. Woronowicz. Positive maps of low dimensional matrix algebras. *Rep. Math. Phys.*, 10(2):165–183, 1976.
- [95] D. Cariello. Separability for weakly irreducible matrices. *Quantum Inf. Comput.*, 14(15-16).
- [96] D. Cariello. Does symmetry imply PPT property? *Quantum Inf. Comput.*, 15(9-10):812–824, 2015.
- [97] A. Defant and K. Floret. *Tensor norms and operator ideals*, volume 176. Elsevier, 1992.
- [98] R.A. Ryan. *Introduction to tensor products of Banach spaces*. Springer Science & Business Media, 2013.
- [99] L. Lami, C. Palazuelos, and A. Winter. Ultimate data hiding in quantum mechanics and beyond. *Preprint arXiv:1703.03392*, 2017.
- [100] B. Simon. *Convexity: An Analytic Viewpoint*. Cambridge Tracts in Mathematics. Cambridge University Press, 2011.
- [101] H. Minkowski. *Geometrie der Zahlen*, 1910.
- [102] C. Carathéodory. Über den Variabilitätsbereich der Koeffizienten von Potenzreihen, die gegebene Werte nicht annehmen. *Math. Ann.*, 1907.

- [103] C.D. Aliprantis and R. Tourky. *Cones and Duality*. Graduate studies in mathematics. American Mathematical Society, 2007.
- [104] P. Janotta and R. Lal. Generalized probabilistic theories without the no-restriction hypothesis. *Phys. Rev. A*, 87:052131, 2013.
- [105] S.P. Boyd and L. Vandenberghe. *Convex Optimization*. Berichte über verteilte messsysteme. Cambridge University Press, 2004.
- [106] A.L. Peressini and D.R. Sherbert. Ordered topological tensor products. *Proc. London Math. Soc.*, s3-19(1):177–190, 1969.
- [107] A. Hulanicki and R.R. Phelps. Some applications of tensor products of partially-ordered linear spaces. *J. Funct. Anal.*, 2(2):177–201, 1968.
- [108] E. Chitambar, D. Leung, L. Mančinska, M. Ozols, and A. Winter. Everything you always wanted to know about LOCC (but were afraid to ask). *Commun. Math. Phys.*, 2014.
- [109] M.A. Nielsen and I.L. Chuang. *Quantum Computation and Quantum Information: 10th Anniversary Edition*. Cambridge University Press, Cambridge, 2010.
- [110] I. Bengtsson and K. Życzkowski. *Geometry of Quantum States: An Introduction to Quantum Entanglement*. Cambridge University Press, Cambridge, 2006.
- [111] M. Koecher. Die Geodätischen von Positivitätsbereichen. *Math. Annalen*, 135(3):192–202, 1958.
- [112] E.B. Vinberg. Homogeneous cones. In *Soviet Math. Dokl*, volume 1, pages 787–790, 1960.
- [113] N. Brunner, D. Cavalcanti, S. Pironio, V. Scarani, and S. Wehner. Bell nonlocality. *Rev. Mod. Phys.*, 86(2):419, 2014.
- [114] L. Gurvits. Classical deterministic complexity of Edmonds’ problem and quantum entanglement. *Proc. 35th ACM Symp. on Theory of Computing*, pages 10–19, 2003.
- [115] S. Gharibian. Strong NP-hardness of the quantum separability problem. *Quantum Inf. Comput.*, 10:343–360, 2010.
- [116] A. Peres. Separability criterion for density matrices. *Phys. Rev. Lett.*, 77:1413–1415, 1996.
- [117] M. Horodecki, P. Horodecki, and R. Horodecki. Separability of mixed states: necessary and sufficient conditions. *Phys. Lett. A*, 223(1–2):1–8, 1996.
- [118] M. Horodecki and P. Horodecki. Reduction criterion of separability and limits for a class of distillation protocols. *Phys. Rev. A*, 59:4206–4216, 1999.
- [119] K. Chen and L. A. Wu. A matrix realignment method for recognizing entanglement. *Quantum Inf. Comput.*, 3:193–202, 2003.

- [120] O. Rudolph. On the cross norm criterion for separability. *J. Phys. A*, 36:5825, 2003.
- [121] O. Rudolph. Some properties of the computable cross-norm criterion for separability. *Phys. Rev. A*, 67:032312, 2003.
- [122] P. Horodecki. Separability criterion and inseparable mixed states with positive partial transposition. *Phys. Lett. A*, 232(5):333 – 339, 1997.
- [123] M.A. Nielsen and J. Kempe. Separable states are more disordered globally than locally. *Phys. Rev. Lett.*, 86:5184–5187, 2001.
- [124] L. Gurvits and H. Barnum. Separable balls around the maximally mixed multipartite quantum states. *Phys. Rev. A*, 68:042312, 2003.
