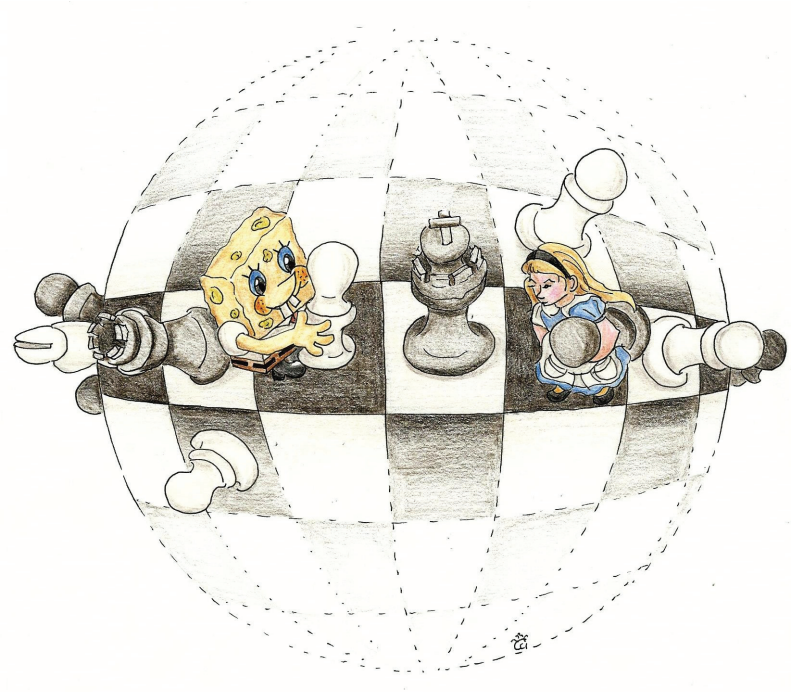
Institut Camille Jordan
UMR 5208 du CNRS

Grande dimension et symétries en théorie quantique de l'information



Cécilia Lancien

Thèse de doctorat

Université Claude Bernard Lyon 1

Université Claude Bernard Lyon 1
École doctorale InfoMath, ED 512
Spécialité : Mathématiques
N. d'ordre NNT : 2016 LYSE1077
En cotutelle avec l'Universitat Autònoma de Barcelona

Grande dimension et symétries en théorie quantique de l'information

High dimension and symmetries in quantum information theory

Thèse de doctorat

Soutenue publiquement le 9 juin 2016 par

Cécilia Lancien

devant le jury composé de:

M. Stéphane Attal	Université Claude Bernard Lyon 1	Examinateur
M. Guillaume Aubrun	Université Claude Bernard Lyon 1	Directeur
M. Florent Benaych-Georges	Université Descartes Paris 5	Examinateur
M. John Calsamiglia	Universitat Autònoma de Barcelona	Examinateur
M. Olivier Guédon	Université Paris-Est Marne-La-Vallée	Examinateur
M. Marius Junge	University of Illinois at Urbana-Champaign	Rapporteur
Mme Stephanie Wehner	Technische Universiteit Delft	Rapporteuse
M. Andreas Winter	Universitat Autònoma de Barcelona	Directeur

Pourquoi fait-on des mathématiques? Parce que c'est intéressant, parce qu'on est curieux, sans doute, mais surtout parce qu'elles sont utiles.

Gilles Godefroy, Les mathématiques mode d'emploi

Abstract / Résumé / Resumen

Abstract

If a one-phrase summary of the subject of this thesis were required, it would be something like: miscellaneous large (but finite) dimensional phenomena in quantum information theory. That said, it could nonetheless be helpful to briefly elaborate. Starting from the observation that quantum physics unavoidably has to deal with high dimensional objects, basically two routes can be taken: either try and reduce their study to that of lower dimensional ones, or try and understand what kind of universal properties might precisely emerge in this regime. We actually do not choose which of these two attitudes to follow here, and rather oscillate between one and the other.

In the first part of this manuscript (Chapters 5 and 6), our aim is to reduce as much as possible the complexity of certain quantum processes, while of course still preserving their essential characteristics. The two types of processes we are interested in are quantum channels and quantum measurements. In both cases, complexity of a transformation is measured by the number of operators needed to describe its action, and proximity of the approximating transformation towards the original one is defined in terms of closeness between the two outputs, whatever the input. We propose universal ways of achieving our quantum channel compression and quantum measurement sparsification goals (based on random constructions) and prove their optimality.

Oppositely, the second part of this manuscript (Chapters 7, 8 and 9) is specifically dedicated to the analysis of high dimensional quantum systems and some of their typical features. Stress is put on multipartite systems and on entanglement-related properties of theirs. We essentially establish the following: as the dimensions of the underlying spaces grow, being barely distinguishable by local observers is a generic trait of multipartite quantum states, and being very rough approximations of separability itself is a generic trait of separability relaxations. On the technical side, these statements stem mainly from average estimates for suprema of Gaussian processes, combined with the concentration of measure phenomenon.

In the third part of this manuscript (Chapters 10 and 11), we eventually come back to a more dimensionality reduction state of mind. This time though, the strategy is to make use of the symmetries inherent to each particular situation we are looking at in order to derive a problem-dependent simplification. By quantitatively relating permutation-symmetry and independence, we are able to show the multiplicative behaviour of several quantities showing up in quantum information theory (such as support functions of sets of states, winning probabilities in multi-player non-local games etc.). The main tool we develop for that purpose is an adaptable de Finetti type result.

Résumé

S'il fallait résumer le sujet de cette thèse en une expression, cela pourrait être quelque chose comme: phénomènes de grande dimension (mais néanmoins finie) en théorie quantique de l'information. Cela étant dit, essayons toutefois de développer brièvement. La physique quantique a inéluctablement affaire à des objets de grande dimension. Partant de cette observation, il y a, en gros, deux stratégies qui peuvent être adoptées: ou bien essayer de ramener leur étude à celle de situations de plus petite dimension, ou bien essayer de comprendre quels sont les comportements universels précisément susceptibles d'émerger dans ce régime. Nous ne donnons ici notre préférence à aucune de ces deux attitudes, mais au contraire oscillons constamment entre l'une et l'autre.

Notre but dans la première partie de ce manuscrit (Chapitres 5 et 6) est de réduire autant que possible la complexité de certains processus quantiques, tout en préservant, évidemment, leurs caractéristiques essentielles. Les deux types de processus auxquels nous nous intéressons sont les canaux quantiques et les mesures quantiques. Dans les deux cas, la complexité d'une transformation est mesurée par le nombre d'opérateurs nécessaires pour décrire son action, tandis que la proximité entre la transformation d'origine et son approximation est définie par le fait que, quel que soit l'état d'entrée, les deux états de sortie doivent être proches l'un de l'autre. Nous proposons des solutions universelles (basées sur des constructions aléatoires) à ces problèmes de compression de canaux quantiques et d'amenuisement de mesures quantiques, et nous prouvons leur optimalité.

La deuxième partie de ce manuscrit (Chapitres 7, 8 et 9) est, au contraire, spécifiquement dédiée à l'analyse de systèmes quantiques de grande dimension et certains de leurs traits typiques. L'accent est mis sur les systèmes multi-partites et leurs propriétés ayant un lien avec l'intrication. Les principaux résultats auxquels nous aboutissons peuvent se résumer de la façon suivante: lorsque les dimensions des espaces sous-jacents augmentent, il est générique pour les états quantiques multi-partites d'être à peine distinguables par des observateurs locaux, et il est générique pour les relaxations de la notion de séparabilité d'en être des approximations très grossières. Sur le plan technique, ces assertions sont établies grâce à des estimations moyennes de suprema de processus gaussiens, combinées avec le phénomène de concentration de la mesure.

Dans la troisième partie de ce manuscrit (Chapitres 10 et 11), nous revenons pour finir à notre état d'esprit de réduction de dimensionnalité. Cette fois pourtant, la stratégie est plutôt: pour chaque situation donnée, tenter d'utiliser au maximum les symétries qui lui sont inhérentes afin d'obtenir une simplification qui lui soit propre. En reliant de manière quantitative symétrie par permutation et indépendance, nous nous retrouvons en mesure de montrer le comportement multiplicatif de plusieurs quantités apparaissant en théorie quantique de l'information (fonctions de support d'ensembles d'états, probabilités de succès dans des jeux multi-joueurs non locaux etc.). L'outil principal que nous développons dans cette optique est un résultat de type de Finetti particulièrement malléable.

Resumen

En unas palabras, el tema de esta tesis se podría resumir como: fenómenos varios en alta (pero finita) dimensión en teoría cuántica de la información. Dicho esto, sin embargo podemos dar algunos detalles de más. Empezando con la observación que la física cuántica ineludiblemente tiene que tratar con objetos de alta dimensión, se pueden seguir esencialmente dos caminos: o intentar reducir su estudio al de otros que tienen dimensión más baja, o intentar comprender qué tipo de comportamiento universal surge precisamente en este régimen. Aquí no elegimos cuál de estas dos posturas hay que adoptar, sino que oscilamos constantemente entre una y la otra.

En la primera parte de este manuscrito (Capítulos 5 y 6), nuestro objetivo es reducir al mínimo posible la complejidad de ciertos procesos cuánticos, preservando sus características esenciales. Los dos tipos de procesos que nos interesan son canales cuánticos y medidas cuánticas. En ambos casos, la complejidad de una transformación se cuantifica con el número de operadores necesarios para describir su acción, y la proximidad entre la transformación de origen y su aproximación se define por el hecho de que, cualquiera que sea el estado de entrada, los respectivos estados de salida deben ser suficientemente similares. Proponemos maneras universales de alcanzar nuestras metas de compresión de canales cuánticos y rarefacción de medidas cuánticas (basadas en construcciones aleatorias) y demostramos su optimalidad.

En contrapartida, la segunda parte de este manuscrito (Capítulos 7, 8 y 9) se dedica específicamente al análisis de sistemos cuánticos de alta dimensión y sus rasgos típicos. El énfasis se pone sobre sistemos multipartidos y sus propiedades de entrelazamiento. En resumen, establecemos principalmente lo siguiente: cuando las dimensiones de los espacios subyacentes aumentan, es genérico para estados cuánticos multipartidos ser prácticamente indistinguible mediante observaciones locales, y es genérico para relajaciones de la noción de separabilidad ser burdas aproximaciones de ella. Desde un punto de vista técnico, estos resultados se derivan de estimaciones de promedio para supremosa de procesos gaussianos, combinadas con el fenómeno de concentración de la medida.

En la tercera parte de este manuscrito (Capítulos 10 y 11), finalmente volvemos a una filsofía de reducción de dimensionalidad. Pero esta vez, nuestra estrategia es utilizar las simetrías inherentes a cada situación particular que consideramos para derivar una simplificación adecuada. Vinculamos de manera cuantitativa simetría por permutación y independencia, lo que nos permite establecer el comportamiento multiplicativo de varias cuantidades que ocurren en teoría cuántica de la información (funciones de soporte de conjuntos de estados, probabilidad de éxito en juegos multi-jugadores no locales etc.). La principal herramienta técnica que desarrollamos con este fin es un resultado de tipo de Finetti muy adaptable.

Acknowledgements / Remerciements

Life is a succession of coincidences, in my scientific case, most lucky ones I must say. When I arrived in Bristol in 2012 to do a first Master's research project with Andreas, the only thing I knew about him was that he was a bad boy wearing sunglasses. When I arrived in Lyon in 2013 to do a second Master's research project with Guillaume, the only thing I knew about him was that he was the successful inventor of a board game. So how on earth could I have guessed from this meagre prior knowledge that it would then be with both of them, science-wise and person-wise, such love at first sight? That being so, I had the inestimable luxury of launching into my PhD already eager and confident that I was in the best possible hands! There are so many things I would like to thank Andreas and Guillaume for that I will unavoidably have to content myself with a random selection here. I am indebted (and immensely grateful) to Andreas for teaching me that solving a problem with basic tools is not shameful (quite the reverse), for inculcating in me (or at least trying to) that if you are not convinced yourself of the interest of what you are doing there is no way the person in front is, for replying at any weird time to any weird math questioning of mine... and also for guiding me up the highest peak in the Pyrenees. I am beholden (and extremely thankful) to Guillaume for exemplifying from every viewpoint the concept of "few words but perfectly chosen words" (which I am obviously still totally unable to follow myself), for never letting pass the slightest bold (hence frequently erroneous) statement of mine, for replying at any non-weird time to any weird math questioning of mine... and also for sharing delightful running discussions. One summarizing word, and then I promise I move on: I thank my two scientific daddies of course as a priority for their availability and dedication during these past few years, for the skilful balance they achieved between guidance and "manage by yourself", for getting me going keenly on fascinating problems, but also for the not so anecdotal fact that they are as madly workaholic and perfectionist as me, so that, interacting with them, I almost felt like a sane person!

Merci à Aurélie qui ne m'a elle non plus pas (trop) prise pour une folle quand je suis venue lui expliquer ce que je souhaitais comme dessin de couverture, et s'est ensuite laissé séduire par les charmes d'Alice in Wonderland et de Sponge Bob!

Having Marius Junge and Stephanie Wehner as referees of this manuscript represents something special for me. My very first contact with quantum information theory, and at the same time my very first researcher's experience, consisted in generalizing a result of Stephanie's. After one week, I was feeling so passionately about the topic that it was already clear to me what I wanted to dedicate my professional life to! Still before even starting my PhD was my quite memorable first encounter with Marius: me giving a talk at the end of which he casually asked "but isn't what you're doing approximating zonoids by zonotopes?". Needless to say this turned out to be true (even though I had absolutely no clue at that time what the hell these objects could be), and gave the questions I was investigated a new dimension! Hence, neither the irreconcilable footballistic disagreement with Stephanie nor the teasing of Marius on my mathematical antecedents alter the great deal of respect I have for them. I am therefore extremely happy to have them both reviewing my PhD manuscript today.

I am also grateful to Stéphane Attal and John Calsamiglia for accepting to be the two "locals" in my PhD committee. I think it is fair to say that Stéphane is the one from whom this whole story originates, as it was thanks to him that I ended up doing my first Master's research project with Andreas. I therefore owe it to him in a sense to be today where I am, all the more given that, since then, he has been continuously taking on a crucial mentor duty for me. As for John, I have rarely met such a scientifically curious person. From interacting with him, even useless mathematicians like me get the feeling that what they are doing is actually interesting and exciting! Seemingly of secondary importance perhaps (but this is also what builds a relationship going far beyond strictly professional), during each of my stays in Barcelona, I could share with him early morning office discussions on basically everything (including hopeless Franco-Spanish administration or politics).

Florent Benaych-Georges et Olivier Guédon me font l'honneur de compléter mon jury, et je les en remercie.

Mon tout premier contact avec l'analyse fonctionnelle asymptotique a été via un cours de Master donné par Olivier. Au vu de mes centres d'intérêt mathématiques actuels, il semble inutile de préciser que cette introduction au sujet a été loin d'être rebutante. Quant à Florent, je me suis retrouvé l'an dernier à étudier plusieurs de ses papiers dont je pensais pouvoir utiliser les résultats. Le projet s'est soldé par un échec, mais ce n'est pas de sa faute, et cela me fait dans tous les cas très plaisir de pouvoir enfin le rencontrer pour de vrai.

Merci aux probabilistes de l'ICJ et de l'UMPA pour leur immense ouverture d'esprit: ils m'ont gentillement accueillie parmi eux, et m'ont même fait m'y sentir à ma place, en dépit du fait que je sois bien loin d'être une véritable probabiliste! La qualité de mes séjours à Lyon doit aussi beaucoup à l'ambiance au sein du groupe de thésards. Que vous soyez 11h30-iste, 12h30-iste ou non-affilié, à tous un grand merci donc. Avec une petite dédicace particulière à l'intention d'Adriane, Benjamin et Xavier, qui m'ont légèrement précédée (et avec qui je pouvais partager états d'âme linguistiques, métaphysiques et mathématiques, respectivement), ainsi que de Christian, Hugo, Luigia, Simon A., Simon B., Sylvain et Tomás, que j'ai légèrement devancés. Enfin, petite dédicace super particulière à l'intention d'Ivan, "ingénieur" comme moi, et de Nadja, "matheuse ultime" quant à elle, mes deux compagnons de tout pendant ces trois années, du scientifique au presque philosophique, en passant par le plus festif: ma vie braconnienne, et tellement au-delà, aurait été bien fade sans vous! La quasi-simultanéité de nos soutenances (la petite tortue que je suis étant logiquement la dernière à passer) est une fort belle conclusion à notre "threesome d'atypiques". Une pensée enfin à l'égard de Dario et Thomas (malgré les cauchemars qu'ils m'ont fait faire sur ma soutenance) pour les sympathiques soirées sportives passées ensemble. Et parfaite transition avec le paragraphe suivant: merci à Daniel, le Catalan-presque-Lyonnais, grâce à qui j'ai eu un toit à Barcelone.

Many thanks to all GIQitos for the emotional "so nice that you're back" or "so sad that you're gone" each time I was coming to or leaving from Barcelona, for the enthusiasm in front of each new cake on average or smelly French cheese of mine, for the passionate Catalunya vs Rest of the World discussions, and for so much more! A specific thought for the Graciosos, Christoph, Gianni, Kk, Ludovico, Mohammad and Sara, who granted me the status of emeritus member of theirs, having at least spiritual, if not physical anymore, presence in this awesome group. Another special thought for the Austrian Dream Team, aka Claude, Marcus and Pauli, for nothing being ever a problem (especially not hosting me anytime anywhere), for everything being always super geil (from French underground music to creepy maths theorems), well basically for being so refreshingly generous and excited! And a last personalised thought for Alex and Milan, my two functional analysis companions in Barcelona. Amongst other, I owe to Alex a now more or less correct Spanish summary of my thesis, and I owe to Milan a few unforgettable hikes! Finally, before closing these Barcelona thanks, let us not forget about expressing our gratitude to some emblematic host institutions there: LIQUID and MAFIA for their "always welcome" motto, Chivuo's, Gasterea, Quimet et al. for their gastro-social federating power!

There are many people without whom my scientific life up to now would not have been the same, precisely because what they brought me goes way beyond the purely scientific field. There is no good manner of ordering them in my acknowledgements so I will just go for chronological order of first encounter. Toby Cubitt was enthusiastically paying me a beer on the day where my first paper was released on the arXiv, and two days later a guy I did not know at this baby mathematician stage, Ashley Montanaro, had already carefully read the paper in question and was sending me detailed comments about it! Since then, both of them have shown (or at least well-pretended to show) constant interest in what I was doing, even enquiring more often than not about the most repelling details of certain proofs. Getting such kind of positive feedback is simply invaluable as a youngster. I met Fernando Brandao and Aram Harrow shortly later. Realising that they were not so much older than me, while the works of theirs that I knew had made me expect old respectful professors, was naturally a bit depressing at first. But them as well have had from then on such a supportive and interactive attitude towards me that I do not hold it against them! David Reeb and Alexander Müller-Hermes are two other scientists I met in my early mathematical youth, and with whom I had constant interchanges ever after. It is indeed a quite pleasant feeling to realise that there are at least two other persons in the world who care about the same insignificant (but annoyingly open) questions as you! The Fall 2013 semester that I spent participating to the thematic programme "Mathematical challenges in quantum information" at the Isaac Newton Institute in Cambridge, at the very beginning of my PhD, was the occasion to make several influential encounters. The one with Nilanjana Datta and Debbie Leung was for sure determining. I would even go as far as seeing it as me projecting myself onto how I would dream to evolve in my future scientific behaviour! The ones with Matthias Christandl and Michael Walter were definitely shaping as well. The two of them patiently filled my group representation gaps and cheerfully kept suggesting new directions when absolutely everything we were working on was collapsing! Subsequent collaborations with Matthias have always been of that same joyful, hence most

pleasant, vein.

Je suis extrêmement reconnaissante envers Ion Nechita pour n'avoir jamais failli à son rôle de mentor. Il m'a en effet dès mes débuts chapeautée, et ce à tous points de vue: passant sans problème des journées entières à répondre à mes questions mathématiques, me mettant sur le devant de la scène lors d'à peu près chacune des conférences qu'il a organisées, et me faisant même découvrir la folle vie nocturne roumaine! Je profite d'ailleurs de cette occasion pour remercier aussi Maria Anastasia Jivulescu, organisatrice en chef (et surtout la plus attentionnée qu'on puisse concevoir) de cette mémorable conférence en Roumanie! Toujours par association d'idées, merci aux autres StoQistes toulousains, Clément Pellegrini et Tristan Benoist, pour leur disponibilité (intra ou extra scientifique) absolument sans faille. Enfin, Benoit Collins et Staszek, mon grand-père mathématique, ont eux aussi été des interlocuteurs constamment disponibles, avec qui j'apprécie toujours échanger des idées et auprès desquels j'ai énormément appris.

The MathQI group in Madrid should be my next home, and I am pretty much looking forward to it, for the very simple reason that I actually already feel at home there. From my first encounter with Angelo, Carlos G.G., Carlos P.C., David P.G., Ignacio, and friends, during the "Intensive month on operator algebras and quantum information" that they were organizing at the ICMAT in Summer 2013, I already knew there were the best conceivable vibes between us, surpassing by far the professional collaboration standpoint!

Ces trois années de thèse (entre bien d'autres choses) n'auraient pas eu la même saveur sans le groupe de grimpeurs de la promo X2009 et ses pièces rapportées (que nous appellerons par la suite Plastik par souci de concision). Je remercie donc Plastik du fond du cœur pour avoir accepté qu'on peut taper une preuve en TeX à 6h du matin et être néanmoins quelqu'un de fréquentable. Plastik, c'est un peu ma deuxième famille, dont je resterai (d'un point de vue scientifique en tout cas) toujours la Maman, puisque première de la troupe à accéder au grade de Docteur. Votre soutien pré-soutenance m'a particulièrement touchée: vos suggestions sur quoi faire le jour J et comment se détendre avant n'ont pas forcément toutes été mises en œuvre, mais ont été appréciées pour leur inventivité! A Guillaume et Harmonie, un immense merci pour votre relecture orthogonalement supplémentaire de ce manuscrit: Harmonie étant probablement hermétique à la beauté mathématique de mes équations (bien que foncièrement convaincue de leur génialitude) mais repérant les moindres fautes d'anglais, Guillaume ne me corrigeant absolument rien mais s'enthousiasmant d'absolument tout!

Enfin, je tiens à dire un grand merci à ma famille, qui a toujours eu la curiosité d'essayer de comprendre ce que je pouvais bien passer mes journées à faire. Merci en particulier à Grounchy et Schnoky pour leur soutien indéfectible: "Alors cet ordi quantique, ça avance?", "Ah ta preuve marche, c'est cool! On sait très bien que demain elle marchera plus, mais t'en fais pas, après-demain elle re-marchera...". And last but not least, merci à mes parents pour m'avoir appris très tôt à faire la différence entre un Banach et un canard. Ils ont bien tenté de me suggérer que j'aurais probablement beaucoup mieux à faire de ma vie que de prouver (voire le plus souvent même pas) des théorèmes dont trois personnes au monde se soucient. Mais il faut croire que l'image qu'ils m'ont renvoyé du chercheur en maths n'a pas été suffisamment décourageante pour me convaincre...

Table of Contents

Page

I Introduction and background . . . 15

Chapter 1 What is this thesis about and how is this manuscript organized? . . . 19

Chapter 2 The mathematics of quantum information theory in a nutshell . . . 23

Chapter 3 Permutation-symmetry and de Finetti type theorems . . . 29

Chapter 4 Asymptotic geometric analysis toolbox . . . 33

II Complexity reduction in quantum information theory . . . 39

Chapter 5 Quantum channel compression . . . 45

Chapter 6 Zonoids and sparsification of quantum measurements . . . 59

III Some aspects of generic entanglement: data-hiding and relaxations of separability in high dimensions . . . 75

Chapter 7 Locally restricted measurements on multipartite quantum systems . . . 81

Chapter 8 Relaxations of separability in multipartite quantum systems . . . 103

Chapter 9 k-extendibility of high-dimensional bipartite quantum states . . . 117

IV Making use of permutation-symmetry to tackle multiplicativity issues . . . 155

Chapter 10 Flexible constrained de Finetti reductions and applications . . . 161

Chapter 11 Parallel repetition and concentration for (sub-)no-signalling games . . . 181

V Outlook and perspectives . . . 193

Bibliography . . . 208

Part I

Introduction and background

Part I – Table of contents

Chapter 1 What is this thesis about and how is this manuscript organized? 19
1.1 Motivations and context of the thesis 19
1.2 Structure of the present manuscript 20
1.3 Notation and conventions 21

Chapter 2 The mathematics of quantum information theory in a nutshell 23
2.1 Quantum states and observables: preparation and measurement 23
2.2 Composite quantum systems: separability vs entanglement 24
2.3 Evolution of a quantum system 25
2.4 Separability criteria on multipartite quantum systems 25
2.5 A few notions from quantum Shannon theory 26

Chapter 3 Permutation-symmetry and de Finetti type theorems 29
3.1 The symmetric subspace: a brief review of some basic notions and facts 29
3.2 Permutation-invariant states in the bipartite finite-dimensional case 30
3.3 Classical and quantum de Finetti type theorems 30

Chapter 4 Asymptotic geometric analysis toolbox 33
4.1 Classical convex geometry 33
4.2 Concentration of measure and deviation inequalities 36

Chapter 1

What is this thesis about and how is this manuscript organized?

1.1 Motivations and context of the thesis

A one-particle quantum system in a pure state is described by a unit vector in some Hilbert space (or to be precise, by the projection onto the corresponding line). For multi-particle systems, the associated Hilbert space is simply given by the tensor product of the individual ones. Its dimension thus grows exponentially with the number of subsystems, making the classical modelling of quantum systems practically unfeasible as soon as more than a few particles are involved.

One way around this curse of dimensionality is to make use of any extra information which may be held, a priori, on the state of the system under consideration. One could know, for instance, that the latter has certain symmetries, and exploit the reduction in the effective number of degrees of freedom that this fact implies. The archetypical example is the following: Assume that a system is composed of n indistinguishable particles (meaning that their labelling is irrelevant), with associated finite-dimensional Hilbert space H. Then, a pure state of such system actually lives in the symmetric subspace $\mathrm{Sym}^n(\mathrm{H})$ of $\mathrm{H}^{\otimes n}$, whose dimension is only polynomial, rather than exponential, in n.

On the other hand, one should not conclude too quickly that having to deal with high dimensional objects is necessarily a misfortune. Indeed, it may also be the case that, as the dimension grows, certain universal behaviours emerge. This is precisely what asymptotic geometric analysis and random matrix theory teach us. In quantum information theory, these subfields of mathematics can serve at least two main purposes. First of all, they may be used as a tool to determine what are the properties that big quantum systems, subject to whatever relevant restrictions, are expected to exhibit. More concretely, one could be interested in knowing what are the typical characteristics of either quantum states or quantum transformations, under several types of constraints such as locality, noise, energy etc. Second of all, they may help in proving the existence of objects having a given property. In this context, the paradigmatic idea is that constructing the latter explicitly might be harder than asserting that a suitably chosen random one will do with overwhelming probability. And with both aims in view, high dimension of the underlying space is an asset, what makes average features become generic.

The purpose of this thesis is therefore clearly two-sided: in some places the focus is on how to reduce the study of large (or even infinite) dimensional situations to that of lower dimensional ones, while in some others the goal is precisely to understand the typical aspects which may arise as the dimension of the studied object grows. The latter objective is clearly the underlying motivation of Part III. As for the former objective, two different routes can be taken to achieve it. The first approach, which is the one followed in Part II, consists in trying to compress the initial data set as much as possible while still preserving the essential information it contains. The second approach consists in exploiting potential additional knowledge on the initial data to argue that they already belong, effectively, to a much smaller set. This is the spirit of Part IV.

All the questions posed here initially arise from quantum information theory. But in order to solve them, several fields of pure mathematics had to get heavily involved. For a start, random matrix theory plays undeniably a prominent role in all our work. Nevertheless, we are not so much preoccupied with the asymptotic study of random matrices, as (free) probabilists usually are, but rather with the non-asymptotic one. Indeed, for the

applications we have in mind, knowing that a certain behaviour emerges almost surely in the regime where the size of the matrix goes to infinity is not so useful: we need instead to understand, for a large but finite size, what is the probability that the properties of the matrix do not deviate too much from their limiting ones. This is where concentration of measure enters the picture, allowing us to assert that, if the function we are looking at is regular enough, then it should be close to its expected value with overwhelming probability as the size of its random input grows. We exploit this key phenomenon under multiple forms throughout these pages. It however appears clear that, before even trying to bound the deviation probability of a given function (and more often than not, also as an end in itself), we have to be able to estimate its average value. For that purpose, we usually call in geometric and probabilistic techniques in Banach space theory. Besides, taking advantage of the symmetries of our problem, is an idea that we try to apply whenever possible. With that goal in view, information theory tools, both classical and quantum, are crucial to us. They are in fact what makes it for instance possible to relate in a quantitative way exchangeability and independence.

1.2 Structure of the present manuscript

Each chapter revolves around a piece of work which either already led to a paper (sometimes published yet, sometimes in preprint form still) or should lead to one shortly. In each case though, modifications have been done on the original version. Occasionally it is just a matter of notation or presentation, to simply fit better with the rest of the manuscript (unfortunately, it is likely that neither complete non-redundancy nor absolute style unity have been achieved between the chapters). But here and there further developments, that were investigated only afterwards, have also been added.

The remainder of this introductory part is dedicated to setting the common ground and tools on which subsequent parts develop. If it achieves the goal it was designed for, Chapter 2 could be renamed "Everything you need to know about quantum physics if you are a mathematician willing to read this manuscript". As for Chapters 3 and 4, they can be seen as the two toolboxes in which we will regularly dig: in Chapter 3 for all the basics about permutation-symmetry, and in Chapter 4 for all the basics about high dimensional convex geometry and deviation inequalities.

As already mentioned, the core scientific material in this thesis is then divided into three parts. Let us try to summarize in a few words the content of each of them, even though the needed definitions in order to do so in a satisfactory way have not been introduced yet (see Chapter 2 for most of them). Precisely for that matter, it is only at the beginning of each part that a proper account is made of its objectives, its main achievements (and techniques to reach them), its non-ignorable difficulties etc.

The main theme of Part II is that of approximating complex processes by simpler ones. In Chapter 5 the question which is investigated is: given a quantum channel, to how much can its number of Kraus operators be brought down, with the constraint that each input state is still sent close to its original output state? In some sense, Chapter 6 deals with a similar issue. Nevertheless, the channels which are now considered are quantum-classical channels (aka quantum measurements), and the approximation requirement is completely different: namely that, for each input state, outcome statistics close to its original ones are obtained. Our objective in both cases is to exhibit universal schemes which attain the maximum doable compression.

The goal of Part III is to study several properties of quantum states (more specifically, entanglement-related properties of multipartite quantum states), and see how (un)common they become as the size of the quantum system grows. Chapter 7 is in line with Chapter 6 since it also focusses on the issue of distinguishing quantum states from their measurement outcome statistics. But this time the question is: how well can local observers, having access only to their own subsystem, typically perform in this task? Then, in both Chapters 8 and 9, the aim is to quantify the average strength of certain semidefinite relaxations of separability. In Chapter 8, stress is put on exporting entanglement detection tools from the bipartite setting, relying on positive maps, to certify genuinely multipartite entanglement. Oppositely, the interest in Chapter 9 is in a purely bipartite and not positive map based separability criterion (more precisely, a complete hierarchy of necessary conditions for bi-separability build on symmetric extensions).

This opens naturally the route to Part IV, where symmetries are exploited to their maximum, with the purpose of reducing the understanding of permutation-invariant scenarios to that of i.i.d. ones. Hence, a major difference with Part II is that we are now making extensive use of the specificities of the situation we have at hand in order to get a highly problem-dependent simplification. Chapter 10 is dedicated to setting a very general and adaptable framework in which one can indeed do so. As a consequence, several quantities arising

in quantum information theory can be shown to exhibit a multiplicative behaviour. Chapter 11 is then entirely devoted to seeing through one important application of the previously developed machinery, namely to the parallel repetition problem for multi-player non-local games.

Perhaps now is the time to fleetingly elaborate on the choice of the thesis title. There are two distinct ways in which we are principally dealing with "high dimension": either because we are looking at one system whose underlying dimension is large, or because we are looking at many copies of a system. Yet, in both situations "symmetries" play a central role. To attack the first problem, it is usually tools from asymptotic geometric analysis that we call on for help. This field of mathematics can be seen as a middle ground between geometry (which traditionally deals with small-dimensional objects) and functional analysis (which traditionally deals with infinite-dimensional objects). And its whole purpose is to identify and exploit approximate symmetries that escaped both the too rigid area of geometry and the too qualitative area of functional analysis. As for the second issue, it is usually tackled via information theory techniques. And it is more often than not making use of the permutation-symmetry of the multi-copy scenario which allows relating asymptotic performances to single-shot ones, and hence get an understanding of them.

This manuscript essentially puts together the material appearing in the following publications or preprints:

- C. Lancien. Quantum channel compression.
  In preparation. (cf. Chapter 5)
- G. Aubrun and C. Lancien. Zonoids and sparsification of quantum measurements.
  Positivity, 20(1):1–23, 2016. arXiv[quant-ph]:1309.6003 (cf. Chapter 6)
- G. Aubrun and C. Lancien. Locally restricted measurements on a multipartite quantum system: data hiding is generic.
  Quant. Inf. Comput., 15(5–6):512–540, 2014. arXiv[quant-ph]:1406.1959. (cf. Chapter 7)
- O. Gühne, M. Huber, C. Lancien and R. Sengupta. Relaxations of separability in multipartite systems: semidefinite programs, witnesses and volumes.
  J. Phys. A: Math. Theor., 48(505302), 2015. arXiv[quant-ph]:1504.01029. (cf. Chapter 8)
- C. Lancien. k-extendibility of high-dimensional bipartite quantum states.
  Preprint. arXiv[quant-ph]:1504.06459. (cf. Chapter 9)
- C. Lancien and A. Winter. Flexible constrained de Finetti reductions and applications.
  Preprint. arXiv[quant-ph]:1605.09013. (cf. Chapter 10)
- C. Lancien and A. Winter. Parallel repetition and concentration for (sub-)no-signalling games via a flexible constrained de Finetti reduction.
  Preprint. arXiv[quant-ph]:1506.07002. (cf. Chapter 11)

Other publications or preprints on which this manuscript does not focus:

- C. Lancien and A. Winter. Distinguishing multi-partite states by local measurements.
  Commun. Math. Phys., 323:555–573, 2013. arXiv[quant-ph]:1206.2884.
- G. Adesso, S. Di Martino, M. Huber, C. Lancien, M. Piani and A. Winter. Should entanglement measures be monogamous or faithful?
  Preprint. arXiv[quant-ph]:1604.02189.

1.3 Notation and conventions

Here is a list of notation that shall be used repeatedly throughout the whole manuscript (it might happen though that some are recalled as the text goes). Oppositely, notation which are needed more sporadically will be introduced only in due time.

Given a complex Hilbert space H, we denote by $\mathscr{L}(\mathrm{H})$ the space of all linear operators on H. When $\mathrm{H} \equiv \mathbf{C}^d$ is finite-dimensional (which will almost always be the case we will be dealing with), we identify $\mathscr{L}(\mathrm{H})$ with the space of $d \times d$ complex matrices. Important subsets of $\mathscr{L}(\mathrm{H})$ include: the group of unitary operators on H, which we denote by $\mathscr{U}(\mathrm{H})$, the space of Hermitian operators on H, which we denote by $\mathscr{H}(\mathrm{H})$, the cone of positive semidefinite operators on H, which we denote by $\mathscr{H}_+(\mathrm{H})$. The notation Id (or Id_H if there is a risk of confusion) stands for the identity operator on H, while the notation $\mathscr{Id}$ (or $\mathscr{Id}_\mathrm{H}$) is used for the identity (super-)operator

on $\mathscr{L}(\mathrm{H})$. Also, for any $X \in \mathscr{L}(\mathrm{H})$, we denote by $X^\dagger$ its adjoint (i.e. transpose conjugate) and by $|X| = \sqrt{X^\dagger X}$ its absolute value.

For each $p \in [1, +\infty]$, we define $\|\cdot\|_p$ as the Schatten p-norm on $\mathscr{L}(\mathrm{H})$, i.e.

$$\forall\ X \in \mathscr{L}(\mathrm{H}),\ \|X\|_p = (\mathrm{Tr}\,|X|^p)^{1/p} \text{ if } p \in [1, +\infty[\text{ and } \|X\|_\infty = \lim_{p \to +\infty} \|X\|_p.$$

Particular instances of interest are the trace class norm $\|\cdot\|_1$, the Hilbert–Schmidt norm $\|\cdot\|_2$ (which is nothing else than the Euclidean norm arising from the Hilbert–Schmidt inner product $\langle X, Y\rangle \mapsto \mathrm{Tr}\left(X^\dagger Y\right)$ on $\mathscr{L}(\mathrm{H})$), and the operator norm $\|\cdot\|_\infty$. Most of the time, we will look at Schatten p-norms not on the full space $\mathscr{L}(\mathrm{H})$ but in restriction to its subspace $\mathscr{H}(\mathrm{H})$. We denote by $B_p(\mathrm{H})$, resp. $S_p(\mathrm{H})$, the corresponding unit ball, resp. sphere, in $\mathscr{H}(\mathrm{H})$. In the special case $p = 2$, we may also use the notation $B_{HS}(\mathrm{H})$ and $S_{HS}(\mathrm{H})$ instead.

We denote by $\|\cdot\|_p$ as well the p-norms on $\mathbf{R}^d$ or $\mathbf{C}^d$. One exception is again the Euclidean norm $\|\cdot\|_2$, for which we shall usually simply use the notation $\|\cdot\|$, while S^{d-1} and $S_{\mathbf{C}^d}$ stand for the corresponding Euclidean unit spheres in $\mathbf{R}^d$ and $\mathbf{C}^d$, respectively.

The notation $|\cdot|$ stands both for the cardinality when applied to a finite set and for the dimension when applied to a finite-dimensional Hilbert space.

Given $n \in \mathbf{N}$, we may use the shorthand notation $[n]$ for $\{1, \ldots, n\}$, and we denote by $\mathfrak{S}(n)$, resp. $\mathfrak{P}(n)$, the set of permutations, resp. partitions, of $[n]$.

In the remainder of this manuscript, we are often interested in the asymptotic regime, when the dimensions of the underlying finite-dimensional Hilbert spaces tend to infinity. In that setting, the letters $C, c, c_0, \ldots$ will always denote (non-negative) numerical constants, independent from any other parameters such as the dimension. The value of these constants may change from occurrence to occurrence. Similarly $c(\varepsilon)$ denotes a constant depending only on the parameter ε. Also, given functions f, g of the underlying dimension d, we will write $f = O(g)$, resp. $f = \Omega(g)$, if there exists $C > 0$ such that, for all $d \in \mathbf{N}$, $f(d) \leqslant Cg(d)$, resp. $f(d) \geqslant g(d)/C$, and we will write $f = \Theta(g)$ if both $f = O(g)$ and $f = \Omega(g)$ hold.

When working with a random variable X, we will use the notation $\mathbf{P}(\mathscr{E}(X))$ to denote the probability of the event $\mathscr{E}(X)$, and the notation $\mathbf{E}\left(f(X)\right)$ to denote the expectation of the function $f(X)$.

More quantum information orientated notation include (see Chapter 2, Sections 2.1 and 2.2, for the corresponding definitions): $\mathscr{D}(\mathrm{H})$ for the set of quantum states on the Hilbert space H, $\mathcal{S}(\mathrm{H}_1 : \cdots : \mathrm{H}_k)$ for the set of separable quantum states on the tensor product Hilbert space $\mathrm{H}_1 \otimes \cdots \otimes \mathrm{H}_k$ (across the k-partite cut $\mathrm{H}_1 : \cdots : \mathrm{H}_k$).

Chapter 2

The mathematics of quantum information theory in a nutshell

The purpose of this chapter is to explain why certain mathematical notions enter the theory of quantum mechanics, by giving an idea of their physical interpretation. There is no attempt to being exhaustive, far from that, the accent being put only on concepts that will then play a central role throughout this manuscript. The reader is for instance referred to the lecture notes [177], Chapter 1, 2 and 3, for a much more comprehensive justification of the correspondence between mathematical objects and physical situations in the quantum mechanical framework. The book [14], Chapters 2 and 3, would be another recommendation for going way deeper into that topic.