- [125] G. Vidal and R. Tarrach. Robustness of entanglement. *Phys. Rev. A*, 59(1):141, 1999.
- [126] A.C. Doherty, P.A. Parrilo, and F.M. Spedalieri. Complete family of separability criteria. *Phys. Rev. A*, 69:022308, 2004.
- [127] N. Johnston. What the operator-Schmidt decomposition tells us about separability. <http://www.njohnston.ca>, 2014.
- [128] M. Christandl, R. König, G. Mitchison, and R. Renner. One-and-a-half quantum de Finetti theorems. *Commun. Math. Phys.*, 2007.
- [129] F.G.S.L. Brandao, M. Christandl, A.W. Harrow, and M. Walter. The mathematics of entanglement. *Preprint arXiv:1604.01790*, 2016.
- [130] K.G.H. Vollbrecht and R.F. Werner. Entanglement measures under symmetry. *Phys. Rev. A*, 64:062307, 2001.
- [131] T. Eggeling. On multipartite symmetric states in Quantum Information Theory. 2003.
- [132] J. Oppenheim and S. Wehner. The uncertainty principle determines the nonlocality of quantum mechanics. *Science*, 330(6007):1072–1074, 2010.
- [133] M. Banik, Md.R. Gazi, S. Ghosh, and G. Kar. Degree of complementarity determines the nonlocality in quantum mechanics. *Phys. Rev. A*, 87:052125, 2013.
- [134] P. Busch, T. Heinosaari, J. Schultz, and N. Stevens. Comparing the degrees of incompatibility inherent in probabilistic physical theories. *EPL*, 103(1):10002, 2013.
- [135] N. Stevens and P. Busch. Steering, incompatibility, and Bell-inequality violations in a class of probabilistic theories. *Phys. Rev. A*, 89:022123, 2014.
- [136] D. Cavalcanti and P. Skrzypczyk. Quantitative relations between measurement incompatibility, quantum steering, and nonlocality. *Phys. Rev. A*, 93:052112, 2016.
- [137] M. Plávala. All measurements in a probabilistic theory are compatible if

- and only if the state space is a simplex. *Phys. Rev. A*, 94:042108, 2016.
- [138] A. Jenčová and M. Plávala. Conditions on the existence of maximally incompatible two-outcome measurements in general probabilistic theory. *Preprint arXiv:1703.09447*, 2017.
- [139] G. Birkhoff. *Lattice Theory*. Number v. 25, pt. 2 in American Mathematical Society colloquium publications. American Mathematical Society, 1940.
- [140] R.R. Phelps. *Lectures on Choquet's Theorem*. Van Nostrand Reinhold mathematical studies. D. Van Nostrand Company, 1966.
- [141] F. Riesz. Sur quelques notions fondamentales dans la theorie generale des operations lineaires. *Ann. Math.*, 41(1):174–206, 1940.
- [142] A.I. Yudin. Solution of two problems on the theory of partially ordered spaces. In *Dokl. Akad. Nauk SSSR*, volume 23, pages 418–422, 1939.
- [143] A.D. Wallace. A fixed-point theorem. *Bull. Amer. Math. Soc.*, 51(6):413–416, 1945.
- [144] T. Andô. On fundamental properties of a Banach space with a cone. *Pacific J. Math.*, 12(4):1163–1169, 1962.
- [145] K. Kraus. *States, effects and operations: fundamental notions of quantum theory*. Springer, 1983.
- [146] P. Lahti and S. Pulmannová. Coexistence vs. functional coexistence of quantum observables. *Rep. Math. Phys.*, 47(2):199–212, 2001.
- [147] P. Lahti. Coexistence and joint measurability in quantum mechanics. *Int. J. Theor. Phys.*, 2003.
- [148] J.F. Clauser, M.A. Horne, A. Shimony, and R.A. Holt. Proposed experiment to test local hidden-variable theories. *Phys. Rev. Lett.*, 23:880–884, 1969.
- [149] A. Fine. Joint distributions, quantum correlations, and commuting observables. *J. Math. Phys.*, 23(7):1306–1310, 1982.
- [150] M.M. Wolf, D. Perez-Garcia, and C. Fernandez. Measurements incompatible in quantum theory cannot be measured jointly in any other no-signaling theory. *Phys. Rev. Lett.*, 103:230402, 2009.
- [151] W. Matthews, S. Wehner, and A. Winter. Distinguishability of quantum states under restricted families of measurements with an application to quantum data hiding. *Commun. Math. Phys.*, 2009.
- [152] C.W. Helstrom. *Quantum detection and estimation theory*. Academic press, 1976.
- [153] G. Kimura, T. Miyadera, and H. Imai. Optimal state discrimination in general probabilistic theories. *Phys. Rev. A*, 79:062306, 2009.