2.1 Quantum states and observables: preparation and measurement

In quantum mechanics, the state of a system is described by a unit vector ψ in a complex separable Hilbert space $(\mathrm{H}, \langle\cdot|\cdot\rangle)$. The physical picture is then the following: There exists a distinguished orthonormal basis $\{e_i,\ i \in I\}$ of H, where I is countable (finite or infinite) and labels the possible "levels" of the system. If $\psi = e_i$ for some $i \in I$, then (as one would expect) the system is said to be in level i. But in general, ψ is in a superposition of the basis vectors, i.e.

$$\psi = \sum_{i \in I} \langle e_i|\psi\rangle e_i,$$

And the only thing that one can tell is that the probability of obtaining outcome $i \in I$ when measuring its level is given by $|\langle e_i|\psi\rangle|^2$. We already see with this description that, for any $\alpha \in \mathbf{C}$ such that $|\alpha|^2 = 1$, states ψ and $\alpha\psi$ cannot be distinguished. It is therefore more accurate to say that the state of the quantum system is characterized by the rank-1 projection $|\psi\rangle\langle\psi|$ on H, which reads

$$|\psi\rangle\langle\psi| = \sum_{i,j \in I} \langle e_i|\psi\rangle\langle\psi|e_j\rangle |e_i\rangle\langle e_j|.$$

And hence, simply rewriting what we just explained, the probability of obtaining outcome $i \in I$ when measuring its level is given by $\mathrm{Tr}\left(|e_i\rangle\langle e_i||\psi\rangle\langle\psi|\right)$.

This describes well the situation where the physical system under consideration can be perfectly prepared and kept in any given target state ψ. Such scenario is of course completely idealistic, and in a more realistic framework noise has to be taken into account. The latter may arise either from imprecisions in the preparation procedure or from interactions of the system of interest with its environment. In both cases, one can only assign probabilities $\{p_x,\ x \in X\}$ to the system being in one amongst the so-called *pure states* $\{|\psi_x\rangle\langle\psi_x|,\ x \in X\}$. And the true state of the system is the so-called *mixed state* ρ defined as

$$\rho = \sum_{x \in X} p_x |\psi_x\rangle\langle\psi_x|.$$

Here again, the formula for the probability of getting outcome $i \in I$ when measuring the system's level reads $\mathrm{Tr}\left(|e_i\rangle\langle e_i|\rho\right)$.

Hence, probabilities appear at two very distinct stages in quantum mechanics: Just as in classical mechanics, they model our ignorance of the precise state in which the system is. But they also occur, more fundamentally, because even having perfect knowledge of the system's state, we cannot (in general) predict with certainty which value a measurement performed on it will yield. That is why the measurement process, as a transition from possibilities to facts, plays such a central role in the quantum theory.

Up to now, we mentioned only one specific measurement, namely the one performed in the canonical basis of H, which determines the level of the system. Its action on a state ρ consists in sending it on one amongst the pure states $\{|e_i\rangle\langle e_i|,\ i \in I\}$, with respective probabilities $\{\mathrm{Tr}\left(|e_i\rangle\langle e_i|\rho\right),\ i \in I\}$. Still looking at this particular measurement, one could however consider the case where it cannot be performed with such perfect accuracy. For instance, it could be that levels which are too close to one another are simply seen as being the same, or that certain outcomes are sometimes mistaken for one another. The resulting "blurred" measurement gives outcomes labelled by $j \in J$, and its action on a state ρ consists in sending it on one amongst the mixed states $\{M_j/\operatorname{Tr} M_j,\ j \in J\}$, with respective probabilities $\{\mathrm{Tr}(M_j\rho),\ j \in J\}$, where for each $j \in J$, there exists a probability distribution $\{p_{j,i},\ i \in I\}$ such that $M_j = \sum_{i\in I} p_{j,i}|e_i\rangle\langle e_i|$.

Let us summarize and formalize the above discussion, focussing for simplicity on the finite-dimensional case, which is the one we shall almost exclusively consider in the sequel. The Hilbert space associated to a d-level quantum system can simply be identified with $\mathbf{C}^d$, equipped with its canonical inner product. The state of such quantum system is entirely characterized by a *density operator* ρ on $\mathbf{C}^d$, which is nothing else than a convex combination (or mixture) of rank-1 projectors on $\mathbf{C}^d$ (or pure states on $\mathbf{C}^d$). Equivalently, ρ has to satisfy the two properties of being positive semidefinite and having trace 1. Denoting by $\mathscr{H}(\mathbf{C}^d)$ the set of Hermitian operators on $\mathbf{C}^d$, the set of density operators on $\mathbf{C}^d$ is thus defined as

$$\mathscr{D}(\mathbf{C}^d) := \mathrm{conv}\{|\psi\rangle\langle\psi|,\ \psi \in \mathbf{C}^d,\ |\langle\psi|\psi\rangle|^2 = 1\} = \{\rho \in \mathscr{H}(\mathbf{C}^d),\ \rho \geqslant 0,\ \mathrm{Tr}\,\rho = 1\}.$$

Besides, a measurement on such quantum system is described by a *Positive Operator-Valued Measure (POVM)*, which is a resolution of the identity on $\mathbf{C}^d$, i.e. a set of positive semidefinite operators summing to the identity on $\mathbf{C}^d$.

Understanding the geometry of the set $\mathscr{D}(\mathbf{C}^d)$ and of certain of its subsets is a wide topic, upon which we shall touch in several places of this manuscript (see [27] or [14], Chapters 2 and 9, for a comprehensive exposition). For now, let us just state some immediate facts. $\mathscr{D}(\mathbf{C}^d)$ is a compact convex set whose extreme points are exactly the pure states, i.e. the rank-1 states. It is included in $\{M \in \mathscr{H}(\mathbf{C}^d),\ \mathrm{Tr}\,M = 1\}$, hyperplane of $\mathscr{H}(\mathbf{C}^d)$, where it has non-empty interior. $\mathscr{D}(\mathbf{C}^d)$ is therefore a convex body of real dimension $d^2 - 1$ (recall that $\mathscr{H}(\mathbf{C}^d)$ has real dimension d^2). Its center of mass is the so-called *maximally mixed state* Id/d.

2.2 Composite quantum systems: separability vs entanglement

Assume now that the quantum system under consideration is composed of k subsystems, with associated Hilbert spaces $\mathrm{H}_1, \ldots, \mathrm{H}_k$. Then, the Hilbert space associated to the global k-partite system is simply the tensor product Hilbert space $\mathrm{H} = \mathrm{H}_1 \otimes \cdots \otimes \mathrm{H}_k$. A state ρ on H is called *separable* if it can be written as a convex combination of product states, i.e. states of the form $\rho_1 \otimes \cdots \otimes \rho_k$ where $\rho_1, \ldots, \rho_k$ are states on $\mathrm{H}_1, \ldots, \mathrm{H}_k$ respectively. Otherwise, it is called *entangled*.

As we shall later see on several occasions, the dichotomy between separability and entanglement is fundamental in quantum information theory (see in particular Chapters 7, 8, 9 and 10). So let us start here with a few basic observations. If a pure state is separable, then it is necessarily product, which means that its local subsystems are completely uncorrelated. In the general mixed case, a separable state may have local subsystems which exhibit correlations, but classical ones only (in the sense that such state can always be prepared by local parties sharing just common randomness).

In the finite-dimensional case, the set of separable states on $\mathbf{C}^{d_1} \otimes \cdots \otimes \mathbf{C}^{d_k} \equiv \mathbf{C}^d$ (i.e. $d = d_1 \times \cdots \times d_k$) is thus defined as

$$\mathscr{S}(\mathbf{C}^{d_1}{:}\cdots{:}\mathbf{C}^{d_k}) := \mathrm{conv}\left\{\rho_1 \otimes \cdots \otimes \rho_k,\ \forall\ 1 \leqslant i \leqslant k,\ \rho_i \in \mathscr{D}(\mathbf{C}^{d_i})\right\}.$$

And the extreme points of $\mathscr{S}(\mathbf{C}^{d_1}{:}\cdots{:}\mathbf{C}^{d_k})$ are actually easy to characterize: these are precisely the pure separable states, i.e. the pure product states. Just as $\mathscr{D}(\mathbf{C}^d)$, $\mathscr{S}(\mathbf{C}^{d_1}{:}\cdots{:}\mathbf{C}^{d_k})$ has real dimension $d^2 - 1$ and its center of mass is the maximally mixed state Id/d (here again, see [27] or [14], Chapters 2 and 9, for a much more exhaustive presentation).

For the sake of clarity, let us focus for now on the bipartite case $\mathrm{H} = \mathrm{H}_1 \otimes \mathrm{H}_2$. Given a state ρ on H, we can define its *reduced state* on the first subsystem $\rho_1 = \mathrm{Tr}_2\,\rho$, which is the state on H_1 characterized by the property that, for any operator M_1 on H_1, $\mathrm{Tr}(\rho_1 M_1) = \mathrm{Tr}(\rho\, M_1 \otimes \mathrm{Id}_2)$. Conversely, given a state ρ_1 on H_1, we say that a state ρ on H is an *extension* of ρ_1 if $\mathrm{Tr}_2\,\rho = \rho_1$, and that it is more precisely a *purification* of ρ_1 if it is additionally pure. For any state ρ_2 on H_2, the state $\rho_1 \otimes \rho_2$ is obviously an extension of ρ_1, and when the latter is pure its extensions are in fact necessarily of this product form (see e.g. [51], Chapter 2). This means in other words that a system in a pure state cannot share any kind of correlations (not even classical ones) with another system.

2.3 Evolution of a quantum system

When talking about the time evolution of a physical system (classical or quantum), two cases have to be distinguished: that of *closed systems* and that of *open systems*. A closed system is one which is considered perfectly isolated from the outside world, and thus undergoes reversible dynamics only. On the contrary, when studying an open system, one takes into account that it may interact with its surrounding, so that irreversibility may arise from this coupling (it is only the system of interest accompanied by its environment which is seen as a whole as closed).

Let us see how these ideas are mathematically formalized when looking at a quantum system, with associated Hilbert space H. If the system is considered closed, the evolutions it may undergo are unitary transformations. This means that there exists a unitary operator U on H such that, if it is in the initial state ρ, then it is brought in the final state $U\rho U^\dagger$. If this time the system is considered open, coupled with some ancilla Hilbert space K, it is the global system $\mathrm{H}\otimes\mathrm{K}$ which is subject to a unitary transformation. Hence, there exist a unitary operator U on $\mathrm{H}\otimes\mathrm{K}$ and a pure state $|\psi\rangle\langle\psi|$ on K such that, at the level of the system of interest, if it is in the initial state ρ, then it is brought in the final state $\mathrm{Tr}_\mathrm{K}\left(U\rho\otimes|\psi\rangle\langle\psi|U^\dagger\right)$.

A crucial observation is that these two situations can be encompassed into one common description, by saying that the evolution of a quantum system is, in full generality, characterized by a so-called *completely positive and trace preserving* (CPTP) map [168]. Let us explain what this means. A linear map $\mathcal{N}$ from operators on H to operators on H′ is *positive* (P) if, for any positive semidefinite operator X on H, $\mathcal{N}(X)$ is a positive semidefinite operator on H′. And it is *completely positive* (CP) if, for any K, the linear map $\mathcal{N}\otimes\mathcal{I}d$ from operators on $\mathrm{H}\otimes\mathrm{K}$ to operators on $\mathrm{H}'\otimes\mathrm{K}$ is positive. Besides, it is called *trace preserving* (TP) if, as the name suggests, for any trace-class operator X on H, $\mathrm{Tr}\,\mathcal{N}(X) = \mathrm{Tr}\,X$. A CPTP map $\mathcal{N}$ is therefore usually referred to as a *quantum channel*, transforming the input state ρ into the output state $\mathcal{N}(\rho)$.

These definitions will be crucial to us in Chapter 5, and to some extent also in Chapter 10. Note that the measurement procedure described before falls into this category: a POVM is a particular type of quantum channel, which maps quantum states to probability distributions, and is thus sometimes called a quantum-classical (QC) channel. These QC channels will be thoroughly studied in Chapters 6 and 7. Oppositely, a preparation procedure, which consists in mapping a probability distribution to a mixture of quantum states, is sometimes called a classical-quantum (CQ) channel. And the fully classical analogue of these notions is simply that of a stochastic map (aka conditional probability distribution), which will play a central role in Chapter 11.

2.4 Separability criteria on multipartite quantum systems

Already in the bipartite case, deciding whether a given state is entangled or separable is known to be, in general, a hard task, both from a mathematical and computational point of view. In fact, even the weak membership problem for separability (i.e. deciding if the state under consideration is far away from or close to the set of separable states) is a NP-hard problem [90, 80]. There exist necessary conditions for separability, though, which are much easier to check. A whole family of them is based on the following easy observation: if ρ is a separable state on $\mathrm{H}_1 \otimes \mathrm{H}_2$, then for any positive map $\mathcal{N}_1$ from operators on H_1 to operators on H_1', $\mathcal{N}_1 \otimes \mathcal{I}d_2(\rho)$ is a positive semidefinite operator on $\mathrm{H}_1 \otimes \mathrm{H}_2$.

Widely studied and used examples of positive (yet not completely positive) maps, giving rise to a corresponding (non trivial) necessary condition for separability, include:

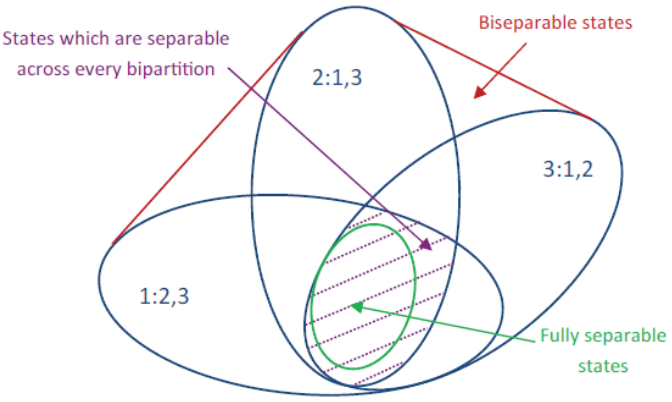
- The transposition map $\mathcal{T} : X \in \mathcal{L}(\mathrm{H}) \mapsto X^T \in \mathcal{L}(\mathrm{H})$ [146, 108].

- The reduction map $\mathscr{R} : X \in \mathscr{L}(\mathrm{H}) \mapsto (\mathrm{Tr}\, X)\mathrm{Id} - X \in \mathscr{L}(\mathrm{H})$ [107].
- The Choi map $\mathscr{C} : X \in \mathscr{L}(\mathrm{H}) \mapsto -X + (|\mathrm{H}| - 1)\Delta(X) + \Delta'(X) \in \mathscr{L}(\mathrm{H})$, where the action of Δ and Δ' on any $X \in \mathscr{L}(\mathrm{H})$ is defined by, for each $1 \leqslant i, j \leqslant |\mathrm{H}|$, $\Delta(X_{i,j}) = \delta_{i,j} X_{i,i}$ and $\Delta'(X_{i,j}) = \delta_{i,j} X_{i-1,i-1}$ [47].

Taken altogether, these separability tests based on positive maps actually define a necessary and sufficient condition for separability: a state ρ on $\mathrm{H}_1 \otimes \mathrm{H}_2$ is separable if and only if, for all positive maps $\mathscr{N}_1$ from operators on H_1 to operators on H_1', $\mathscr{N}_1 \otimes \mathscr{I}\mathscr{d}_2(\rho)$ is a positive semidefinite operator on $\mathrm{H}_1' \otimes \mathrm{H}_2$ [46]. This result is a fundamental one in quantum information theory. It translates in the language of so-called *entanglement witnesses* as follows: If a state ρ on $\mathrm{H}_1 \otimes \mathrm{H}_2$ is entangled, then there exists a block-positive operator M on $\mathrm{H}_1 \otimes \mathrm{H}_2$ such that $\mathrm{Tr}(\rho M) < 0$ [108]. M is thus said to witness the entanglement of ρ, since for any separable state σ on $\mathrm{H}_1 \otimes \mathrm{H}_2$, we have $\mathrm{Tr}(\sigma M) \geqslant 0$. In other words, the operator M defines a hyperplane in $\mathscr{H}(\mathrm{H}_1 \otimes \mathrm{H}_2)$ which separates the convex body $\mathcal{S}(\mathrm{H}_1{:}\mathrm{H}_2)$ from the point ρ. One given entanglement witness can only detect the entanglement in a small fraction of states (those on the same side as ρ of the hyperplane it defines). However, entanglement witnesses carry the advantage of being experimentally easily accessible.

In the multipartite case, the picture becomes even more complex. Indeed, checking separability of a given state across all bipartite splitting of the subsystems is not enough to guarantee that it is fully separable. Motivated by this easy observation, one can define a hierarchy of relaxations of the latter notion: a state on $\mathrm{H}_1 \otimes \cdots \otimes \mathrm{H}_k$ is ℓ-separable, where $2 \leqslant \ell \leqslant k$, if it is a convex combination of states, each of which is separable across a given splitting of the k subsystems into ℓ groups (note that, in general, there is no splitting of the k subsystems into ℓ groups across which an ℓ-separable state is separable). k-separability is then just full separability, while a state which is not 2-separable is called *genuinely multipartite entangled*. These various notions of separability are illustrated in the tripartite case $\mathrm{H} = \mathrm{H}_1 \otimes \mathrm{H}_2 \otimes \mathrm{H}_3$ in Figure 2.1 (where $x : y, z$ denotes the set of states on H which are separable across the bipartite cut $\mathrm{H}_x{:}\mathrm{H}_y \otimes \mathrm{H}_z$).

Figure 2.1: Several degrees of separability on a tripartite system



All these notions concerning entanglement detection will be crucial in several places of this manuscript, most notably in Chapters 8 and 9, but also to a lesser extent in Chapters 7 and 10.

2.5 A few notions from quantum Shannon theory

Classical information theory is concerned with quantifying the amount of information and correlations present in probability distributions. By analogy, quantum information theory is concerned with doing the same for density operators. We gather here some basic definitions and facts from the quantum Shannon theory that we shall later need here and there. The reader is for instance referred to the book [176], Chapter 11, for a complete expounding of these notions.

The *von Neumann entropy* of a quantum state ρ is defined as

$$S(\rho) := -\mathrm{Tr}(\rho \log \rho).$$

It is always non-negative, and equal to 0 if and only if ρ is a pure state. In the case where the underlying Hilbert space H is finite-dimensional, it is upper-bounded by $\log |\mathrm{H}|$, with equality if and only if ρ is the maximally mixed state $\mathrm{Id}/|\mathrm{H}|$.

The *mutual information* of a state ρ, on a bipartite Hilbert space $\mathrm{H}_1 \otimes \mathrm{H}_2$, is defined as

$$I(\mathrm{H}_1{:}\mathrm{H}_2)(\rho) := S(\rho_1) + S(\rho_2) - S(\rho).$$

By sub-additivity of the quantum entropy [5], it is always non-negative, and equal to 0 if and only if $\rho = \rho_1 \otimes \rho_2$, i.e. if and only if there is no correlation (not even classical ones) between the two subsystems.

The *conditional mutual information* of a state ρ, on a tripartite Hilbert space $\mathrm{H}_1 \otimes \mathrm{H}_2 \otimes \mathrm{H}_3$, is defined as

$$I(\mathrm{H}_1{:}\mathrm{H}_2|\mathrm{H}_3)(\rho) := S(\rho_{13}) + S(\rho_{23}) - S(\rho_3) - S(\rho).$$

By strong sub-additivity of the quantum entropy [134], it is always non-negative.

Finally, the *relative entropy* between states ρ and σ is defined as

$$D(\rho\|\sigma) := \mathrm{Tr}(\rho(\log\rho - \log\sigma)).$$

It is always non-negative, and infinite whenever the support of ρ and the orthogonal of the support of σ have a non-zero intersection. For a state ρ on a bipartite Hilbert space $\mathrm{H}_1 \otimes \mathrm{H}_2$, we have the notable property that $D(\rho\|\rho_1 \otimes \rho_2) = I(\mathrm{H}_1{:}\mathrm{H}_2)(\rho)$.

The main reason why we shall later be interested in these entropic quantities is because several measures of the amount of entanglement present in multipartite quantum systems can be built from them. We will need all these definitions chiefly in Chapters 5 and 10.

Chapter 3

Permutation-symmetry and de Finetti type theorems

This chapter collects together some standard definitions and results revolving around permutation-symmetry in quantum information theory. The reader is referred to the review [93] or to the lecture notes [174], Chapter 22, for a much more satisfying treatment of this material. Here, we focus only on what we will make use of in this manuscript (most notably in Chapters 9, 10 and 11, but also more punctually in Chapter 7, and at a technical rather than fundamental level, in Chapters 5 and 6).