- [154] J. Walgate, A.J. Short, L. Hardy, and V. Vedral. Local distinguishability

- of multipartite orthogonal quantum states. *Phys. Rev. Lett.*, 85:4972–4975, 2000.
- [155] S. Virmani, M.F. Sacchi, M.B. Plenio, and D. Markham. Optimal local discrimination of two multipartite pure states. *Phys. Lett. A*, 288(2):62–68, 2001.
- [156] W. Matthews and A. Winter. On the Chernoff distance for asymptotic LOCC discrimination of bipartite quantum states. *Commun. Math. Phys.*, 2009.
- [157] C.H. Bennett, G. Brassard, C. Crépeau, R. Jozsa, A. Peres, and W.K. Wootters. Teleporting an unknown quantum state via dual classical and Einstein-Podolsky-Rosen channels. *Phys. Rev. Lett.*, 70:1895–1899, 1993.
- [158] R.R. Tucci. Quantum entanglement and conditional information transmission. *Preprint quant-ph/9909041*, 1999.
- [159] E.H. Lieb and M.B. Ruskai. A fundamental property of quantum-mechanical entropy. *Phys. Rev. Lett.*, 1973.
- [160] E.H. Lieb and M.B. Ruskai. Proof of the strong subadditivity of quantum mechanical entropy. *J. Math. Phys.*, 1973.
- [161] E.H. Lieb. Convex trace functions and the Wigner-Yanase-Dyson conjecture. *Adv. Math.*, 1973.
- [162] F.G.S.L. Brandão, M. Christandl, and J. Yard. Faithful squashed entanglement. *Commun. Math. Phys.*, 2011.
- [163] K. Li and A. Winter. Relative entropy and squashed entanglement. *Commun. Math. Phys.*, 2014.
- [164] K. Li and A. Winter. Squashed entanglement, k -extendibility, quantum Markov chains, and recovery maps. *Preprint arXiv:1410.4184*, 2014.
- [165] O. Fawzi and R. Renner. Quantum conditional mutual information and approximate Markov chains. *Commun. Math. Phys.*, 2015.
- [166] R.L. Hudson and G.R. Moody. Locally normal symmetric states and an analogue of de Finetti’s theorem. *Z. Wahrscheinlichkeit*, 33(4):343–351, 1976.
- [167] C.M. Caves, C.A. Fuchs, and R. Schack. Unknown quantum states: The quantum de Finetti representation. *J. Math. Phys.*, 43(9):4537–4559, 2002.
- [168] R. König and R. Renner. A de Finetti representation for finite symmetric quantum states. *J. Math. Phys.*, 46(12):122108, 2005.
- [169] F.G.S.L. Brandao and A.W. Harrow. Quantum de Finetti theorems under local measurements with applications. In *Proc. 45th ACM Symp. on Theory of Computing, STOC ’13*, pages 861–870, New York, NY, USA, 2013. ACM.
- [170] B. De Finetti. La prévision: ses lois logiques, ses sources subjectives. In

- Ann. Inst. H. Poincaré*, volume 7, pages 1–68, 1937.
- [171] F.G.S.L. Brandão, M. Christandl, and J. Yard. A quasipolynomial-time algorithm for the quantum separability problem. In *Proc. 43rd ACM Symp. on Theory of Computing*, pages 343–352. ACM, 2011.
 - [172] G. Pisier. Counterexamples to a conjecture of Grothendieck. *Acta Math.*, 151(1):181–208, 1983.
 - [173] G. Pisier. Finite rank projections on Banach spaces and a conjecture of Grothendieck. In *Proc. Intern. Congr. Math.*, volume 2, pages 1027–1039, 1983.
 - [174] P. Erdos and J. Spencer. Probabilistic methods in combinatorics. *AMC*, 10:12, 1974.
 - [175] P. Frankl, V. Rödl, and R.M. Wilson. The number of submatrices of a given type in a Hadamard matrix and related results. *J. Comb. Theory B*, 44(3):317–328, 1988.
 - [176] N. Alon and J.H. Spencer. *The probabilistic method*. John Wiley & Sons, 2016.
 - [177] N. Alon. Approximating sparse binary matrices in the cut-norm. *Linear Algebra Appl.*, 486:409–418, 2015.
 - [178] W. de Launey and D.M. Gordon. A comment on the Hadamard conjecture. *J. Comb. Theory A*, 95(1):180–184, 2001.
 - [179] S. Szarek. On the best constants in the Khinchin inequality. *Studia Math.*, 58(2):197–208, 1976.
 - [180] U. Haagerup. The best constants in the Khintchine inequality. *Studia Math.*, 70(3):231–283, 1981.
 - [181] R. Latała and K. Oleszkiewicz. On the best constant in the Khintchine–Kahane inequality. *Studia Math.*, 109(1):101–104, 1994.
 - [182] C. Eltschka and J. Siewert. Quantifying entanglement resources. *J. Phys. A*, 47(42):424005, 2014.
 - [183] J.K. Korbicz, M.L. Almeida, J. Bae, M. Lewenstein, and A. Acín. Structural approximations to positive maps and entanglement-breaking channels. *Phys. Rev. A*, 78:062105, 2008.
 - [184] D. Chruściński and A. Kossakowski. Bell diagonal states with maximal Abelian symmetry. *Phys. Rev. A*, 82:064301, 2010.
 - [185] S.M. Hashemi Rafsanjani, M. Huber, C.J. Broadbent, and J.H. Eberly. Genuinely multipartite concurrence of n -qubit x matrices. *Phys. Rev. A*, 86:062303, 2012.
 - [186] C. Eltschka and J. Siewert. Negativity as an estimator of entanglement dimension. *Phys. Rev. Lett.*, 111(10):100503, 2013.
 - [187] L.E. Buchholz, T. Moroder, and O. Gühne. Evaluating the geomet-

- ric measure of multipartite entanglement. *Ann. Phys. (Berlin)*, 528(3-4):278–287, 2016.
- [188] N. Yu. Separability of a mixture of Dicke states. *Phys. Rev. A*, 94:060101, 2016.
- [189] L. Moravčíková and M. Ziman. Entanglement-annihilating and entanglement-breaking channels. *J. Phys. A*, 43(27):275306, 2010.
- [190] M. Horodecki, P.W. Shor, and M.B. Ruskai. Entanglement breaking channels. *Rev. Math. Phys.*, 15(06):629–641, 2003.
- [191] C.H. Bennett, D.P. DiVincenzo, and J.A. Smolin. Capacities of quantum erasure channels. *Phys. Rev. Lett.*, 78(16):3217, 1997.
- [192] C.H. Bennett, P.W. Shor, J.A. Smolin, and A.V. Thapliyal. Entanglement-assisted classical capacity of noisy quantum channels. *Phys. Rev. Lett.*, 83(15):3081, 1999.
- [193] D. Bruß, L. Faoro, C. Macchiavello, and G.M. Palma. Quantum entanglement and classical communication through a depolarizing channel. *J. Mod. Opt.*, 47(2-3):325–331, 2000.
- [194] C. King. The capacity of the quantum depolarizing channel. *IEEE Trans. Inf. Theory*, 49(1):221–229, 2003.
- [195] L. Lami and Marcus Huber. Bipartite depolarizing maps. *J. Math. Phys.*, 57(9):092201, 2016.
- [196] K.G.H. Vollbrecht and M.M. Wolf. Activating distillation with an infinitesimal amount of bound entanglement. *Phys. Rev. Lett.*, 88:247901, 2002.
- [197] M. Piani and C.E. Mora. Class of positive-partial-transpose bound entangled states associated with almost any set of pure entangled states. *Phys. Rev. A*, 75:012305, 2007.
- [198] R.A. Horn and C.R. Johnson. *Matrix Analysis*. Cambridge University Press, 1990.
- [199] R.A. Horn and C.R. Johnson. *Topics in Matrix Analysis*. Topics in Matrix Analysis. Cambridge University Press, 1994.
- [200] Satoshi Ishizaka. Bound entanglement provides convertibility of pure entangled states. *Phys. Rev. Lett.*, 93:190501, 2004.
- [201] M.-D. Choi. Completely positive linear maps on complex matrices. *Linear Algebra Appl.*, 10(3):285 – 290, 1975.
- [202] W.F. Stinespring. Positive functions on C^* -algebras. *Proc. Am. Math. Soc.*, 6(2):211–216, 1955.
- [203] D. Chruściński and A. Kossakowski. Multipartite invariant states. i. unitary symmetry. *Phys. Rev. A*, 73:062314, 2006.
- [204] M. Huber and M. Plesch. Purification of genuine multipartite entangle-

- ment. *Phys. Rev. A*, 83(6):062321, 2011.
- [205] A. Sanpera, D. Bruß, and M. Lewenstein. Schmidt-number witnesses and bound entanglement. *Phys. Rev. A*, 63:050301, 2001.
- [206] A. Müller-Hermes, D. Reeb, and M.M. Wolf. Positivity of linear maps under tensor powers. *J. Math. Phys.*, 57(1):015202, 2016.
- [207] N.J. Cerf, G. Leuchs, and E.S. Polzik. *Quantum information with continuous variables of atoms and light*. Imperial College Press, 2007.
- [208] C. Weedbrook, S. Pirandola, R. García-Patrón, N.J. Cerf, T.C. Ralph, J.H. Shapiro, and S. Lloyd. Gaussian quantum information. *Rev. Mod. Phys.*, 84:621–669, 2012.
- [209] G. Adesso, S. Ragy, and A.R. Lee. Continuous variable quantum information: Gaussian states and beyond. *Open Syst. Inf. Dyn.*, 21(01n02):1440001, 2014.