3.1 The symmetric subspace: a brief review of some basic notions and facts

Let H be a Hilbert space and $\{|i\rangle,\ 1 \leqslant i \leqslant |\mathrm{H}|\}$ be an orthonormal basis of H (with the convention that $\{1, \ldots, |\mathrm{H}|\} = \mathbf{N}$ whenever H is infinite-dimensional). For any $n \in \mathbf{N}$ and any permutation $\pi \in \mathfrak{S}(n)$, denote by $U(\pi) \in \mathcal{U}(\mathrm{H}^{\otimes n})$ the associated permutation unitary on $\mathrm{H}^{\otimes n}$, characterized by

$$\forall\ 1 \leqslant i_1, \ldots, i_n \leqslant |\mathrm{H}|,\ U(\pi)|i_1\rangle \otimes \cdots \otimes |i_n\rangle = |i_{\pi(1)}\rangle \otimes \cdots \otimes |i_{\pi(n)}\rangle.$$

Note that this definition is actually independent of the basis. The n-symmetric subspace of $\mathrm{H}^{\otimes n}$ can then be defined as the simultaneous $+1$-eigenspace of all $U(\pi)$'s,

$$\begin{aligned}\mathrm{Sym}^n(\mathrm{H}) &:= \left\{|\psi\rangle \in \mathrm{H}^{\otimes n}\ :\ \forall\ \pi \in \mathfrak{S}(n),\ U(\pi)|\psi\rangle = |\psi\rangle\right\} \\ &= \mathrm{Span}\left\{|v_{i_1,\ldots,i_n}\rangle = \sum_{\pi\in\mathfrak{S}(n)} |i_{\pi(1)}\rangle \otimes \cdots \otimes |i_{\pi(n)}\rangle\ :\ 1 \leqslant i_1 \leqslant \cdots \leqslant i_n \leqslant |\mathrm{H}|\right\}.\end{aligned}$$

The orthogonal projector onto $\mathrm{Sym}^n(\mathrm{H})$ may thus be written as

$$P_{\mathrm{Sym}^n(\mathrm{H})} = \frac{1}{n!}\sum_{\pi\in\mathfrak{S}(n)} U(\pi) = \sum_{1\leqslant i_1\leqslant\cdots\leqslant i_n\leqslant|\mathrm{H}|} |\psi_{i_1,\ldots,i_n}\rangle\langle\psi_{i_1,\ldots,i_n}|,$$

where for each $1 \leqslant i_1 \leqslant \cdots \leqslant i_n \leqslant |\mathrm{H}|$, $|\psi_{i_1,\ldots,i_n}\rangle$ denotes the unit vector having same direction as $|v_{i_1,\ldots,i_n}\rangle$. In the case where H is finite-dimensional, this can also be re-written as

$$P_{\mathrm{Sym}^n(\mathrm{H})} = \binom{n+|\mathrm{H}|-1}{n}\int_{|\psi\rangle\in S_{\mathrm{H}}} |\psi\rangle\langle\psi|^{\otimes n}\mathrm{d}\psi,$$

where $\mathrm{d}\psi$ stands for the uniform probability measure on the unit sphere S_{H} of H. This is due to Schur's Lemma, since $\mathrm{Sym}^n(\mathrm{H})$ is an irreducible representation of the commutant action of $\{U(\pi),\ \pi \in \mathfrak{S}(n)\}$, which is that of the local unitaries $\{U^{\otimes n},\ U \in \mathcal{U}(\mathrm{H})\}$ (see e.g. [165] for more group representation background).

A state ρ on $\mathrm{H}^{\otimes n}$ is said to be *permutation-invariant* if it satisfies $U(\pi)\rho U(\pi)^\dagger = \rho$ for all $\pi \in \mathfrak{S}(n)$. This condition is actually equivalent to the existence of a unit vector $|\psi\rangle \in \mathrm{Sym}^n(\mathrm{H} \otimes \mathrm{H}')$, where $\mathrm{H}' \equiv \mathrm{H}$, such that $\rho = \mathrm{Tr}_{\mathrm{H}'}|\psi\rangle\langle\psi|$. That is why permutation-invariant states are also sometimes simply referred to as *symmetric* states.

3.2 Permutation-invariant states in the bipartite finite-dimensional case

The permutation-invariant states on $(\mathbf{C}^d)^{\otimes 2}$ are the so-called *Werner states*. These are mixtures of the (fully) symmetric state π_s and the (fully) antisymmetric state π_a on $(\mathbf{C}^d)^{\otimes 2}$ (which are defined as the renormalized projectors onto the symmetric and antisymmetric subspaces of $(\mathbf{C}^d)^{\otimes 2}$, respectively). Concretely, set

$$\begin{aligned}
\mathrm{Sym}^2(\mathbf{C}^d) &:= \mathrm{Span}\left\{|i_1, i_2\rangle + |i_2, i_1\rangle \ : \ 1 \leqslant i_1 \leqslant i_2 \leqslant d\right\}, \\
\mathrm{Asym}^2(\mathbf{C}^d) &:= \mathrm{Span}\left\{|i_1, i_2\rangle - |i_2, i_1\rangle \ : \ 1 \leqslant i_1 < i_2 \leqslant d\right\}.
\end{aligned}$$

We then have the orthogonal direct sum decomposition

$$\left(\mathbf{C}^d\right)^{\otimes 2} = \mathrm{Sym}^2(\mathbf{C}^d) \overset{\perp}{\oplus} \mathrm{Asym}^2(\mathbf{C}^d), \text{ with } \begin{cases} \dim(\mathrm{Sym}^2(\mathbf{C}^d)) = d(d+1)/2 \\ \dim(\mathrm{Asym}^2(\mathbf{C}^d)) = d(d-1)/2 \end{cases}.$$

The two states π_s and π_a on $(\mathbf{C}^d)^{\otimes 2}$ are therefore defined as

$$\pi_s := \frac{2}{d(d+1)} P_{\mathrm{Sym}^2(\mathbf{C}^d)} \text{ and } \pi_a := \frac{2}{d(d-1)} P_{\mathrm{Asym}^2(\mathbf{C}^d)}.$$

And we define next, for each $0 \leqslant \lambda \leqslant 1$, the Werner state $\rho_\lambda := \lambda\pi_s + (1-\lambda)\pi_a$ on $(\mathbf{C}^d)^{\otimes 2}$.

This one-parameter family of states was introduced in [175] in order to understand the relation between entanglement and violation of Bell inequalities. It has received considerable interest since then. Indeed, due to their symmetry property, quantities which may be hard to compute in general become much easier to analyze for Werner states. And still, permutation-invariance is an assumption which is more often than not natural to make, so that they encompass a wide enough range of situations. For instance, separability vs entanglement is simple to characterize for Werner states, because it coincides with their being positive under partial transposition or not (see Chapter 2, Section 2.4, for definitions). The result is that ρ_λ is separable for $1/2 \leqslant \lambda \leqslant 1$ and entangled for $0 \leqslant \lambda < 1/2$ [175].

3.3 Classical and quantum de Finetti type theorems

The motivation behind all de Finetti type theorems is to reduce the study of permutation-invariant scenarios to that of i.i.d. ones. This is precisely what the seminal classical finite de Finetti theorem, proved by Diaconis and Freedman in [60], enables. Indeed, it tells us that the marginal probability distribution (in a few random variables) of an exchangeable probability distribution (in a lot of random variables) is well-approximated by a convex combination of product probability distributions. More precisely, we have the quantitative error-bound provided by Theorem 3.3.1 below.

Theorem 3.3.1 (Classical finite de Finetti theorem, [60]). *Let $P^{(n)}$ be an exchangeable probability distribution in n random variables, meaning that, for any $\pi \in \mathfrak{S}(n)$, $P^{(n)} \circ \pi = P^{(n)}$. For any $k \leqslant n$, denote by $P^{(k)}$ the marginal probability distribution of $P^{(n)}$ in k random variables. Then, there exists a probability distribution μ on the set of probability distributions in 1 random variable such that*

$$\left\| P^{(k)} - \int_Q Q^{\otimes k} \mathrm{d}\mu(Q) \right\|_1 \leqslant \frac{k^2}{n}.$$

The first quantum analogue of this result was established by Christandl, König, Mitchison and Renner in [48]. The statement is of similar spirit: the reduced state (on a few subsystems) of a permutation-invariant state (on a lot of subsystems) is well-approximated by a convex combination of product states. Again, the worst-case error can be quantitatively upper-bounded, which is the content of Theorem 3.3.2 below.

Theorem 3.3.2 (Quantum finite de Finetti theorem, [48]). *Let $\rho^{(n)}$ be a permutation-invariant state on $(\mathbf{C}^d)^{\otimes n}$, meaning that, for any $\pi \in \mathfrak{S}(n)$, $U(\pi)\rho^{(n)}U(\pi)^\dagger = \rho^{(n)}$. For any $k \leqslant n$, denote by $\rho^{(k)} = \mathrm{Tr}_{(\mathbf{C}^d)^{\otimes n-k}} \rho^{(n)}$ the reduced state of $\rho^{(n)}$ on $(\mathbf{C}^d)^{\otimes k}$. Then, there exists a probability distribution μ on the set of states on $\mathbf{C}^d$ such that*

$$\left\| \rho^{(k)} - \int_\sigma \sigma^{\otimes k} \mathrm{d}\mu(\sigma) \right\|_1 \leqslant \frac{2kd^2}{n}.$$

Note that the main difference between these two theorems is that the classical one applies to probability distributions on an infinite size alphabet while the quantum one only applies to quantum states on a finite-dimensional Hilbert space. And it is known that the appearance of the dimension in the upper bound is not an artefact of the proof techniques: there actually exist permutation-invariant states $\rho^{(n)}$ on $(\mathbf{C}^d)^{\otimes n}$ which are such that the distance of $\rho^{(k)}$ to the set of separable states on $(\mathbf{C}^d)^{\otimes k}$ scales as kd/n (see [48], Section C, or [143], Section III, for such examples).

In Chapter 9, a hierarchy of necessary conditions for separability, based on symmetric extensions, is studied in extensive depth. And it is Theorem 3.3.2 under this precise form which allows to prove that this hierarchy converges to separability. However, in many applications, one does not actually need such a strong approximation result: knowing only that the considered permutation-invariant state can be upper bounded by a convex combination of tensor power states (up to some multiplicative factor C) is enough. Indeed, assume that you have such an operator-ordering, and that you know that a given order-preserving linear form f satisfies $f \leqslant \epsilon$, for some $0 < \epsilon < 1$, on 1-particle states. Then, you can conclude that $f^{\otimes n} \leqslant C\epsilon^n$ on permutation-invariant n-particle states, which decays exponentially to 0 with n. These de Finetti results of different kind are usually referred to as *de Finetti reductions* or *post-selection lemmas*. Establishing and applying appropriate variants of them is at the heart of Chapters 10 and 11.

Chapter 4

Asymptotic geometric analysis toolbox

This chapter gathers several notions from asymptotic geometric analysis that will be used in many places of this manuscript. Section 4.1 introduces standard definitions, notation and results from classical convex geometry, in particular related to estimating the size of a convex body, which is something that we will have to do at various occasions. Section 4.2 summarizes two basic incarnations of the concentration of measure phenomenon (taking the view point of either functions on a sphere or sums of independent random variables), on which we shall build later more elaborate deviation estimates, tailored to our specific needs. Additional functional analytic tools (e.g. from random matrix theory or from the local theory of Banach spaces), which play a more sporadic role in this manuscript, will be introduced only in due time.

4.1 Classical convex geometry

The standard convex geometry concepts expounded in this section will be crucial tools in (at least part of) Chapters 6, 7, 8 and 9. The reader is e.g. referred to the lecture notes [20] or [173] for a detailed and accessible presentation.

4.1.1 Some vocabulary

We work in the Euclidean space $\mathbf{R}^n$, where we denote by $\|\cdot\|$ the Euclidean norm, and by B^n, resp. S^{n-1}, the associated unit ball, resp. sphere. We denote by $\mathrm{vol}_n(\cdot)$ or simply $\mathrm{vol}(\cdot)$ the n-dimensional Lebesgue measure. A *convex body* $K \subset \mathbf{R}^n$ is a convex compact set with non-empty interior. A convex body K is *symmetric* if $K = -K$.

The *gauge* (or *Minkowski functional*) associated to a convex body K in $\mathbf{R}^n$ is the function $\|\cdot\|_K$ defined for $x \in \mathbf{R}^n$ by

$$\|x\|_K := \inf\{t \geqslant 0 \ : \ x \in tK\}.$$

This is a norm if and only if K is symmetric. In such case, this actually defines a one-to-one correspondence between norms on $\mathbf{R}^n$ and symmetric convex bodies in $\mathbf{R}^n$: K is simply the unit ball for $\|\cdot\|_K$. This identification has the following elementary properties: $K \subset L$ if and only if $\|\cdot\|_K \geqslant \|\cdot\|_L$, and for all $t > 0$, $\|\cdot\|_{tK} = \|\cdot\|_K / t$.

If $K \subset \mathbf{R}^n$ is a convex body with origin in its interior, the *polar* of K is the convex body K° defined as

$$K^\circ := \{y \in \mathbf{R}^n \ : \ \forall\, x \in K,\ \langle x, y\rangle \leqslant 1\}.$$

In the symmetric case, the norms $\|\cdot\|_K$ and $\|\cdot\|_{K^\circ}$ are dual to each other, meaning that

$$\forall\, x \in \mathbf{R}^n,\ \|x\|_{K^\circ} = \sup\{\langle x, y\rangle \ : \ \|y\|_K \leqslant 1\} = \sup\{\langle x, y\rangle \ : \ y \in K\}.$$

If $x \in S^{n-1}$, we shall sometimes call the quantity defined above the *support function* of K in the direction x, and use the notation

$$h_K(x) := \|x\|_{K^\circ}.$$

Note that $h_K(x)$ is the distance from the origin to the hyperplane tangent to K in the direction x.

Two global invariants associated to a convex body $K \subset \mathbf{R}^n$, the *volume radius* and the *mean width*, will play an important role in many of our proofs.

Definition 4.1.1. *The volume radius of a convex body $K \subset \mathbf{R}^n$ is defined as*

$$\operatorname{vrad}(K) := \left(\frac{\operatorname{vol} K}{\operatorname{vol} B^n}\right)^{1/n}.$$

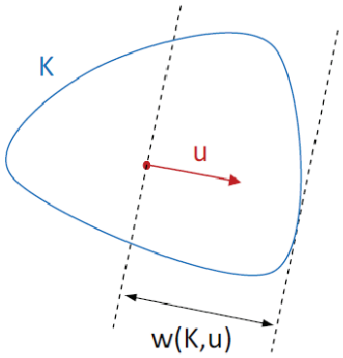
In words, $\operatorname{vrad}(K)$ *is the radius of the Euclidean ball with same volume as* K*.*

Definition 4.1.2. *The mean width of a subset $K \subset \mathbf{R}^n$ is defined as*

$$w(K) := \int_{S^{n-1}} \max_{x \in K} \langle u, x \rangle \, \mathrm{d}\sigma(u),$$

where σ is the uniform probability measure on S^{n-1}. In words, $w(K)$ is the average, for u uniformly distributed on S^{n-1}, of the (half-)width of K in the direction u (see Figure 4.1).

Figure 4.1: Width of the convex body K in the direction u



If K is a convex body, we have

$$w(K) = \int_{S^{n-1}} h_K(u) \, \mathrm{d}\sigma(u) = \int_{S^{n-1}} \|u\|_{K^\circ} \, \mathrm{d}\sigma(u).$$

Moreover, one can check that the mean width satisfies the additivity property $w(K+L) = w(K) + w(L)$, for any $K, L \subset \mathbf{R}^n$ bounded.

The inequality below (see e.g. [148], Corollary 1.4, for a proof) is a fundamental result which compares the volume radius and the mean width. It asserts that, among sets of given volume, the mean width is minimized for Euclidean balls.

Theorem 4.1.3 (Urysohn inequality)**.** *For any convex body $K \subset \mathbf{R}^n$, we have*

$$\operatorname{vrad}(K) \leqslant w(K).$$

It is convenient to compute the mean width using Gaussian rather than spherical integration. Let G be a standard Gaussian vector in $\mathbf{R}^n$, i.e. such that its coordinates, in any orthonormal basis, are independent and following a Gaussian distribution with mean 0 and variance 1. Denoting $\gamma_n = \mathbf{E}\,\|G\| \sim_{n\to+\infty} \sqrt{n}$, we have, for any compact set $K \subset \mathbf{R}^n$,

$$w_G(K) := \mathbf{E} \max_{x \in K} \langle G, x \rangle = \gamma_n w(K).$$

The Gaussian mean width has the advantage over the spherical one of being independent from the ambient dimension. Indeed, if $K \subset E$ is a compact set in a subspace E of $\mathbf{R}^n$, the value of $w_G(K)$ does not depend on whether it is computed in E or in $\mathbf{R}^n$, contrary to $w(K)$. It is also usually easier to compute. For example, it allows to compute the mean width of a segment: if $u \in S^{n-1}$ is a unit vector, then

$$\alpha_n := w(\operatorname{conv}\{\pm u\}) = \frac{1}{\gamma_n}\sqrt{\frac{2}{\pi}} \underset{n\to+\infty}{\sim} \sqrt{\frac{2}{\pi n}}.$$

It also shows how to control the mean width of an intersection or a projection. Let $K \subset \mathbf{R}^n$ be a compact set, and $E \subset \mathbf{R}^n$ be a k-dimensional subspace. Denoting by P_E the orthogonal projection onto E, we have $w_G(P_E K) \leqslant w_G(K)$, and therefore

$$w(K \cap E) \leqslant w(P_E K) \leqslant \frac{\gamma_n}{\gamma_k} w(K). \tag{4.1}$$

We will also need the two following lemmas, which are incarnations of the familiar "union bound". Lemma 4.1.4 appears for example in [131], Chapter 3, under the equivalent formulation via suprema of Gaussian processes. Lemma 4.1.5 on the other hand, seems to appear nowhere in the literature, even under a different phrasing. Since it is a result that we will use at several occasions (in Chapters 7 and 8) we write a short proof of it for completeness. The route we follow is the "usual" one in that setting, detailed e.g. in [131], Chapter 3, to which the reader is referred for the details that we may pass over.

Lemma 4.1.4 (Bounding the mean width of a polytope)**.** *Let $v_1, \dots, v_N$ be points in $\mathbf{R}^n$ such that $v_i \in \lambda B^n$ for every index $1 \leqslant i \leqslant N$, for some $\lambda \geqslant 0$. Then,*

$$w\left(\operatorname{conv}(v_1, \dots, v_N)\right) \leqslant \lambda \sqrt{\frac{2 \ln N}{n}}.$$

Lemma 4.1.5 (Bounding the mean width of a union)**.** *Let $K_1, \dots, K_N$ be convex sets in $\mathbf{R}^n$ such that $K_i \subset \lambda B^n$ for every index $1 \leqslant i \leqslant N$, for some $\lambda \geqslant 0$. Then,*

$$w\left(\operatorname{conv}\left(\bigcup_{i=1}^N K_i\right)\right) \leqslant 2\left(\max_{1 \leqslant i \leqslant N} w(K_i) + \lambda \sqrt{\frac{2 \ln N}{n}}\right).$$

Proof. By homogeneity, it is enough to prove Lemma 4.1.5 in the case $\lambda = 1$. Now, for G a standard Gaussian in $\mathbf{R}^n$, we have that, for any $\delta \geqslant 0$ (to be fixed later),

$$\mathbf{E} \max_{1 \leqslant i \leqslant N} \left| h_{K_i}(G) - w_G(K_i) \right| \leqslant \delta + \sum_{i=1}^N \int_\delta^{+\infty} \mathbf{P}\left(\left| h_{K_i}(G) - w_G(K_i) \right| > t \right) \mathrm{d}t.$$

Yet, by the Gaussian concentration inequality (see e.g. [131], Chapter 1, and also Section 4.2 below for a more detailed exposition of closely related results), we know that, for each $1 \leqslant i \leqslant N$,

$$\forall\, t > 0,\ \mathbf{P}\left(\left| h_{K_i}(G) - w_G(K_i) \right| > t \right) \leqslant e^{-t^2/2}.$$

This is because, by the assumption $K_i \subset B^n$, h_{K_i} is a 1-Lipschitz function:

$$\forall\, G, H \in \mathbf{R}^n,\ \left| h_{K_i}(G) - h_{K_i}(H) \right| \leqslant h_{K_i}(G - H) \leqslant \|G - H\|.$$

We therefore have in the end

$$\mathbf{E} \max_{1 \leqslant i \leqslant N} \left| h_{K_i}(G) - w_G(K_i) \right| \leqslant \delta + N \int_\delta^{+\infty} e^{-t^2/2} \mathrm{d}t \leqslant \delta + N \sqrt{\frac{\pi}{2}} e^{-\delta^2/2}.$$

Choosing $\delta = \sqrt{2 \ln N}$ in the above inequality eventually yields

$$\mathbf{E} \max_{1 \leqslant i \leqslant N} h_{K_i}(G) \leqslant 2 \max_{1 \leqslant i \leqslant N} w_G(K_i) + \sqrt{2 \ln N} + \sqrt{\frac{\pi}{2}} \leqslant 2\left(\max_{1 \leqslant i \leqslant N} w_G(K_i) + \sqrt{2 \ln N} \right).$$

And going back to spherical rather than Gaussian variables gives precisely the advertised result. □

4.1.2 Some volume inequalities

We will use several times (in Chapters 7, 8 and 9) the following result, established by Milman and Pajor.