- [210] A. Serafini. *Quantum Continuous Variables: A Primer of Theoretical Methods*. CRC Press, Taylor & Francis Group, 2017.
- [211] A. Furusawa, J.L. Sørensen, S.L. Braunstein, C.A. Fuchs, H.J. Kimble, and E.S. Polzik. Unconditional quantum teleportation. *Science*, 282(5389):706–709, 1998.
- [212] J.F. Sherson, H. Krauter, R.K. Olsson, B. Julsgaard, K. Hammerer, I. Cirac, and E.S. Polzik. Quantum teleportation between light and matter. *Nature*, 443(7111):557, 2006.
- [213] H. Krauter, D. Salart, C.A. Muschik, J.M. Petersen, H. Shen, T. Fernholz, and E.S. Polzik. Deterministic quantum teleportation between distant atomic objects. *Nat. Phys.*, 9(7), 2013.
- [214] F. Grosshans, G. Van Assche, J. Wenger, R. Brouri, N.J. Cerf, and P. Grangiera. Quantum key distribution using gaussian-modulated coherent states. *Nature (London)*, 421:238–241, 2003.
- [215] The LIGO Scientific Collaboration. A gravitational wave observatory operating beyond the quantum shot-noise limit. *Nat. Phys.*, 7(12):962–965, 2011.
- [216] The LIGO Scientific Collaboration. Enhanced sensitivity of the LIGO gravitational wave detector by using squeezed states of light. *Nat. Photonics*, 7(8):613–619, 2013.
- [217] V.I. Arnold. *Mathematical methods of classical mechanics*. Springer Verlag, New York, 1978.
- [218] G. Giedke, B. Kraus, M. Lewenstein, and J.I. Cirac. Entanglement criteria for all bipartite Gaussian states. *Phys. Rev. Lett.*, 87:167904, 2001.
- [219] A. Serafini, G. Adesso, and F. Illuminati. Unitarily localizable entanglement of Gaussian states. *Phys. Rev. A*, 71:032349, 2005.
- [220] A.S. Holevo and R.F. Werner. Evaluating capacities of bosonic Gaussian

- channels. *Phys. Rev. A*, 63:032312, 2001.
- [221] A. Botero and B. Reznik. Modewise entanglement of Gaussian states. *Phys. Rev. A*, 67:052311, 2003.
- [222] G. Giedke, J. Eisert, I.J. Cirac, and M.B. Plenio. Entanglement transformations of pure Gaussian states. *Quantum Inf. Comput.*, 3(3):211–223, 2003.
- [223] J. Eisert, S. Scheel, and M.B. Plenio. Distilling Gaussian states with Gaussian operations is impossible. *Phys. Rev. Lett.*, 89:137903, 2002.
- [224] J. Fiurášek. Gaussian transformations and distillation of entangled Gaussian states. *Phys. Rev. Lett.*, 89:137904, 2002.
- [225] G. Giedke and I.J. Cirac. Characterization of Gaussian operations and distillation of Gaussian states. *Phys. Rev. A*, 66:032316, 2002.
- [226] C.H. Bennett, G. Brassard, S. Popescu, B. Schumacher, J.A. Smolin, and W.K. Wootters. Purification of noisy entanglement and faithful teleportation via noisy channels. *Phys. Rev. Lett.*, 76:722–725, 1996.
- [227] M. Horodecki, P. Horodecki, and R. Horodecki. Mixed-state entanglement and distillation: Is there a “bound” entanglement in nature? *Phys. Rev. Lett.*, 80:5239–5242, 1998.
- [228] G. Giedke, L.-M. Duan, I.J. Cirac, and P. Zoller. Distillability criterion for all bipartite Gaussian states. *Quantum Inf. Comput.*, 1(3):79–86, 2001.
- [229] L. Lami, A. Serafini, and G. Adesso. Gaussian entanglement revisited. *Preprint arXiv:1612.05215*, 2016.
- [230] G. Giedke, B. Kraus, M. Lewenstein, and J.I. Cirac. Separability properties of three-mode Gaussian states. *Phys. Rev. A*, 64:052303, 2001.
- [231] G. Adesso and F. Illuminati. Entanglement in continuous-variable systems: recent advances and current perspectives. *J. Phys. A*, 40(28):7821, 2007.
- [232] F. Zhang. *The Schur complement and its applications*, volume 4. Springer Science & Business Media, 2006.
- [233] P. Hyllus and J. Eisert. Optimal entanglement witnesses for continuous-variable systems. *New J. Phys.*, 8(4):51, 2006.
- [234] G. Adesso and R. Simon. Strong subadditivity for log-determinant of covariance matrices and its applications. *J. Phys. A*, 49(34):34LT02, 2016.
- [235] L. Lami, C. Hirche, G. Adesso, and A. Winter. Schur complement inequalities for covariance matrices and monogamy of quantum correlations. *Phys. Rev. Lett.*, 117:220502, 2016.