Theorem 4.1.6 (Milman–Pajor inequality, [140], Corollary 3)**.** *If K, L are convex bodies in $\mathbf{R}^n$ with the same center of mass, then*

$$\operatorname{vrad}(K \cap L) \operatorname{vrad}(K - L) \geqslant \operatorname{vrad}(K) \operatorname{vrad}(L),$$

where $K - L = \{x - y \ :\ x \in K,\ y \in L\}$ stands for the Minkowski sum of the convex bodies K and $-L$.

Choosing $K = -L$ in Theorem 4.1.6 yields the following corollary.

Corollary 4.1.7. *If K is a convex body in $\mathbf{R}^n$ with center of mass at the origin, then*

$$\mathrm{vrad}(K \cap -K) \geqslant \frac{1}{2}\mathrm{vrad}(K),$$

and more generally for any orthogonal transformation θ,

$$\mathrm{vrad}(K \cap \theta(K)) \geqslant \frac{1}{2}\frac{\mathrm{vrad}(K)^2}{w(K)}.$$

The latter inequality is simply because, on the one hand, $\mathrm{vrad}(\theta(K)) = \mathrm{vrad}(K)$, and on the other hand, by Theorem 4.1.3, $\mathrm{vrad}(K - \theta(K)) \leqslant w(K - \theta(K)) = w(K) + w(\theta(K)) = 2w(K)$.

We will typically use Corollary 4.1.7 in the following way: if K is a convex body with center of mass at the origin which satisfies a "reverse" Urysohn inequality, i.e. $\mathrm{vrad}(K) \geqslant \alpha w(K)$ for some constant $0 < \alpha < 1$, we can conclude that the volume radius of $K \cap \theta(K)$ is comparable to the volume radius of K.

Another volume inequality which will be useful to us (in Chapter 7) is the one below, due to Rogers and Shepard.

Theorem 4.1.8 (Rogers–Shephard inequality, [153])**.** *Let $u \in S^{n-1}$, $h > 0$, and consider the affine hyperplane*

$$H = \{x \in \mathbf{R}^n \ : \ \langle x, u\rangle = h\}.$$

Let K be a convex body inside H and $L = \mathrm{conv}(K, -K)$. Then,

$$2h\,\mathrm{vol}_{n-1}(K) \leqslant \mathrm{vol}_n(L) \leqslant 2h\,\mathrm{vol}_{n-1}(K)\frac{2^{n-1}}{n}.$$

Consequently,

$$\mathrm{vrad}(L) \simeq h^{1/n}\mathrm{vrad}(K)^{1-1/n}. \tag{4.2}$$

We can infer from equation (4.2) that for sets K with "reasonable" volume (which will be the case of all the sets that we will consider) $\mathrm{vrad}(K)$ and $\mathrm{vrad}(L)$ are comparable.

4.2 Concentration of measure and deviation inequalities

4.2.1 From individual to global concentration estimates

Levy's Lemma guarantees that, in high dimension, regular enough functions typically do not deviate much from their average behaviour. We will use repeatedly variations and refinements of this general paradigm (in Chapter 7 a multi-variate version, in Chapter 8 a Gaussian rather than spherical version, in Chapter 9 a local version). So let us recall precisely the seminal formulation here.

Lemma 4.2.1 (Levy's Lemma for Lipschitz functions on the sphere, [132])**.** *Let $n \in \mathbf{N}$. For any L-Lipschitz function $f : S^{n-1} \to \mathbf{R}$ and any $t > 0$, if x is uniformly distributed on S^{n-1}, then*

$$\mathbf{P}(|f(x) - \mathbf{E}\,f| > t) \leqslant e^{-cnt^2/L^2},$$

where $c > 0$ is a universal constant.

Combining such type of individual deviation probability estimates and a discretization with *nets* of reasonable size, one can then usually derive global deviation probability estimates via the union bound. Let us specify what we mean.

Definition 4.2.2. *Fix $\varepsilon > 0$, and let $\|\cdot\|_\sharp, \|\cdot\|_\flat$ be two norms in $\mathbf{R}^n$, with associated unit balls $B_\sharp, B_\flat$. $\mathscr{A}$ is an ε-net for $\|\cdot\|_\flat$ within $B_\sharp$ if $\mathscr{A} \subset B_\sharp$ and, for all $x \in B_\sharp$, there exists $x' \in \mathscr{A}$ such that $\|x - x'\|_\flat \leqslant \varepsilon$.*

Usually, one is interested in having such a discretized version of a given unit ball with as few elements as possible. Now, just observing that an ε-separated set with maximal cardinality actually forms an ε-net, it is easy to get the following cardinality upper bound by simply comparing volumes.

Lemma 4.2.3 (Bounding the cardinality of nets via a volumetric argument, [148], Lemmas 4.16 and 4.10)**.** *Fix $\varepsilon > 0$, and let $\|\cdot\|_\sharp, \|\cdot\|_\flat$ be two norms in $\mathbf{R}^n$, with associated unit balls $B_\sharp, B_\flat$. Then, there exists an ε-net $\mathcal{A}$ for $\|\cdot\|_\flat$ within $B_\sharp$ satisfying*

$$|\mathcal{A}| \leqslant \frac{\mathrm{vol}\left((2/\varepsilon)B_\sharp + B_\flat\right)}{\mathrm{vol}\left(B_\flat\right)}.$$

So in particular, there exists an ε-net $\mathcal{A}$ for $\|\cdot\|_\sharp$ within $B_\sharp$ such that $|\mathcal{A}| \leqslant (1 + 2/\varepsilon)^n$.

Subsequently, making use of both Lemmas 4.2.1 and 4.2.3 in a careful way, one gets the celebrated Dvoretzky's Theorem, which is quoted below. This is precisely the kind of strategy we adopt in Chapters 5 and 6 to go from individual to global concentration phenomenon.

Lemma 4.2.4 (Dvoretzky's Theorem for Lipschitz functions on the sphere, [139, 83, 162])**.** *Let $n \in \mathbf{N}$. For any symmetric L-Lipschitz function $f : S^{n-1} \to \mathbf{R}$ and any $t > 0$, if H is a uniformly distributed Cnt^2/L^2-dimensional subspace of $\mathbf{R}^n$, with $C > 0$ a universal constant, then*

$$\mathbf{P}\left(\exists\, x \in H \cap S^{n-1} :\ |f(x) - \mathbf{E}\, f| > t\right) \leqslant e^{-cnt^2/L^2},$$

where $c > 0$ is a universal constant.

This tangible version of Dvoretzky's theorem is essentially due to Milman [139], but with a dependence on t in the maximal dimension of H which scales as $t^2/\log(1/t)$. This extra logarithmic factor was later removed by Gordon in [83] and by Schechtman in [162] (the latter proof is based on concentration of measure, like the original one, while the former proof uses instead comparison inequalities for Gaussian processes).

4.2.2 ψ_α random variables

The reader is for instance referred to the review [42], Chapter 1, for a complete presentation of the theory of Orlicz spaces. A particular instance of these are the so-called L_{ψ_α} spaces, of which we gather here only the few properties that we will need to exploit.

For any $\alpha \geqslant 1$, a random variable X is called a ψ_α random variable if its ψ_α-norm $\|X\|_{\psi_\alpha}$ is finite. The latter may be defined in several equivalent ways, the most standard one being through the Orlicz function $x \mapsto \exp(x^\alpha) - 1$. This characterization is not the one that will be most practical for us though, and we shall rather use the one via the growth of absolute moments, which leads to the equivalent norm (see e.g. [42], Corollary 1.1.6)

$$\|X\|_{\psi_\alpha} = \sup_{p \in \mathbf{N}} \frac{\left(\mathbf{E}\,|X|^p\right)^{1/p}}{p^{1/\alpha}}.$$

A ψ_1, resp. ψ_2, random variable is also referred to as as sub-exponential, resp. sub-gaussian. Indeed, the ψ_1-norm, resp. ψ_2-norm, of a random variable quantifies the exponential, resp. gaussian, decay of its tail.

As a crucial tool in several of our coming reasonings (in Chapters 5 and 6), we will need the Bernstein-type deviation inequality for a sum of independent ψ_1 random variables which is quoted below (see e.g. [42], Theorem 1.2.5, for a proof).

Theorem 4.2.5 (Bernstein's inequality)**.** *Let $X_1, \ldots, X_N$ be N independent centered ψ_1 random variables. Setting $M = \max_{1 \leqslant i \leqslant N} \|X_i\|_{\psi_1}$ and $\sigma^2 = \sum_{1 \leqslant i \leqslant N} \left(\|X_i\|_{\psi_1}\right)^2 / N$, we have*

$$\forall\ t > 0,\ \mathbf{P}\left(\left|\frac{1}{N}\sum_{i=1}^N X_i\right| > t\right) \leqslant 2\exp\left(-cN\min\left(\frac{t^2}{\sigma^2}, \frac{t}{M}\right)\right),$$

where $c > 0$ is a universal constant.

We see in Theorem 4.2.5 above that, for a sum of independent ψ_1 random variables, there are two distinct regimes which enter the picture in the behaviour of the tail: sub-gaussian for moderate deviations (when the central limit phenomenon dominates) and sub-exponential for large deviations (when the prominent role is played by the tails of the individual variables).

Part II

Complexity reduction in quantum information theory

A central issue in asymptotic geometric analysis is that of quantifying how close a given, potentially complex, convex body is to one which is easier to describe. There are several reasons why a convex body may be considered simple: because there is an efficient way of deciding membership for it (algorithmic point of view), because it is the unit ball for a Euclidean norm or because it has few vertices/faces (geometric point of view), because it can be covered with few Euclidean balls (entropic point of view) etc. There are also several notions of distance between convex bodies that one might consider depending on the context. In any case, two sample (dual) take-home messages from asymptotic geometric analysis are: given a high dimensional convex body, firstly most of its information is already encoded in its projection onto a lower dimensional subspace, and secondly most of its complexity disappears when looking at its section by a lower dimensional subspace. Mathematically, these two results are known, respectively, as the Johnson–Lindenstrauss lemma and as the Dvoretzky theorem. From an information theory standpoint, they have a clear data-compression interpretation, with an achievability side (preservation of almost all information when projecting onto a subspace of dimension above a certain value) and a converse side (loss of almost all information when intersecting with a subspace of dimension below a certain value).

Similar considerations appear in many contexts of quantum information theory. Phrased very generally, a natural wonder would usually be: given an ideal process, with many (even potentially infinitely many) degrees of freedom, is it possible to approximate it by a more realistic one, i.e. one which can be described with few parameters? Or in other words: by just allowing some small error, can we execute a task which potentially requires a lot of resources with much less resources? In this part, we take a closer look at two such problems: the question of compressing a quantum channel into one with a small environment is treated in Chapter 5, and the question of sparsifying a quantum measurement into one with few outcomes is treated in Chapter 6. Here is a summary of the main results in each of them.

Chapter 5 deals with the following problem: given a quantum channel $\mathcal{N}$, find a quantum channel $\widehat{\mathcal{N}}$ with environment as small as possible such that, for any input state ρ, the output states $\mathcal{N}(\rho)$ and $\widehat{\mathcal{N}}(\rho)$ are close to one another. We investigate different notions of closeness: standard ones quantified in terms of Schatten p-norm distance, but also stronger ones defined in terms of operator ordering. Those are especially well-suited for then deriving closeness results involving, for instance, entropic output quantities. In brief, our main result is that any channel $\mathcal{N}$ with input and output dimensions d can be approximated, say in $(1\to1)$-norm distance, by a channel $\widehat{\mathcal{N}}$ with environment dimension $O(d\log d)$, hence much smaller than the a priori environment dimension d^2 of $\mathcal{N}$. In the case where $\mathcal{N}$ is sufficiently noisy (meaning that all its output states are sufficiently mixed), this result can be improved to $O(d)$. Such dimensional dependence is shown to be optimal, contrary to the general one for which we do not know whether or not the $\log d$ factor can be removed. On the technical side, all our statements stem, as a first crucial step, from large deviation inequalities for sums of independent sub-exponential random variables.

In Chapter 6, attention is focussed on a particular kind of quantum channels, namely quantum-classical channels (i.e. POVMs). However, the notion of approximation which is investigated there is completely different from that of Chapter 5. Let us specify what we mean. One can associate to any POVM M its so-called *distinguishability norm* $\|\cdot\|_{\mathrm{M}}$, which quantifies how well it performs in the task of discriminating two quantum states. Given a POVM M, with potentially many (or even infinitely many) outcomes, we are interested in constructing a POVM M$'$, with as few outcomes as possible, behaving almost as the POVM M in terms of distinguishability norm (in the sense that $\|\cdot\|_{\mathrm{M}} \simeq \|\cdot\|_{\mathrm{M}'}$, up to some small error). This notion of closeness between POVMs thus has an operational significance: two POVMs are comparable if they yield comparable biases on any pair of states to be discriminated. Our first realization is that the unit ball for a POVM's distinguishability norm is a very particular object: precisely, its polar is, in general, a *zonoid*, and in the case of a discrete POVM, a *zonotope*. What kind of convex body are these? A zonotope is the Minkowski sum of a finite number of segments, while a zonoid is the limit of a sequence of zonotopes (in Hausdorff distance). So our problem can almost be rephrased as: to approximate a zonoid by a zonotope, how many segments are needed? And this turns out to be a well-studied topic in classical convex geometry. Exporting these ideas from the local theory of Banach spaces, we are able to prove the following main results: on $\mathbf{C}^d$, a POVM which is symmetric enough (such as e.g. the uniform POVM, the most symmetric one) can be sparsified by a POVM with $O(d^2)$ outcomes, and any POVM can be sparsified by a sub-POVM with $O(d^2\log d)$ outcomes. The dimensional dependence in the former result is optimal, and we leave it open whether or not the latter result can be improved. What is more, we can also deal with the multipartite setting, after considering the appropriate notion of tensor product for zonoids. We establish the, desirable but not a priori obvious, fact that local POVMs can be sparsified locally.

Looking at it more closely afterwards, there are analogies between Chapters 5 and 6 which go beyond their common complexity reduction motivation. To begin with, we propose in both cases a universal random construction which has the property of working optimally in very balanced situations (such as the fully randomizing channel or the uniform POVM), but maybe not as well in very unbalanced ones. Hence the question of coming up with more adaptive schemes in one instance or the other. Note that this issue is far from being specific to the two problems we are dealing with. On the contrary, it is for instance a recurrent one when trying to embed a Banach space into another one almost isometrically: how can their geometry be taken into account in a clever way? Furthermore, the routes we follow in both chapters in order to establish our main theorems have the exact same spirit (which is, admittedly, a very standard one): first we prove concentration for one fixed data point (which follows from certain random variables having a nice sub-exponential behaviour), and then we derive concentration for all data points by discretizing our data set with a well-chosen net. The reader is referred to Chapter 4, Section 4.2, for the archetypal example of this strategy, namely the derivation of Dvoretzky's theorem from Levy's lemma.

Part II – Table of contents

Chapter 5 Quantum channel compression 45
5.1 Introduction 45
5.2 Quantum channel approximation: definitions and already known facts 46
5.3 Statement of the main results 47
5.4 Proof of the main results 49
5.5 Consequences and applications 53
5.6 Discussion 57

Chapter 6 Zonoids and sparsification of quantum measurements 59
6.1 Introduction 59
6.2 POVMs and distinguishability norms 60
6.3 POVMs and zonoids 61
6.4 Local POVMs and tensor products of zonoids 64
6.5 Sparsifying POVMs 66
6.6 Sparsifying local POVMs 69
6.7 Proof of the main theorem concerning the sparsification of the uniform POVM 70
6.8 Proof of the main theorem concerning the sparsification of any POVM 72

Chapter 5

Quantum channel compression

Based on "Quantum channel compression" [127].

We study the problem of approximating a quantum channel by one with as few Kraus operators as possible (in the sense that, for any input state, the output states of the two channels should be close to one another). Our main result is that any quantum channel mapping states on some input Hilbert space A to states on some output Hilbert space B can be compressed into one with order $d \log d$ Kraus operators, where $d = \max(|\mathrm{A}|, |\mathrm{B}|)$, hence much less than $|\mathrm{A}||\mathrm{B}|$. In the case where the channel's outputs are all very mixed, this can be improved to order d. We discuss the optimality of this result as well as some consequences.

5.1 Introduction

Quantum channels are the most general framework in which the transformations that a quantum system may undergo are described. These are defined as completely positive and trace preserving (CPTP) maps from the set of bounded operators on some input Hilbert space A to the set of bounded operators on some output Hilbert space B. Indeed, to be a physically valid evolution in the open system setting, a linear map $\mathcal{N}$ has to preserve quantum states (i.e. positive semi-definiteness and unit-trace conditions) even when tensorized with the identity map $\mathcal{I}d$ on an auxiliary system. The reader is referred to Chapter 2, Section 2.3, for a more developed exposition.

In the remainder of this chapter, we shall use the general notation introduced in Chapter 1, Section 1.3 (in particular concerning operators and operator-norms on Hilbert spaces). Also, given a finite-dimensional Hilbert space H (which will be the case of all the Hilbert spaces we will deal with in the sequel) we shall denote by $|\mathrm{H}|$ its dimension.

So assume from now on that the Hilbert spaces A and B are finite-dimensional. Then, we know by Choi's representation theorem [46] that a CPTP map $\mathcal{N} : \mathcal{L}(\mathrm{A}) \to \mathcal{L}(\mathrm{B})$ can always be written as

$$\mathcal{N} : X \in \mathcal{L}(\mathrm{A}) \mapsto \sum_{i=1}^{s} K_i X K_i^\dagger \in \mathcal{L}(\mathrm{B}), \tag{5.1}$$

where the operators $K_i : \mathrm{A} \to \mathrm{B}$, $1 \leqslant i \leqslant s$, are called the Kraus operators of $\mathcal{N}$ and satisfy the normalization relation $\sum_{i=1}^{s} K_i^\dagger K_i = \mathrm{Id}_\mathrm{A}$. The minimal $s \in \mathbf{N}$ such that $\mathcal{N}$ can be decomposed in the Kraus form (5.1) is called the Kraus rank of $\mathcal{N}$, which we shall denote by $r_K(\mathcal{N})$. By Stinespring's dilatation theorem [168], another alternative way of characterizing a CPTP map $\mathcal{N} : \mathcal{L}(\mathrm{A}) \to \mathcal{L}(\mathrm{B})$ is as follows

$$\mathcal{N} : X \in \mathcal{L}(\mathrm{A}) \mapsto \mathrm{Tr}_\mathrm{E}\left(VXV^\dagger\right) \in \mathcal{L}(\mathrm{B}), \tag{5.2}$$

for some environment Hilbert space E and some isometry $V : \mathrm{A} \hookrightarrow \mathrm{B} \otimes \mathrm{E}$ (i.e. $V^\dagger V = \mathrm{Id}_\mathrm{A}$). In such picture, $r_K(\mathcal{N})$ is then nothing else than the minimal environment dimension $|\mathrm{E}| \in \mathbf{N}$ such that $\mathcal{N}$ may be expressed in the Stinespring form (5.2). It may be worth pointing out that there is a lot of freedom in representation (5.1): two sets of Kraus operators $\{K_i,\ 1 \leqslant i \leqslant s\}$ and $\{L_i,\ 1 \leqslant i \leqslant s\}$ give rise to the same quantum channel as soon as there exists a unitary U on $\mathbf{C}^s$ such that, for all $1 \leqslant i \leqslant s$, $L_i = \sum_{j=1}^{s} U_{ij} K_j$. On the contrary, representation (5.2) is essentially unique, up to the (usually irrelevant) transformation $V \mapsto (\mathrm{Id} \otimes U)V$, for U a unitary on E. That is why we will often prefer working with the latter than with the former.

Yet another way of viewing the Kraus rank of a CPTP map $\mathcal{N} : \mathscr{L}(\mathrm{A}) \to \mathscr{L}(\mathrm{B})$ is as the rank of its associated Choi-Jamiolkowski state. Denoting by ψ a maximally entangled state on $\mathrm{A} \otimes \mathrm{A}$, i.e. $|\psi\rangle = \sum_{i=1}^{|\mathrm{A}|} |ii\rangle/\sqrt{|\mathrm{A}|}$ for $\{|i\rangle,\ 1 \leqslant i \leqslant |\mathrm{A}|\}$ an orthonormal basis of A, the latter is defined as the state $\tau(\mathcal{N}) = \mathcal{Id} \otimes \mathcal{N}(\psi)$ on $\mathrm{A} \otimes \mathrm{B}$. Consequently, it holds that any quantum channel from A to B has Kraus rank at most $|\mathrm{A}||\mathrm{B}|$. And the extremal such quantum channels have Kraus rank less than $|\mathrm{A}|$. In particular, the case $r_K(\mathcal{N}) = 1$ corresponds to $\mathcal{N}$ being a unitary, hence reversible, evolution, whereas whenever $r_K(\mathcal{N}) > 1$, one can view $\mathcal{N}$ as a noisy summary of a unitary evolution on a larger system. The Kraus rank of a quantum channel can thus legitimately be seen as a measure of its "complexity": it quantifies the minimal amount of ancillary resources needed to implement it (or equivalently the amount of degrees of freedom in it that one is ignorant of). A natural question in this context would therefore be: given any quantum channel, is it possible to reduce its complexity while not affecting too much its action, or in other words to find a channel with much smaller Kraus rank which approximately simulates it?

One last definition we shall need concerning CP maps is the following: the conjugate (or dual) of a CP map $\mathcal{N} : \mathscr{L}(\mathrm{A}) \to \mathscr{L}(\mathrm{B})$ is the CP map $\mathcal{N}^* : \mathscr{L}(\mathrm{B}) \to \mathscr{L}(\mathrm{A})$ defined by

$$\forall\ X \in \mathscr{L}(\mathrm{A}),\ \forall\ Y \in \mathscr{L}(\mathrm{B}),\ \mathrm{Tr}(\mathcal{N}(X)Y) = \mathrm{Tr}(X\mathcal{N}^*(Y)).$$

It is characterized as well by saying that $\{K_i,\ 1 \leqslant i \leqslant s\}$ is a set of Kraus operators for $\mathcal{N}$ if and only if $\{L_i = K_i^\dagger,\ 1 \leqslant i \leqslant s\}$ is a set of Kraus operators for $\mathcal{N}^*$. Hence obviously, $\mathcal{N}$ and $\mathcal{N}^*$ have same Kraus rank, while the trace-preservingness condition $\sum_{i=1}^s K_i^\dagger K_i = \mathrm{Id}$ for $\mathcal{N}$ is equivalent to the unitality condition $\sum_{i=1}^s L_i L_i^\dagger = \mathrm{Id}$ for $\mathcal{N}^*$.

5.2 Quantum channel approximation: definitions and already known facts

Before going any further, we need to specify a bit what we mean by "approximating a quantum channel", since indeed, several definitions of approximation may be considered. In our setting, the most natural one is probably that of approximation in (1→1)-norm: given CPTP maps $\mathcal{N}, \widehat{\mathcal{N}} : \mathscr{L}(\mathrm{A}) \to \mathscr{L}(\mathrm{B})$, we will say that $\widehat{\mathcal{N}}$ is an ε-approximation of $\mathcal{N}$ in (1→1)-norm, where $\varepsilon > 0$ is some fixed parameter, if

$$\forall\ \rho \in \mathscr{D}(\mathrm{A}),\ \left\|\widehat{\mathcal{N}}(\rho) - \mathcal{N}(\rho)\right\|_1 \leqslant \varepsilon. \tag{5.3}$$

One could think at first sight that an even more natural error quantification in such context would be in terms of completely-bounded (1→1)-norm (aka diamond norm). That is, in order to call $\widehat{\mathcal{N}}$ an ε-approximation of $\mathcal{N}$, we would require that, for any Hilbert space A',

$$\forall\ \rho \in \mathscr{D}(\mathrm{A} \otimes \mathrm{A}'),\ \left\|\widehat{\mathcal{N}} \otimes \mathcal{Id}(\rho) - \mathcal{N} \otimes \mathcal{Id}(\rho)\right\|_1 \leqslant \varepsilon. \tag{5.4}$$

Nevertheless, this notion of approximation is too strong for our purposes. Indeed, if $\mathcal{N}$ and $\widehat{\mathcal{N}}$ satisfy equation (5.4), it implies in particular that their associated Choi-Jamiolkowski states have to be ε-close in trace-norm distance. And this, in general, is possible only if $\mathcal{N}$ and $\widehat{\mathcal{N}}$ have a number of Kraus operators which scale the same: for instance, if $\tau(\mathcal{N}) = P/r_K(\mathcal{N})$ with P a projector on a $r_K(\mathcal{N})$-dimensional subspace of $\mathrm{A} \otimes \mathrm{B}$, then $\|\tau(\widehat{\mathcal{N}}) - \tau(\mathcal{N})\|_1 \leqslant \varepsilon$ can hold only if $r_K(\widehat{\mathcal{N}}) \geqslant (1 - \varepsilon/2)\, r_K(\mathcal{N})$. So no environment dimensionality reduction can be achieved in that sense.

The question of quantum channel compression has already been studied in one specific case, which is the one of the fully randomizing (or depolarizing) channel. Let us recall what is known there. The fully randomizing channel $\mathcal{R} : \mathscr{L}(\mathrm{A}) \to \mathscr{L}(\mathrm{A})$ is the CPTP map with same input and output spaces defined by

$$\mathcal{R} : X \in \mathscr{L}(\mathrm{A}) \mapsto (\mathrm{Tr}\, X)\frac{\mathrm{Id}}{|\mathrm{A}|} \in \mathscr{L}(\mathrm{A}),$$

so that, in particular, all input states $\rho \in \mathscr{D}(\mathrm{A})$ are sent to the maximally mixed state $\mathrm{Id}/|\mathrm{A}| \in \mathscr{D}(\mathrm{A})$. $\mathcal{R}$ has maximal Kraus rank $|\mathrm{A}|^2$ (because $\tau(\mathcal{R})$ is simply $\mathrm{Id}/|\mathrm{A}|^2$, and hence has rank $|\mathrm{A}|^2$). This was of course to be expected, if adhering to the intuitive idea that the bigger is the Kraus rank of channel, the noisier is the channel. One possible minimal Kraus decomposition for $\mathcal{R}$ is

$$\mathcal{R} : X \in \mathscr{L}(\mathrm{A}) \mapsto \frac{1}{|\mathrm{A}|^2} \sum_{i,j=1}^{|\mathrm{A}|} V_{ij} X V_{ij}^\dagger \in \mathscr{L}(\mathrm{A}),$$

where for each $1 \leqslant i,j \leqslant |\mathrm{A}|$, $V_{ij} = \Sigma_x^j \Sigma_z^k$ with Σ_x, Σ_z the generalized Pauli operators on A. It was initially established in [99] and later improved in [9] that there exist almost randomizing channels with drastically smaller Kraus rank. More specifically, the following was proved: for any $0 < \varepsilon < 1$, the CPTP map $\mathscr{R}$ can be ε-approximated in $(1 \to 1)$-norm by a CPTP map $\widehat{\mathscr{R}}$ with Kraus rank at most $C|\mathrm{A}|/\varepsilon^2$, where $C > 0$ is a universal constant. Actually, something stronger was established, namely

$$\forall\ \rho \in \mathscr{D}(\mathrm{A}),\ \left\|\widehat{\mathscr{R}}(\rho) - \mathscr{R}(\rho)\right\|_\infty \leqslant \frac{\varepsilon}{|\mathrm{A}|},$$

which obviously implies that, for any $1 \leqslant p \leqslant \infty$, $\widehat{\mathscr{R}}$ is an ε-approximation of $\mathscr{R}$ in $(1 \to p)$-norm, in the sense that

$$\forall\ \rho \in \mathscr{D}(\mathrm{A}),\ \left\|\widehat{\mathscr{R}}(\rho) - \mathscr{R}(\rho)\right\|_p \leqslant \frac{\varepsilon}{|\mathrm{A}|^{1-1/p}}. \tag{5.5}$$

The question we investigate here is whether such kind of statement actually holds true for any channel. Note however that, for a channel which is not the fully randomizing one, the notion of approximation in Schatten p-norm appearing in equation (5.5) is maybe not what we would expect as being the "correct" one. In fact, it would seem more accurate to quantify closeness in terms of relative error. Hence, given a CPTP map $\mathscr{N} : \mathscr{L}(\mathrm{A}) \to \mathscr{L}(\mathrm{B})$, we would rather be interested in finding a CPTP map $\widehat{\mathscr{N}} : \mathscr{L}(\mathrm{A}) \to \mathscr{L}(\mathrm{B})$ with Kraus rank as small as possible, and such that

$$\forall\ \rho \in \mathscr{D}(\mathrm{A}),\ \left\|\widehat{\mathscr{N}}(\rho) - \mathscr{N}(\rho)\right\|_p \leqslant \varepsilon \left\|\mathscr{N}(\rho)\right\|_p.$$

5.3 Statement of the main results

Theorem 5.3.1. *Fix $0 < \varepsilon < 1$ and let $\mathscr{N} : \mathscr{L}(\mathrm{A}) \to \mathscr{L}(\mathrm{B})$ be a CPTP map with Kraus rank $|\mathrm{E}| \geqslant |\mathrm{A}|, |\mathrm{B}|$. Then, there exists a CP map $\widehat{\mathscr{N}} : \mathscr{L}(\mathrm{A}) \to \mathscr{L}(\mathrm{B})$ with Kraus rank at most $C \max(|\mathrm{A}|, |\mathrm{B}|) \log(|\mathrm{E}|/\varepsilon)/\varepsilon^2$ (where $C > 0$ is a universal constant) and such that*

$$\forall\ \rho \in \mathscr{D}(\mathrm{A}),\ -\varepsilon\left(\mathscr{N}(\rho) + \frac{\mathrm{Id}}{|\mathrm{B}|}\right) \leqslant \widehat{\mathscr{N}}(\rho) - \mathscr{N}(\rho) \leqslant \varepsilon\left(\mathscr{N}(\rho) + \frac{\mathrm{Id}}{|\mathrm{B}|}\right). \tag{5.6}$$

Remark 5.3.2. *Note that if $\widehat{\mathscr{N}}$ satisfies equation (5.6), then it especially implies that it approximates $\mathscr{N}$ in any Schatten-norm in the following sense*

$$\forall\ p \in \mathbf{N},\ \forall\ \rho \in \mathscr{D}(\mathrm{A}),\ \left\|\widehat{\mathscr{N}}(\rho) - \mathscr{N}(\rho)\right\|_p \leqslant \varepsilon\left(\left\|\mathscr{N}(\rho)\right\|_p + \frac{1}{|\mathrm{B}|^{1-1/p}}\right).$$

In particular, we have the worth pointing out $(1 \to 1)$-norm approximation of $\mathscr{N}$ by $\widehat{\mathscr{N}}$

$$\forall\ \rho \in \mathscr{D}(\mathrm{A}),\ \left\|\widehat{\mathscr{N}}(\rho) - \mathscr{N}(\rho)\right\|_1 \leqslant 2\varepsilon,$$

in which we can further impose that $\widehat{\mathscr{N}}$ is strictly, and not just up to an error 2ε, trace preserving (cf. the proof of Theorem 5.3.1).

One important question at that point is the one of optimality in Theorem 5.3.1. A first obvious observation to make in order to answer it is the following: if a CP map has Kraus-rank s, then it necessarily sends rank 1 inputs on rank at most s outputs. This is of course informative only if s is smaller than the output space dimension. But as we shall see, having this is mind will be useful to prove that certain channels cannot be compressed further than as guaranteed by Theorem 5.3.1.

Our constructions will be based on the existence of so-called tight normalized frames. Namely, for any $N, d \in \mathbf{N}$ with $N \geqslant d$, there exist unit vectors $|\psi_1\rangle, \ldots, |\psi_N\rangle$ in $\mathbf{C}^d$ such that

$$\frac{1}{N}\sum_{k=1}^{N} |\psi_k\rangle\langle\psi_k| = \frac{\mathrm{Id}}{d}.$$

Denoting by $\{|j\rangle,\ 1 \leqslant j \leqslant d\}$ an orthonormal basis of $\mathbf{C}^d$, a possible way of constructing such vectors is e.g. to make the choice

$$\forall\ 1 \leqslant k \leqslant N,\ |\psi_k\rangle = \frac{1}{\sqrt{d}}\sum_{j=1}^{d} e^{2i\pi jk/N}|j\rangle. \tag{5.7}$$

Note that if this so, then any basis vector $|j\rangle$, $1 \leqslant j \leqslant d$, is such that, for each $1 \leqslant k \leqslant N$, $|\langle\psi_k|j\rangle|^2 = 1/d$.

Let us now come back to our objective. What we want to exhibit here are CPTP maps $\mathcal{N} : \mathscr{L}(\mathrm{A}) \to \mathscr{L}(\mathrm{B})$ with either one or the other of the following two properties: if a CP map $\widehat{\mathcal{N}} : \mathscr{L}(\mathrm{A}) \to \mathscr{L}(\mathrm{B})$ satisfies

$$\forall\ R \in \mathscr{H}_+(\mathrm{A}),\ (1-\varepsilon)\mathcal{N}(R) - \varepsilon(\mathrm{Tr}\, R)\frac{\mathrm{Id}}{|\mathrm{B}|} \leqslant \widehat{\mathcal{N}}(R) \leqslant (1+\varepsilon)\mathcal{N}(R) + \varepsilon(\mathrm{Tr}\, R)\frac{\mathrm{Id}}{|\mathrm{B}|}, \tag{5.8}$$

then it necessarily has to be such that either $r_K(\widehat{\mathcal{N}}) \geqslant |\mathrm{A}|$ or $r_K(\widehat{\mathcal{N}}) \geqslant |\mathrm{B}|$. Besides, note that the CP maps $\mathcal{N}, \widehat{\mathcal{N}}$ fulfilling condition (5.8) above is equivalent to the conjugate CP maps $\mathcal{N}^*, \widehat{\mathcal{N}}^*$ fulfilling condition (5.9) below

$$\forall\ R \in \mathscr{H}_+(\mathrm{B}),\ (1-\varepsilon)\mathcal{N}^*(R) - \varepsilon(\mathrm{Tr}\, R)\frac{\mathrm{Id}}{|\mathrm{B}|} \leqslant \widehat{\mathcal{N}}^*(R) \leqslant (1+\varepsilon)\mathcal{N}^*(R) + \varepsilon(\mathrm{Tr}\, R)\frac{\mathrm{Id}}{|\mathrm{B}|}. \tag{5.9}$$

Depending on what we want to establish, it will be more convenient to work with either one or the other of these requirements.

Assume first of all that $|\mathrm{B}| \geqslant |\mathrm{A}|$, and consider $\mathcal{M} : \mathscr{L}(\mathrm{A}) \to \mathscr{L}(\mathrm{B})$ a so-called quantum-classical channel (aka measurement). More specifically, define the CPTP map

$$\mathcal{M} : X \in \mathscr{L}(\mathrm{A}) \mapsto \frac{|\mathrm{A}|}{|\mathrm{B}|}\sum_{i=1}^{|\mathrm{B}|}\langle\psi_i|X|\psi_i\rangle|x_i\rangle\langle x_i| \in \mathscr{L}(\mathrm{B}), \tag{5.10}$$

where $\{|x_i\rangle,\ 1 \leqslant i \leqslant |\mathrm{B}|\}$ is an orthonormal basis of B and $|\psi_1\rangle, \ldots, |\psi_{|\mathrm{B}|}\rangle$ are unit vectors of A, defined in terms of an orthonormal basis $\{|j\rangle,\ 1 \leqslant j \leqslant |\mathrm{A}|\}$ of A as by equation (5.7). Just to connect with the considerations in Chapter 6, note that this tight normalized frame assumption implies that $\{(|\mathrm{A}|/|\mathrm{B}|)|\psi_i\rangle\langle\psi_i|\}_{1\leqslant i\leqslant|\mathrm{B}|}$ forms a rank-1 POVM on A (hence a posteriori the justification of the denomination for $\mathcal{M}$). Setting, for each $1 \leqslant i \leqslant |\mathrm{B}|$, $K_i = \sqrt{|\mathrm{A}|/|\mathrm{B}|}\,|x_i\rangle\langle\psi_i|$, we can clearly re-write $\mathcal{M} : X \in \mathscr{L}(\mathrm{A}) \mapsto \sum_{i=1}^{|\mathrm{B}|} K_i X K_i^\dagger \in \mathscr{L}(\mathrm{B})$, so $r_K(\mathcal{M}) \leqslant |\mathrm{B}|$. And what we actually want to show is that it is even impossible to approximate $\mathcal{M}$ in the sense of Theorem 5.3.1 with strictly less than $|\mathrm{B}|$ Kraus operators. Observe that by construction, say, $|1\rangle$ is such that, for each $1 \leqslant i \leqslant |\mathrm{B}|$, $|\langle\psi_i|1\rangle|^2 = 1/|\mathrm{A}|$, so that $\mathcal{M}(|1\rangle\langle 1|) = \mathrm{Id}/|\mathrm{B}|$. Yet, assume that $\widehat{\mathcal{M}} : \mathscr{L}(\mathrm{A}) \to \mathscr{L}(\mathrm{B})$ is a CPTP map such that $\mathcal{M}, \widehat{\mathcal{M}}$ fulfill equation (5.8) for some $0 < \varepsilon < 1/2$. Then, the l.h.s. of equation (5.8) yields in particular, $\widehat{\mathcal{M}}(|1\rangle\langle 1|) \geqslant (1-2\varepsilon)\,\mathrm{Id}/|\mathrm{B}|$, so that $\widehat{\mathcal{M}}(|1\rangle\langle 1|)$ has to have full rank. And therefore, it cannot be that $r_K(\widehat{\mathcal{M}}) < |\mathrm{B}|$.

Assume now that $|\mathrm{A}| \geqslant |\mathrm{B}|$, and consider $\mathcal{N} : \mathscr{L}(\mathrm{A}) \to \mathscr{L}(\mathrm{B})$ a so-called classical-quantum channel. More specifically, define the CPTP map

$$\mathcal{N} : X \in \mathscr{L}(\mathrm{A}) \mapsto \sum_{i=1}^{|\mathrm{A}|}\langle x_i|X|x_i\rangle|\psi_i\rangle\langle\psi_i| \in \mathscr{L}(\mathrm{B}), \tag{5.11}$$

where $\{|x_i\rangle,\ 1 \leqslant i \leqslant |\mathrm{A}|\}$ is an orthonormal basis of A and $|\psi_1\rangle, \ldots, |\psi_{|\mathrm{A}|}\rangle$ are unit vectors in B. Setting, for each $1 \leqslant i \leqslant |\mathrm{A}|$, $K_i = |\psi_i\rangle\langle x_i|$, we can clearly re-write $\mathcal{N} : X \in \mathscr{L}(\mathrm{A}) \mapsto \sum_{i=1}^{|\mathrm{A}|} K_i X K_i^\dagger \in \mathscr{L}(\mathrm{B})$, so $r_K(\mathcal{N}) \leqslant |\mathrm{A}|$. Now, we want to show that, at least for certain choices of $|\psi_1\rangle, \ldots, |\psi_{|\mathrm{A}|}\rangle$, it is even impossible to approximate $\mathcal{N}$ in the sense of Theorem 5.3.1 with strictly less than $|\mathrm{A}|$ Kraus operators. For that, we impose that they are defined in terms of an orthonormal basis $\{|j\rangle,\ 1 \leqslant j \leqslant |\mathrm{B}|\}$ of B as by equation (5.7). Since the conjugate of $\mathcal{N}$ is the CP unital map

$$\mathcal{N}^* : X \in \mathscr{L}(\mathrm{B}) \mapsto \sum_{i=1}^{|\mathrm{A}|}\langle\psi_i|X|\psi_i\rangle|x_i\rangle\langle x_i| \in \mathscr{L}(\mathrm{A}),$$

we have in this case that $\mathcal{M} = (|\mathrm{B}|/|\mathrm{A}|)\mathcal{N}^*$ is precisely of the form (5.10) (with the roles of A and B switched). Hence, as we already showed, if $\widehat{\mathcal{M}} : \mathscr{L}(\mathrm{B}) \to \mathscr{L}(\mathrm{A})$ is a CPTP map such that $\mathcal{M}, \widehat{\mathcal{M}}$ fulfill equation (5.8) (with the roles of A and B switched) for some $0 < \varepsilon < 1/2$, then it cannot be that $r_K(\widehat{\mathcal{M}}) < |\mathrm{A}|$. This means equivalently that if $\widehat{\mathcal{N}}^* : \mathscr{L}(\mathrm{B}) \to \mathscr{L}(\mathrm{A})$ is a CP map such that $\mathcal{N}^*, \widehat{\mathcal{N}}^*$ fulfill equation (5.9) for some $0 < \varepsilon < 1/2$, then it cannot be that $r_K(\widehat{\mathcal{N}}) = r_K(\widehat{\mathcal{N}}^*) < |\mathrm{A}|$.