- [236] L. Lami, C. Hirche, G. Adesso, and A. Winter. From log-determinant inequalities to Gaussian entanglement via recoverability theory. *IEEE*

- Trans. Inf. Theory*, Preprint, 2017.
- [237] B.V. Rajarama Bhat, K.R. Parthasarathy, and R. Sengupta. On the equivalence of separability and extendability of quantum states. *Rev. Math. Phys.*, 29(04):1750012, 2017.
- [238] B. Kraus, J.I. Cirac, S. Karnas, and M. Lewenstein. Separability in $2 \times N$ composite quantum systems. *Phys. Rev. A*, 61:062302, 2000.
- [239] J. Eisert and M.B. Plenio. Introduction to the basics of entanglement theory in continuous-variable systems. *Int. J. Quantum Inf.*, 01(04):479–506, 2003.
- [240] R. Bhatia. *Positive definite matrices*. Princeton university press, 2009.
- [241] W.N. Anderson and R.J. Duffin. Series and parallel addition of matrices. *J. Math. Anal. Appl.*, 26(3):576–594, 1969.
- [242] T. Ando. Concavity of certain maps on positive definite matrices and applications to hadamard products. *Linear Algebra Appl.*, 26:203 – 241, 1979.
- [243] W. Pusz and S.L. Woronowicz. Functional calculus for sesquilinear forms and the purification map. *Rep. Math. Phys.*, 8(2):159 – 170, 1975.
- [244] R. Simon, N. Mukunda, and B. Dutta. Quantum-noise matrix for multimode systems: $U(n)$ invariance, squeezing, and normal forms. *Phys. Rev. A*, 49:1567–1583, 1994.
- [245] Arvind, B. Dutta, N. Mukunda, and R. Simon. The real symplectic groups in quantum mechanics and optics. *Pramana*, 1995.
- [246] J. Williamson. On the algebraic problem concerning the normal forms of linear dynamical systems. *Am. J. Math.*, 58(1):141–163, 1936.
- [247] R. Simon, S. Chaturvedi, and V. Srinivasan. Congruences and canonical forms for a positive matrix: Application to the Schweinler-Wigner extremum principle. *J. Math. Phys.*, 40(7):3632–3642, 1999.
- [248] L. Chen and Y.-X. Chen. Rank-three bipartite entangled states are distillable. *Phys. Rev. A*, 78:022318, 2008.
- [249] J. Manuceau and A. Verbeure. Quasi-free states of the C.C.R.-algebra and Bogoliubov transformations. *Commun. Math. Phys.*, 1968.
- [250] G. Adesso, A. Serafini, and F. Illuminati. Quantification and scaling of multipartite entanglement in continuous variable systems. *Phys. Rev. Lett.*, 93:220504, 2004.
- [251] G. Adesso and F. Illuminati. Genuine multipartite entanglement of symmetric Gaussian states: Strong monogamy, unitary localization, scaling behavior, and molecular sharing structure. *Phys. Rev. A*, 78:042310, 2008.
- [252] R. Bhatia and T. Jain. On symplectic eigenvalues of positive definite matrices. *J. Math. Phys.*, 56(11):112201, 2015.

- [253] M. Kuś and K. Życzkowski. Geometry of entangled states. *Phys. Rev. A*, 63:032307, 2001.
- [254] S. Arunachalam, N. Johnston, and V. Russo. Is absolute separability determined by the partial transpose? *Preprint arXiv:1405.5853*, 2014.
- [255] R. Hildebrand. Positive partial transpose from spectra. *Phys. Rev. A*, 76:052325, 2007.
- [256] F. Verstraete, K. Audenaert, and B. De Moor. Maximally entangled mixed states of two qubits. *Phys. Rev. A*, 64:012316, 2001.
- [257] N. Johnston. Separability from spectrum for qubit-qudit states. *Phys. Rev. A*, 88:062330, 2013.
- [258] J. Eisert, T. Tyc, T. Rudolph, and B.C. Sanders. Gaussian quantum marginal problem. *Commun. Math. Phys.*, 2008.
- [259] H.M. Wiseman, S.J. Jones, and A.C. Doherty. Steering, entanglement, nonlocality, and the Einstein-Podolsky-Rosen paradox. *Phys. Rev. Lett.*, 98:140402, 2007.
- [260] G. Adesso, D. Girolami, and A. Serafini. Measuring Gaussian quantum information and correlations using the Rényi entropy of order 2. *Phys. Rev. Lett.*, 109:190502, 2012.
- [261] Y. Xiang, I. Kogias, G. Adesso, and Q. He. Multipartite Gaussian steering: monogamy constraints and cryptographical applications. *Preprint arXiv:1603.08173*, 2016.
- [262] *Speakable and Unspeakable in Quantum Mechanics: Collected Papers on Quantum Philosophy*.