Summarizing, we just established that $n \geqslant \max(|\mathrm{A}|, |\mathrm{B}|)$ is for sure necessary in Theorem 5.3.1. But it is not clear whether or not the $\log|\mathrm{E}|$ factor can be removed. In the case of "well-behaved" channels, whose range is only composed of sufficiently mixed states, we can answer affirmatively, which is the content of Theorem 5.3.3 below. However, we leave the question open in general.

Theorem 5.3.3. *Fix $0 < \varepsilon < 1$ and let $\mathcal{N} : \mathscr{L}(\mathrm{A}) \to \mathscr{L}(\mathrm{B})$ be a CPTP map with Kraus rank $|\mathrm{E}| \geqslant |\mathrm{A}|, |\mathrm{B}|$. Then, there exists a CP map $\widehat{\mathcal{N}} : \mathscr{L}(\mathrm{A}) \to \mathscr{L}(\mathrm{B})$ with Kraus rank at most $C \max(|\mathrm{A}|, |\mathrm{B}|)/\varepsilon^2$ (where $C > 0$ is a universal constant) and such that*

$$\sup_{\rho \in \mathscr{D}(\mathrm{A})} \left\| \widehat{\mathcal{N}}(\rho) - \mathcal{N}(\rho) \right\|_\infty \leqslant \varepsilon \sup_{\rho \in \mathscr{D}(\mathrm{A})} \|\mathcal{N}(\rho)\|_\infty .$$

5.4 Proof of the main results

As a crucial step in establishing Theorems 5.3.1 and 5.3.3, we will need a large deviation inequality for sums of independent ψ_1 (aka sub-exponential) random variables, known as Bernstein's inequality. All needed definitions and results concerning ψ_1 random variables are gathered in Chapter 4, Section 4.2. Our application of Bernstein's inequality (recalled as Theorem 4.2.5 in Chapter 4, Section 4.2) to a suitably chosen sum of independent ψ_1 random variables will yield Proposition 5.4.1 below. Note that in the latter, as well as in several other places in the remainder of this chapter, we shall use the following shorthand notation, whenever no confusion is at risk: given a unit vector ϕ in $\mathbf{C}^n$, we also denote by ϕ the corresponding pure state $|\phi\rangle\langle\phi|$ on $\mathbf{C}^n$.

Proposition 5.4.1. *Let $\mathcal{N} : \mathscr{L}(\mathrm{A}) \to \mathscr{L}(\mathrm{B})$ be a CPTP map with Kraus rank $|\mathrm{E}|$, defined by*

$$\forall\ \rho \in \mathscr{D}(\mathrm{A}),\ \mathcal{N}(\rho) = \mathrm{Tr}_{\mathrm{E}} \left[V \rho V^\dagger \right], \tag{5.12}$$

for some isometry $V : \mathrm{A} \hookrightarrow \mathrm{B} \otimes \mathrm{E}$.

For any given unit vector φ in E define next the CP map $\mathcal{N}_\varphi : \mathscr{L}(\mathrm{A}) \to \mathscr{L}(\mathrm{B})$ by

$$\forall\ \rho \in \mathscr{D}(\mathrm{A}),\ \mathcal{N}_\varphi(\rho) = |\mathrm{E}|\ \mathrm{Tr}_{\mathrm{E}} \left[(\mathrm{Id} \otimes \varphi) V \rho V^\dagger (\mathrm{Id} \otimes \varphi) \right]. \tag{5.13}$$

Now, fix unit vectors x in A, y in B, and pick random unit vectors $\varphi_1, \ldots, \varphi_n$ in E, independently and uniformly. Then,

$$\forall\ 0 < \varepsilon < 1,\ \mathbf{P} \left(\left| \frac{1}{n} \sum_{i=1}^n \langle y | \mathcal{N}_{\varphi_i}(x) | y \rangle - \langle y | \mathcal{N}(x) | y \rangle \right| > \varepsilon \langle y | \mathcal{N}(x) | y \rangle \right) \leqslant e^{-cn\varepsilon^2},$$

where $c > 0$ is a universal constant.

In order to derive this concentration result, we will need first of all an estimate on the ψ_1-norm of a certain random variable appearing in our construction. This is the content of Lemma 5.4.2 below.

Lemma 5.4.2. *Fix $d, s \in \mathbf{N}$. Let σ be a state on $\mathbf{C}^d \otimes \mathbf{C}^s$ and y be a unit vector in $\mathbf{C}^d$. Next, for φ a uniformly distributed unit vector in $\mathbf{C}^s$ define the random variable*

$$X_\varphi(\sigma, y) = \mathrm{Tr} \left[y \otimes \varphi\, \sigma \right].$$

Then, $X_\varphi(\sigma, y)$ is a ψ_1 random variable with mean and ψ_1-norm satisfying

$$\mathbf{E}\, X_\varphi(\sigma, y) = \frac{1}{s} \mathrm{Tr} \left[y \otimes \mathrm{Id}\, \sigma \right] \quad \text{and} \quad \|X_\varphi(\sigma, y)\|_{\psi_1} \leqslant \frac{1}{s} \mathrm{Tr} \left[y \otimes \mathrm{Id}\, \sigma \right]. \tag{5.14}$$

Proof. To begin with, recall that, for any $p \in \mathbf{N}$, we have, for φ a uniformly distributed unit vector in $\mathbf{C}^s$,

$$\mathbf{E}\, \varphi^{\otimes p} = \frac{1}{\binom{s+p-1}{p}} P_{\mathrm{Sym}^p(\mathbf{C}^s)},$$

where $P_{\mathrm{Sym}^p(\mathbf{C}^s)}$ denotes the orthogonal projector onto the completely symmetric subspace of $(\mathbf{C}^s)^{\otimes p}$ (see [93] and Chapter 3, Section 3.1, of this manuscript for further details).

Now, setting $\sigma_y = \mathrm{Tr}_{\mathbf{C}^d} \left[y \otimes \mathrm{Id}\, \sigma \right]$, positive operator on $\mathbf{C}^s$, we see that $X_\varphi(\sigma, y) = \mathrm{Tr} \left[\varphi\, \sigma_y \right]$. Hence, we clearly have for a start the first statement in equation (5.14), namely

$$\mathbf{E}\, X_\varphi(\sigma, y) = \frac{1}{s} \mathrm{Tr} \left[\mathrm{Id}\, \sigma_y \right] = \frac{1}{s} \mathrm{Tr} \left[y \otimes \mathrm{Id}\, \sigma \right].$$

What is more, for any $p \in \mathbf{N}$, $|X_\varphi(\sigma, y)|^p = (\mathrm{Tr} \left[\varphi\, \sigma_y \right])^p = \mathrm{Tr} \left[\varphi^{\otimes p}\, \sigma_y^{\otimes p} \right]$. And therefore,

$$\mathbf{E}\, |X_\varphi(\sigma, y)|^p = \frac{1}{\binom{s+p-1}{p}} \mathrm{Tr} \left[P_{\mathrm{Sym}^p(\mathbf{C}^s)} \sigma_y^{\otimes p} \right] \leqslant \frac{1}{\binom{s+p-1}{p}} \mathrm{Tr} \left[\sigma_y^{\otimes p} \right] \leqslant \left(\frac{p}{s} \mathrm{Tr} \left[\sigma_y \right] \right)^p,$$

where the last inequality is simply by the rough bounds $p! \leqslant p^p$ and $(s+p-1)!/(s-1)! \geqslant s^p$. So in the end, we get as wanted the second statement in equation (5.14), namely

$$\|X_\varphi(\sigma, y)\|_{\psi_1} = \sup_{p \in \mathbf{N}} \frac{\left(\mathbf{E}\,|X_\varphi(\sigma, y)|^p\right)^{1/p}}{p} \leqslant \frac{1}{s} \operatorname{Tr}\left[y \otimes \operatorname{Id} \sigma\right].$$

This concludes the proof of Lemma 5.4.2. □

Proof of Proposition 5.4.1. Note first of all that we can obviously re-write

$$\langle y|\mathcal{N}(x)|y\rangle = \operatorname{Tr}\left[y \otimes \operatorname{Id} V x V^\dagger\right] \text{ and } \forall\ \varphi \in S_{\mathrm{E}},\ \langle y|\mathcal{N}_\varphi(x)|y\rangle = |\mathrm{E}|\ \operatorname{Tr}\left[y \otimes \varphi\, V x V^\dagger\right].$$

Next, for each $1 \leqslant i \leqslant n$, define the random variable $Y_i = \langle y|\mathcal{N}_{\varphi_i}(x)|y\rangle$. By Lemma 5.4.2, combined with the observation just made above, we know that these are independent ψ_1 random variables with mean $\langle y|\mathcal{N}(x)|y\rangle$ and ψ_1-norm upper bounded by $\langle y|\mathcal{N}(x)|y\rangle$. So by Bernstein's inequality, recalled as Theorem 4.2.5 in Chapter 4, Section 4.2, we get that

$$\forall\ t > 0,\ \mathbf{P}\left(\left|\frac{1}{n}\sum_{i=1}^n Y_i - \langle y|\mathcal{N}(x)|y\rangle\right| > t\right) \leqslant \exp\left(-c_0 n \min\left(\frac{t^2}{\langle y|\mathcal{N}(x)|y\rangle^2}, \frac{t}{\langle y|\mathcal{N}(x)|y\rangle}\right)\right),$$

where $c_0 > 0$ is a universal constant. And hence,

$$\forall\ 0 < \varepsilon < 1,\ \mathbf{P}\left(\left|\frac{1}{n}\sum_{i=1}^n Y_i - \langle y|\mathcal{N}(x)|y\rangle\right| > \varepsilon\langle y|\mathcal{N}(x)|y\rangle\right) \leqslant e^{-c_0 n \varepsilon^2},$$

which is precisely the result announced in Proposition 5.4.1. □

Having at hand the "fixed x, y" concentration inequality of Proposition 5.4.1, we can now get its "for all x, y" counterparts by a standard net-argument. It appears as the following Propositions 5.4.3 and 5.4.4. Note that the approach is very similar to the one leading to the derivation of Dvoretzky's theorem from Levy's lemma, recalled in Chapter 4, Section 4.2.

Proposition 5.4.3. *Let $\mathcal{N} : \mathscr{L}(\mathrm{A}) \to \mathscr{L}(\mathrm{B})$ be a CPTP map, as characterized by equation (5.12), and for each unit vector φ in E define the CP map $\mathcal{N}_\varphi : \mathscr{L}(\mathrm{A}) \to \mathscr{L}(\mathrm{B})$ as in equation (5.13). Next, for $\varphi_1, \ldots, \varphi_n$ independent uniformly distributed unit vectors in E, set $\mathcal{N}_{\varphi^{(n)}} = \left(\sum_{i=1}^n \mathcal{N}_{\varphi_i}\right)/n$. Then, for any $0 < \varepsilon < 1$,*

$$\mathbf{P}\left(\forall\ x \in S_{\mathrm{A}}, y \in S_{\mathrm{B}},\ \left|\langle y|\mathcal{N}_{\varphi^{(n)}}(x) - \mathcal{N}(x)|y\rangle\right| \leqslant \varepsilon\langle y|\mathcal{N}(x)|y\rangle + \frac{\varepsilon}{|\mathrm{B}|}\right) \geqslant 1 - \left(\frac{24|\mathrm{E}||\mathrm{B}|}{\varepsilon}\right)^{2(|\mathrm{A}|+|\mathrm{B}|)} e^{-cn\varepsilon^2},$$

where $c > 0$ is a universal constant.

Proof. Fix $0 < \alpha, \beta < 1$ and consider $\mathscr{A}_\alpha, \mathscr{B}_\beta$ minimal α, β-nets within the unit spheres of A, B, so that by a standard volumetric argument $|\mathscr{A}_\alpha| \leqslant (3/\alpha)^{2|\mathrm{A}|}, |\mathscr{B}_\beta| \leqslant (3/\beta)^{2|\mathrm{B}|}$ (see Lemma 4.2.3 in Chapter 4, Section 4.2). Then, by Proposition 5.4.1 and the union bound, we get that, for any $\varepsilon > 0$,

$$\mathbf{P}\left(\forall\ x \in \mathscr{A}_\alpha, y \in \mathscr{B}_\beta,\ \left|\langle y|\mathcal{N}_{\varphi^{(n)}}(x) - \mathcal{N}(x)|y\rangle\right| \leqslant \varepsilon\langle y|\mathcal{N}\left(x\right)|y\rangle\right) \geqslant 1 - \left(\frac{3}{\alpha}\right)^{2|\mathrm{A}|}\left(\frac{3}{\beta}\right)^{2|\mathrm{B}|} e^{-cn\varepsilon^2}. \tag{5.15}$$

Now, fix $\varepsilon > 0$ and suppose that $\mathscr{C} : \mathscr{L}(\mathrm{A}) \to \mathscr{L}(\mathrm{B})$ is a Hermiticity-preserving map which is such that

$$\forall\ x \in \mathscr{A}_\alpha,\ \forall\ y \in \mathscr{B}_\beta,\ |\langle y|\mathscr{C}(x)|y\rangle| \leqslant \varepsilon\langle y|\mathcal{N}\left(x\right)|y\rangle. \tag{5.16}$$

Assume that $\mathscr{C}$ additionally satisfies the boundedness property

$$\forall\ x \in S_{\mathrm{A}},\ \forall\ y \in S_{\mathrm{B}},\ |\langle y|\mathscr{C}(x)|y\rangle| \leqslant |\mathrm{E}|. \tag{5.17}$$

Note that if $\mathscr{C}$ is Hermicity-preserving, then for any x, y, $\sup_{v,v'} |\langle y|\mathscr{C}(|v\rangle\langle v'|)|y\rangle| = \sup_v |\langle y|\mathscr{C}(|v\rangle\langle v|)|y\rangle|$ and $\sup_{w,w'} |\langle w|\mathscr{C}(|x\rangle\langle x|)|w'\rangle| = \sup_w |\langle w|\mathscr{C}(|x\rangle\langle x|)|w\rangle|$ (this is because for any X, $\mathscr{C}(X^\dagger) = \mathscr{C}(X)^\dagger$). Hence, it will be useful to us later on to keep in mind that assumption (5.17) is actually equivalent to

$$\forall\ x, x' \in S_{\mathrm{A}},\ \forall\ y, y' \in S_{\mathrm{B}},\ \begin{cases} |\langle y|\mathscr{C}(|x\rangle\langle x'|)|y\rangle| \leqslant |\mathrm{E}| \\ |\langle y|\mathscr{C}(|x\rangle\langle x|)|y'\rangle| \leqslant |\mathrm{E}| \end{cases}.$$

Then, for any unit vectors $x \in S_{\mathrm{A}}$, $y \in S_{\mathrm{B}}$, we know by definition that there exist $\tilde{x} \in \mathscr{A}_\alpha$, $\tilde{y} \in \mathscr{B}_\beta$ such that $\|x - \tilde{x}\| \leqslant \alpha$, $\|y - \tilde{y}\| \leqslant \beta$. Hence, first of all

$$\begin{aligned}|\langle y|\mathscr{C}(|x\rangle\langle x|)|y\rangle| &\leqslant |\langle \tilde{y}|\mathscr{C}(|x\rangle\langle x|)|\tilde{y}\rangle| + |\langle y-\tilde{y}|\mathscr{C}(|x\rangle\langle x|)|\tilde{y}\rangle| + |\langle y|\mathscr{C}(|x\rangle\langle x|)|y-\tilde{y}\rangle| \\ &\leqslant |\langle \tilde{y}|\mathscr{C}(|x\rangle\langle x|)|\tilde{y}\rangle| + 2\beta|\mathrm{E}|,\end{aligned}$$

where the second inequality follows from the boundedness property (5.17) of $\mathscr{C}$, combined with the fact that $\|y - \tilde{y}\| \leqslant \beta$. Then similarly, because $\|x - \tilde{x}\| \leqslant \alpha$,

$$\begin{aligned}|\langle \tilde{y}|\mathscr{C}(|x\rangle\langle x|)|\tilde{y}\rangle| &\leqslant |\langle \tilde{y}|\mathscr{C}(|\tilde{x}\rangle\langle \tilde{x}|)|\tilde{y}\rangle| + |\langle \tilde{y}|\mathscr{C}(|x-\tilde{x}\rangle\langle \tilde{x}|)|\tilde{y}\rangle| + |\langle \tilde{y}|\mathscr{C}(|x\rangle\langle x-\tilde{x}|)|\tilde{y}\rangle| \\ &\leqslant |\langle \tilde{y}|\mathscr{C}(|\tilde{x}\rangle\langle \tilde{x}|)|\tilde{y}\rangle| + 2\alpha|\mathrm{E}|.\end{aligned}$$

Putting together the two previous upper bounds, we see that we actually have

$$|\langle y|\mathscr{C}(|x\rangle\langle x|)|y\rangle| \leqslant |\langle \tilde{y}|\mathscr{C}(|\tilde{x}\rangle\langle \tilde{x}|)|\tilde{y}\rangle| + 2|\mathrm{E}|(\alpha+\beta) \leqslant \varepsilon\langle \tilde{y}|\mathcal{N}(|\tilde{x}\rangle\langle \tilde{x}|)|\tilde{y}\rangle + 2|\mathrm{E}|(\alpha+\beta),$$

where the second inequality is by assumption (5.16) on $\mathscr{C}$. Now, arguing just as before (using this time that $\mathcal{N}$ satisfies the boundedness property $|\langle y|\mathcal{N}(|x\rangle\langle x'|)|y'\rangle| \leqslant 1$ for any $x, x' \in S_{\mathrm{A}}$ and $y, y' \in S_{\mathrm{B}}$), we get

$$\langle \tilde{y}|\mathcal{N}(|\tilde{x}\rangle\langle \tilde{x}|)|\tilde{y}\rangle \leqslant \langle y|\mathcal{N}(|x\rangle\langle x|)|y\rangle + 2(\alpha+\beta).$$

So eventually, what we obtain is

$$|\langle y|\mathscr{C}(x)|y\rangle| \leqslant \varepsilon\big(\langle y|\mathcal{N}(x)|y\rangle + 2(\alpha+\beta)\big) + 2|\mathrm{E}|(\alpha+\beta) \leqslant \varepsilon\langle y|\mathcal{N}(x)|y\rangle + 4|\mathrm{E}|(\alpha+\beta).$$

Therefore, choosing $\alpha = \beta = \varepsilon/(8|\mathrm{E}||\mathrm{B}|)$ (and observing that, by the way $\mathcal{N}_{\varphi^{(n)}}$ is constructed, $\mathcal{N}_{\varphi^{(n)}} - \mathcal{N}$ fulfills condition (5.17)), it follows from equation (5.15) that, for any $\varepsilon > 0$,

$$\mathbf{P}\left(\forall\, x \in S_{\mathrm{A}}, y \in S_{\mathrm{B}},\ \left|\langle y|\mathcal{N}_{\varphi^{(n)}}(x) - \mathcal{N}(x)|y\rangle\right| \leqslant \varepsilon\langle y|\mathcal{N}\left(x\right)|y\rangle + \frac{\varepsilon}{|\mathrm{B}|}\right) \geqslant 1 - \left(\frac{24|\mathrm{E}||\mathrm{B}|}{\varepsilon}\right)^{2(|\mathrm{A}|+|\mathrm{B}|)} e^{-cn\varepsilon^2},$$

which is exactly what we wanted to show. □

Proposition 5.4.4. *Let $\mathcal{N} : \mathscr{L}(\mathrm{A}) \to \mathscr{L}(\mathrm{B})$ be a CPTP map, as characterized by equation (5.12), and for each unit vector φ in E define the CP map $\mathcal{N}_\varphi : \mathscr{L}(\mathrm{A}) \to \mathscr{L}(\mathrm{B})$ as in equation (5.13). Next, for $\varphi_1, \ldots, \varphi_n$ independent uniformly distributed unit vectors in E, set $\mathcal{N}_{\varphi^{(n)}} = \big(\sum_{i=1}^n \mathcal{N}_{\varphi_i}\big)/n$. Then, for any $0 < \varepsilon < 1$,*

$$\mathbf{P}\left(\sup_{x\in S_{\mathrm{A}}, y\in S_{\mathrm{B}}} \left|\langle y|\mathcal{N}_{\varphi^{(n)}}(x) - \mathcal{N}(x)|y\rangle\right| \leqslant \varepsilon \sup_{x\in S_{\mathrm{A}}, y\in S_{\mathrm{B}}} \langle y|\mathcal{N}(x)|y\rangle\right) \geqslant 1 - 225^{|\mathrm{A}|+|\mathrm{B}|} e^{-cn\varepsilon^2},$$

where $c > 0$ is a universal constant.