- [263] M.M. Wolf, G. Giedke, O. Krüger, R.F. Werner, and J.I. Cirac. Gaussian entanglement of formation. *Phys. Rev. A*, 69(5):052320, 2004.
- [264] G. Giedke, M.M. Wolf, O. Krüger, R.F. Werner, and J.I. Cirac. Entanglement of formation for symmetric Gaussian states. *Phys. Rev. Lett.*, 91(10):107901, 2003.
- [265] D. Gross and M. Walter. Stabilizer information inequalities from phase space distributions. *J. Math. Phys.*, 54(8):082201, 2013.
- [266] S.-W. Ji, M.S. Kim, and H. Nha. Quantum steering of multimode Gaussian states by Gaussian measurements: monogamy relations and the Peres conjecture. *J. Phys. A*, 48(13):135301, 2015.
- [267] R. Gallego and L. Aolita. Resource theory of steering. *Phys. Rev. X*, 5:041008, 2015.
- [268] A.B. Sainz, N. Brunner, D. Cavalcanti, P. Skrzypczyk, and T. Vértesi. Postquantum steering. *Phys. Rev. Lett.*, 115:190403, 2015.
- [269] G. Vidal and R.F. Werner. Computable measure of entanglement. *Phys. Rev. A*, 65:032314, 2002.

- [270] M.B. Plenio. Logarithmic negativity: A full entanglement monotone that is not convex. *Phys. Rev. Lett.*, 95:090503, 2005.
- [271] I. Kogias and G. Adesso. Einstein-Podolsky-Rosen steering measure for two-mode continuous variable states. *J. Opt. Soc. Am. B*, 32(4):A27–A33.
- [272] Sabine Wollmann, Nathan Walk, Adam J Bennet, Howard M Wiseman, and Geoff J Pryde. Observation of genuine one-way einstein-podolsky-rosen steering. *Phys. Rev. Lett.*, 116(16):160403, 2016.
- [273] S.-W. Ji, J. Lee, J. Park, and H. Nha. Quantum steering of Gaussian states via non-Gaussian measurements. *Sci. Rep.*, 6:29729, 2016.
- [274] L. Barnett, A.B. Barrett, and A.K. Seth. Granger causality and transfer entropy are equivalent for Gaussian variables. *Phys. Rev. Lett.*, 103(23):238701, 2009.
- [275] M.A. de Gosson. *Symplectic Geometry and Quantum Mechanics*. Operator Theory: Advances and Applications. Birkhäuser Basel, 2006.
- [276] T.M. Cover and A. Thomas. Determinant inequalities via information theory. *SIAM J. Matrix Anal. Appl.*, 9(3):384–392, 1988.
- [277] L. Henderson and V. Vedral. Classical, quantum and total correlations. *J. Phys. A*, 34(35):6899, 2001.
- [278] V. Vedral. Classical correlations and entanglement in quantum measurements. *Phys. Rev. Lett.*, 90:050401, 2003.
- [279] B. Groisman, S. Popescu, and A. Winter. Quantum, classical, and total amount of correlations in a quantum state. *Phys. Rev. A*, 72:032317, 2005.
- [280] B. Schumacher and M.D. Westmoreland. Quantum mutual information and the one-time pad. *Phys. Rev. A*, 74:042305, 2006.
- [281] D.P. DiVincenzo, M. Horodecki, D.W. Leung, J.A. Smolin, and B.M. Terhal. Locking classical correlations in quantum states. *Phys. Rev. Lett.*, 92:067902, 2004.
- [282] P. Hayden, D.W. Leung, and A. Winter. Aspects of generic entanglement. *Commun. Math. Phys.*, 2006.
- [283] V. Giovannetti, R. García-Patrón, N. J. Cerf, and A. S. Holevo. *Nat. Photonics*.
- [284] G. Adesso, T.R. Bromley, and M. Cianciaruso. Measures and applications of quantum correlations. *J. Phys. A*, 49(47):473001, 2016.
- [285] T. Hiroshima, G. Adesso, and F. Illuminati. Monogamy inequality for distributed Gaussian entanglement. *Phys. Rev. Lett.*, 98:050503, 2007.
- [286] G. Adesso and F. Illuminati. Strong monogamy of bipartite and genuine multipartite entanglement: The Gaussian case. *Phys. Rev. Lett.*, 99:150501, 2007.

- [287] C. Lancien, S. Di Martino, M. Huber, M. Piani, G. Adesso, and A. Winter. Should entanglement measures be monogamous or faithful? *Phys. Rev. Lett.*, 117:060501, 2016.
- [288] R. Raussendorf and H.J. Briegel. A one-way quantum computer. *Phys. Rev. Lett.*, 86:5188–5191, 2001.
- [289] D. Gross. Hudson’s theorem for finite-dimensional quantum systems. *J. Math. Phys.*, 47(12):122107, 2006.
- [290] A. Dembo, T.M. Cover, and J.A. Thomas. Information theoretic inequalities. *IEEE Trans. Inf. Theory*, 37(6):1501–1518, 1991.
- [291] A. Wehrl. General properties of entropy. *Rev. Mod. Phys.*, 50:221–260, 1978.
- [292] R.W. Yeung. A framework for linear information inequalities. *IEEE Trans. Inf. Theory*, 43(6):1924–1934, 1997.
- [293] D.W. Robinson and D. Ruelle. Mean entropy of states in classical statistical mechanics. *Commun. Math. Phys.*, 1967.
- [294] O.E. Lanford III and D.W. Robinson. Mean entropy of states in quantum-statistical mechanics. *J. Math. Phys.*, 9(7):1120–1125, 1968.
- [295] T. Ando and D. Petz. Gaussian Markov triplets approached by block matrices. *Acta Sci. Math. (Szeged)*, 75(1-2):329–345, 2009.
- [296] F. Hiai and D. Petz. *Introduction to matrix analysis and applications*. Springer Science & Business Media, 2014.
- [297] K. Audenaert, F. Hiai, and D. Petz. Strongly subadditive functions. *Acta Math. Hung.*, 2010.
- [298] M. Lewin and J. Sabin. A family of monotone quantum relative entropies. *Lett. Math. Phys.*, 104(6):691–705, 2014.
- [299] R. Bhatia. *Matrix Analysis*. Graduate Texts in Mathematics. Springer New York, 2013.
- [300] T.H. Chan. Balanced information inequalities. *IEEE Trans. Inf. Theory*, 49(12):3261–3267, 2003.
- [301] Z. Zhang and R.W. Yeung. On characterization of entropy function via information inequalities. *IEEE Trans. Inf. Theory*, 44(4):1440–1452, 1998.
- [302] R. Dougherty, C. Freiling, and K. Zeger. Six new non-shannon information inequalities. In *ISIT 2006*, pages 233–236. IEEE, 2006.
- [303] F. Matus. Infinitely many information inequalities. In *ISIT 2007*, pages 41–44. IEEE, 2007.
- [304] C.R. Johnson and W.W. Barrett. Determinantal inequalities for positive definite matrices. *Discrete Math.*, 119(1-3):97–106, 1993.
- [305] B. Hassibi and S. Shadbakht. The entropy region for three gaussian random variables. In *ISIT 2008*, pages 2634–2638. IEEE, 2008.

- [306] S. Shadbakht and B. Hassibi. On the entropy region of Gaussian random variables. *Preprint arXiv:1112.0061*, 2011.
- [307] N. Linden, M. Mosonyi, and A. Winter. The structure of rényi entropic inequalities. In *Proc. Royal Soc. A*, volume 469, page 20120737. The Royal Society, 2013.
- [308] M. Tomamichel. *Quantum Information Processing with Finite Resources: Mathematical Foundations*, volume 5. Springer, 2015.
- [309] D. Petz. Sufficient subalgebras and the relative entropy of states of a von neumann algebra. *Commun. Math. Phys.*, 1986.
- [310] F.G.S.L. Brandão, A.W. Harrow, J. Oppenheim, and S. Strelchuk. Quantum conditional mutual information, reconstructed states, and state redistribution. *Phys. Rev. Lett.*, 115:050501, 2015.
- [311] T. Cooney, C. Hirche, C. Morgan, J.P. Olson, K.P. Seshadreesan, J. Watrous, and M.M. Wilde. Operational meaning of quantum measures of recovery. *Phys. Rev. A*, 94(2):022310, 2016.
- [312] L. Lami, S. Das, and M.M. Wilde. Approximate reversal of quantum Gaussian dynamics. *Preprint arXiv:1702.04737*, 2017.
- [313] M.A. Woodbury. Inverting modified matrices. *Memorandum Rept.*, 42, 1950.
- [314] Müller-Lennert M., F. Dupuis, O. Szehr, S. Fehr, and Tomamichel. On quantum rényi entropies: A new generalization and some properties. *J. Math. Phys.*, 54(12):122203, 2013.
- [315] K.M.R. Audenaert. Comparisons between quantum state distinguishability measures. *Quantum Inf. Comput.*, 14(1-2).
- [316] I. Devetak, A.W. Harrow, and A.J. Winter. A resource framework for quantum shannon theory. *IEEE Trans. Inf. Theory*, 54(10):4587–4618, 2008.
- [317] K.P. Seshadreesan, M. Berta, and M.M. Wilde. Rényi squashed entanglement, discord, and relative entropy differences. *J. Phys. A*, 48(39):395303, 2015.
- [318] L. Mišta and R. Tatham. Gaussian intrinsic entanglement. *Phys. Rev. Lett.*, 117:240505, 2016.