Proof. We will argue in a way very similar to what was done in the proof of Proposition 5.4.3, and hence skip some of the details here. Again, fix $0 < \alpha, \beta < 1/4$ and consider $\mathscr{A}_\alpha, \mathscr{B}_\beta$ minimal α, β-nets within the unit spheres of A, B. Now, fix $\varepsilon > 0$ and suppose that $\mathscr{C} : \mathscr{L}(\mathrm{A}) \to \mathscr{L}(\mathrm{B})$ is a Hermiticity-preserving map which is such that,

$$\forall\, x \in \mathscr{A}_\alpha,\ \forall\, y \in \mathscr{B}_\beta,\ |\langle y|\mathscr{C}(x)|y\rangle| \leqslant \varepsilon\langle y|\mathcal{N}\left(x\right)|y\rangle.$$

Then, for any unit vectors $x \in S_{\mathrm{A}}$, $y \in S_{\mathrm{B}}$,

$$\begin{aligned}|\langle y|\mathscr{C}(|x\rangle\langle x|)|y\rangle| \leqslant \varepsilon\big(&\langle y|\mathcal{N}(|x\rangle\langle x|)|y\rangle + 2\alpha \sup_{v,v'}\langle y|\mathcal{N}(|v\rangle\langle v'|)|y\rangle + 2\beta \sup_{w,w'}\langle w|\mathcal{N}(|\tilde{x}\rangle\langle \tilde{x}|)|w'\rangle\big) \\ &+ 2\alpha \sup_{v,v'}|\langle \tilde{y}|\mathscr{C}(|v\rangle\langle v'|)|\tilde{y}\rangle| + 2\beta \sup_{w,w'}|\langle w|\mathscr{C}(|x\rangle\langle x|)|w'\rangle|,\end{aligned}$$

where $\tilde{x} \in \mathscr{A}_\alpha$, $\tilde{y} \in \mathscr{B}_\beta$ are such that $\|x - \tilde{x}\| \leqslant \alpha$, $\|y - \tilde{y}\| \leqslant \beta$. And consequently, taking supremum over unit vectors $x \in S_{\mathrm{A}}$, $y \in S_{\mathrm{B}}$, we get

$$\sup_{x,y}|\langle y|\mathscr{C}(x)|y\rangle| \leqslant \varepsilon(1 + 2(\alpha+\beta)) \sup_{x,y}\langle y|\mathcal{N}(x)|y\rangle + 2(\alpha+\beta)\sup_{x,y}|\langle y|\mathscr{C}(x)|y\rangle|,$$

that is equivalently,

$$\sup_{x,y}|\langle y|\mathscr{C}(x)|y\rangle| \leqslant \varepsilon\,\frac{1 + 2(\alpha+\beta)}{1 - 2(\alpha+\beta)} \sup_{x,y}\langle y|\mathcal{N}(x)|y\rangle.$$

Therefore, choosing $\alpha = \beta = 1/5$, so that $(1+2(\alpha+\beta))/(1-2(\alpha+\beta)) = 9$ and $3/\alpha = 3/\beta = 15$, we eventually obtain that, for any $0 < \varepsilon < 1$,

$$\mathbf{P}\left(\sup_{x\in S_{\mathrm{A}}, y\in S_{\mathrm{B}}} \left|\langle y|\mathcal{N}_{\varphi^{(n)}}(x) - \mathcal{N}(x)|y\rangle\right| \leqslant 9\varepsilon \sup_{x\in S_{\mathrm{A}}, y\in S_{\mathrm{B}}} \langle y|\mathcal{N}(x)|y\rangle\right) \geqslant 1 - 15^{2(|\mathrm{A}|+|\mathrm{B}|)} e^{-cn\varepsilon^2},$$

which, after relabelling 9ε in ε, implies precisely the result announced in Proposition 5.4.4. □

Proof of Theorem 5.3.1. Because operator-ordering is preserved by convex combinations, it follows from Proposition 5.4.3 that there exists a universal constant $c > 0$ such that, for any $\varepsilon > 0$,

$$\mathbf{P}\left(\forall\ \rho \in \mathscr{D}(\mathrm{A}),\ \left|\mathcal{N}_{\varphi^{(n)}}(\rho) - \mathcal{N}(\rho)\right| \leqslant \varepsilon\left(\mathcal{N}(\rho) + \frac{\mathrm{Id}}{|\mathrm{B}|}\right)\right) \geqslant 1 - \left(\frac{24|\mathrm{E}||\mathrm{B}|}{\varepsilon}\right)^{2(|\mathrm{A}|+|\mathrm{B}|)} e^{-cn\varepsilon^2}.$$

The r.h.s. of the latter inequality becomes larger than, say, $1/2$ as soon as $n \geqslant C\max(|\mathrm{A}|, |\mathrm{B}|)\log(|\mathrm{E}|/\varepsilon)/\varepsilon^2$, for $C > 0$ some universal constant.

Recapitulating, what we have shown sofar is that there exists a completely positive map $\mathcal{N}^{(n)}$ with Kraus rank $n \leqslant C\max(|\mathrm{A}|, |\mathrm{B}|)\log(|\mathrm{E}|/\varepsilon)/\varepsilon^2$, for $C > 0$ some universal constant, such that,

$$\forall\ \rho \in \mathscr{D}(\mathrm{A}),\ -\varepsilon\left(\mathcal{N}(\rho) + \frac{\mathrm{Id}}{|\mathrm{B}|}\right) \leqslant \mathcal{N}^{(n)}(\rho) - \mathcal{N}(\rho) \leqslant \varepsilon\left(\mathcal{N}(\rho) + \frac{\mathrm{Id}}{|\mathrm{B}|}\right). \tag{5.18}$$

In particular, equation (5.18) implies that, for any $\rho \in \mathscr{D}(\mathrm{A})$, $\left|\mathrm{Tr}\left(\mathcal{N}^{(n)}(\rho)\right) - 1\right| \leqslant 2\varepsilon$, so that $\mathcal{N}^{(n)}$ is almost trace preserving, up to an error 2ε. As a consequence of equation (5.18), we also have

$$\forall\ \rho \in \mathscr{D}(\mathrm{A}),\ \left\|\mathcal{N}^{(n)}(\rho) - \mathcal{N}(\rho)\right\|_1 \leqslant 2\varepsilon, \tag{5.19}$$

and to get only such trace-norm approximation, it is actually possible to impose that $\mathcal{N}^{(n)}$ is strictly trace preserving. Indeed, denote by $\{K_1, \ldots, K_n\}$ a set of Kraus operators for $\mathcal{N}^{(n)}$, and set $S = \sum_{i=1}^n K_i^\dagger K_i$. Equation (5.19) guarantees that $\|S - \mathrm{Id}\|_\infty \leqslant 2\varepsilon$, so that S is in particular invertible, as soon as $\varepsilon < 1/2$. Hence, assume in the sequel that, in fact, $\varepsilon < 1/4$, and consider the completely positive map $\widehat{\mathcal{N}}^{(n)}$ having $\{K_1 S^{-1/2}, \ldots, K_n S^{-1/2}\}$ as a set of Kraus operators, which means that $\widehat{\mathcal{N}}^{(n)}(\cdot) = \mathcal{N}^{(n)}(S^{-1/2}\cdot S^{-1/2})$. The latter is trace preserving by construction, and such that

$$\forall\ \rho \in \mathscr{D}(\mathrm{A}),\ \left\|\widehat{\mathcal{N}}^{(n)}(\rho) - \mathcal{N}(\rho)\right\|_1 \leqslant \left\|\widehat{\mathcal{N}}^{(n)}(\rho) - \mathcal{N}^{(n)}(\rho)\right\|_1 + \left\|\mathcal{N}^{(n)}(\rho) - \mathcal{N}(\rho)\right\|_1 \leqslant 5\varepsilon. \tag{5.20}$$

Indeed, for any $\rho \in \mathscr{D}(\mathrm{A})$, we have the chain of inequalities

$$\|S^{-1/2}\rho S^{-1/2} - \rho\|_1 \leqslant \left(\|S^{-1/2}\|_\infty + \|\mathrm{Id}\|_\infty\right)\|\rho\|_1\|S^{-1/2} - \mathrm{Id}\|_\infty \leqslant (1 + 2\varepsilon + 1)\,2\varepsilon \leqslant 3\varepsilon,$$

where the first inequality follows from the triangle and Hölder inequalities (after simply noticing that, setting $\Delta = \mathrm{Id} - S^{-1/2}$, we can rewrite $S^{-1/2}\rho S^{-1/2} - \rho$ as $\Delta\rho\mathrm{Id} + S^{-1/2}\rho\Delta$), while the second inequality is because, for any $0 < x < 1/4$, $(1+2x)^{-1/2} \geqslant 1 - x$ and $(1-2x)^{-1/2} \leqslant 1 + 2x$, so that $\|S^{-1/2}\|_\infty \leqslant 1 + 2\varepsilon$ and $\|S^{-1/2} - \mathrm{Id}\|_\infty \leqslant 2\varepsilon$. This implies that, for any $\rho \in \mathscr{D}(\mathrm{A})$,

$$\left\|\widehat{\mathcal{N}}^{(n)}(\rho) - \mathcal{N}^{(n)}(\rho)\right\|_1 = \left\|\mathcal{N}^{(n)}\left(S^{-1/2}\rho S^{-1/2} - \rho\right)\right\|_1 \leqslant 3\varepsilon,$$

which, combined with (5.19), justifies the last inequality in (5.20).

This concludes the proof of Theorem 5.3.1 and of Remark 5.3.2 following it. □

Proof of Theorem 5.3.3. By extremality of pure states amongst all states, it follows from Proposition 5.4.4 that there exists a universal constant $c > 0$ such that, for any $\varepsilon > 0$,

$$\mathbf{P}\left(\sup_{\rho\in\mathscr{D}(\mathrm{A})} \left\|\mathcal{N}_{\varphi^{(n)}}(\rho) - \mathcal{N}(\rho)\right\|_\infty \leqslant \varepsilon \sup_{\rho\in\mathscr{D}(\mathrm{A})} \|\mathcal{N}(\rho)\|_\infty\right) \geqslant 1 - 144^{|\mathrm{A}|+|\mathrm{B}|} e^{-cn\varepsilon^2}.$$

The r.h.s. of the latter inequality becomes larger than, say, $1/2$ as soon as $n \geqslant C\max(|\mathrm{A}|, |\mathrm{B}|)/\varepsilon^2$, for $C > 0$ some universal constant. And the proof of Theorem 5.3.3 is thus complete. □

5.5 Consequences and applications

5.5.1 Approximation in terms of output entropies or fidelities

This section gathers some (more or less straightforward) corollaries of Theorem 5.3.1 concerning approximation of quantum channels in other distance measures than the (1→1)-norm distance mostly studied up to now.

Given a state ρ on some Hilbert space H, we define, for any $p \in]1,\infty[$, its Rényi entropy of order p as

$$S_p(\rho) = -\frac{p}{p-1}\log\|\rho\|_p,$$

and the latter definition is extended by continuity to $p \in \{1,\infty\}$ as

$$S_1(\rho) = S(\rho) = -\operatorname{Tr}(\rho\log\rho) \text{ and } S_\infty(\rho) = -\log\lambda_{\max}(\rho).$$

Rényi p-entropies thus measure the amount of information present in a quantum state, generalizing the case $p = 1$ of the von Neumann entropy introduced in Chapter 2, Section 2.5. Besides, given states ρ, σ on some Hilbert space H, their fidelity is defined as $F(\rho,\sigma) = \|\sqrt{\rho}\sqrt{\sigma}\|_1$.

Now, given a channel $\mathcal{N}$, from some input Hilbert space A to some output Hilbert space B, it is quite important to understand quantities such as its minimum output Rényi p-entropy, i.e. $S_p^{\min}(\mathcal{N}) = \min_{\rho\in\mathscr{D}(\mathrm{A})} S_p\big(\mathcal{N}(\rho)\big)$, or its maximum output fidelity with a fixed state σ on B, i.e. $F^{\max}(\mathcal{N},\sigma) = \max_{\rho\in\mathscr{D}(\mathrm{A})} F\big(\mathcal{N}(\rho),\sigma\big)$. Hence the interest of having a channel $\widehat{\mathcal{N}}$ which is simpler than $\mathcal{N}$ but nevertheless shares approximately the same $S_p^{\min}$ and $F^{\max}(\cdot,\sigma)$.

Proposition 5.5.1. *Let $\mathcal{N} : \mathscr{L}(\mathrm{A}) \to \mathscr{L}(\mathrm{B})$ be a CPTP map, and assume that the CP map $\widehat{\mathcal{N}} : \mathscr{L}(\mathrm{A}) \to \mathscr{L}(\mathrm{B})$ satisfies*

$$\forall\ \rho \in \mathscr{D}(\mathrm{A}),\ (1-\varepsilon)\mathcal{N}(\rho) - \varepsilon\frac{\mathrm{Id}}{|\mathrm{B}|} \leqslant \widehat{\mathcal{N}}(\rho) \leqslant (1+\varepsilon)\mathcal{N}(\rho) + \varepsilon\frac{\mathrm{Id}}{|\mathrm{B}|}, \tag{5.21}$$

for some $0 < \varepsilon < 1/2$. Then, for any $p \in]1,\infty]$, we have

$$\forall\ \rho \in \mathscr{D}(\mathrm{A}),\ S_p\big(\mathcal{N}(\rho)\big) - \frac{p}{p-1}2\varepsilon \leqslant S_p\big(\widehat{\mathcal{N}}(\rho)\big) \leqslant S_p\big(\mathcal{N}(\rho)\big) + \frac{p}{p-1}4\varepsilon.$$

Hence, for any $p \in]1,\infty]$, $\widehat{\mathcal{N}}$ is close to $\mathcal{N}$ in terms of output p-entropies, in the sense that

$$\forall\ \rho \in \mathscr{D}(\mathrm{A}),\ \Big|S_p\big(\widehat{\mathcal{N}}(\rho)\big) - S_p\big(\mathcal{N}(\rho)\big)\Big| \leqslant \frac{p}{p-1}4\varepsilon.$$

Proof. Setting $\sigma = \mathcal{N}(\rho)$, $\widehat{\sigma} = \widehat{\mathcal{N}}(\rho)$ and $\tau = \mathrm{Id}/|\mathrm{B}|$, we can re-write equation (5.21) as the two inequalities

$$\widehat{\sigma} \leqslant (1+\varepsilon)\sigma + \varepsilon\tau \text{ and } \sigma \leqslant \frac{1}{1-\varepsilon}\widehat{\sigma} + \frac{\varepsilon}{1-\varepsilon}\tau \leqslant (1+2\varepsilon)\widehat{\sigma} + 2\varepsilon\tau.$$

By operator monotonicity and the triangle inequality for $\|\cdot\|_p$, these imply the two estimates

$$\|\widehat{\sigma}\|_p \leqslant (1+\varepsilon)\|\sigma\|_p + \varepsilon\|\tau\|_p \text{ and } \|\sigma\|_p \leqslant (1+2\varepsilon)\|\widehat{\sigma}\|_p + 2\varepsilon\|\tau\|_p. \tag{5.22}$$

Now, from the first inequality in equation (5.22), we get

$$\log\|\widehat{\sigma}\|_p \leqslant \log\big((1+\varepsilon)\|\sigma\|_p + \varepsilon\|\tau\|_p\big) \leqslant \log(1+\varepsilon) + \log\|\sigma\|_p + \frac{\varepsilon}{1+\varepsilon}\frac{\|\tau\|_p}{\|\sigma\|_p} \leqslant \log\|\sigma\|_p + 2\varepsilon,$$

where we used first that log is non-decreasing, then twice that $\log(1+x) \leqslant x$, and finally that $\|\sigma\|_p \leqslant \|\tau\|_p$. Similarly, we derive from the second inequality in equation (5.22) that

$$\log\|\sigma\|_p \leqslant \log\|\widehat{\sigma}\|_p + 4\varepsilon.$$

Multiplying the two previous inequalities by $-p/(p-1) < 0$, we eventually obtain

$$S_p(\widehat{\sigma}) \geqslant S_p(\sigma) - \frac{p}{p-1}2\varepsilon \text{ and } S_p(\sigma) \geqslant S_p(\widehat{\sigma}) - \frac{p}{p-1}4\varepsilon.$$

And all the conclusions of Proposition 5.5.1 follow. □

Proposition 5.5.2. *Let $\mathcal{N} : \mathscr{L}(\mathrm{A}) \to \mathscr{L}(\mathrm{B})$ be a CPTP map, and assume that the CP map $\widehat{\mathcal{N}} : \mathscr{L}(\mathrm{A}) \to \mathscr{L}(\mathrm{B})$ satisfies*

$$\forall\ \rho \in \mathscr{D}(\mathrm{A}),\ \left\| \widehat{\mathcal{N}}(\rho) - \mathcal{N}(\rho) \right\|_1 \leqslant \frac{2\varepsilon}{\log|\mathrm{B}|}, \tag{5.23}$$

for some $0 < \varepsilon < 1/2$. Then, we have

$$\forall\ \rho \in \mathscr{D}(\mathrm{A}),\ \left| S\big(\widehat{\mathcal{N}}(\rho)\big) - S\big(\mathcal{N}(\rho)\big) \right| \leqslant \varepsilon + \frac{2\varepsilon}{\log|\mathrm{B}|} + \sqrt{\frac{\varepsilon}{\log|\mathrm{B}|}}.$$

Hence, $\widehat{\mathcal{N}}$ is close to $\mathcal{N}$ in terms of output entropies, in the sense that

$$\forall\ \rho \in \mathscr{D}(\mathrm{A}),\ \left| S\big(\widehat{\mathcal{N}}(\rho)\big) - S\big(\mathcal{N}(\rho)\big) \right| \leqslant 4\sqrt{\varepsilon}.$$

Proof. By Fannes-Audenaert inequality [19], equation (5.23) implies that

$$\left| S\big(\widehat{\mathcal{N}}(\rho)\big) - S\big(\mathcal{N}(\rho)\big) \right| \leqslant \varepsilon - \frac{\varepsilon}{\log|\mathrm{B}|} \log\left(\frac{\varepsilon}{\log|\mathrm{B}|}\right) - \left(1 - \frac{\varepsilon}{\log|\mathrm{B}|}\right) \log\left(1 - \frac{\varepsilon}{\log|\mathrm{B}|}\right).$$

Now, for any $0 < x < 1/2$, we have on the one hand $x \log(1/x) \leqslant \sqrt{x}$, while we have on the other hand $\log(1/(1-x)) \leqslant \log(1+2x) \leqslant 2x$ so that $(1-x)\log(1/(1-x)) \leqslant 2x$. All the conclusions of Proposition 5.5.2 then follow. □

Theorem 5.5.3. *Fix $0 < \varepsilon < 1$ and let $\mathcal{N} : \mathscr{L}(\mathrm{A}) \to \mathscr{L}(\mathrm{B})$ be a CPTP map with Kraus rank $|\mathrm{E}| \geqslant |\mathrm{A}|, |\mathrm{B}|$. Then, there exists a CP map $\widehat{\mathcal{N}} : \mathscr{L}(\mathrm{A}) \to \mathscr{L}(\mathrm{B})$ with Kraus rank at most $C \max(|\mathrm{A}|, |\mathrm{B}|) \log(|\mathrm{E}|/\varepsilon)/\varepsilon^2$ (where $C > 0$ is a universal constant) and such that*

$$\forall\ p \in]1, \infty],\ \forall\ \rho \in \mathscr{D}(\mathrm{A}),\ \left| S_p\big(\widehat{\mathcal{N}}(\rho)\big) - S_p\big(\mathcal{N}(\rho)\big) \right| \leqslant \frac{p}{p-1}\varepsilon.$$

Besides, there exists a CPTP map $\widehat{\mathcal{N}}' : \mathscr{L}(\mathrm{A}) \to \mathscr{L}(\mathrm{B})$ with Kraus rank at most $C \max(|\mathrm{A}|, |\mathrm{B}|) \log^5(|\mathrm{E}|/\varepsilon^2)/\varepsilon^4$ (where $C > 0$ is a universal constant) and such that

$$\forall\ \rho \in \mathscr{D}(\mathrm{A}),\ \left| S\big(\widehat{\mathcal{N}}'(\rho)\big) - S\big(\mathcal{N}(\rho)\big) \right| \leqslant \varepsilon. \tag{5.24}$$

Proof. This is a direct consequence of Theorem 5.3.1, combined with Propositions 5.5.1 and 5.5.2. □

We already argued about optimality in Theorem 5.3.1, showing that there indeed exist some CPTP maps $\mathcal{N} : \mathscr{L}(\mathrm{A}) \to \mathscr{L}(\mathrm{B})$ for which at least $|\mathrm{A}|$ or $|\mathrm{B}|$ Kraus operators are needed to approximate them in the sense of equation (5.6). We will now establish that, even to get the weaker notion of approximation of equation (5.24), a Kraus-rank of at least $|\mathrm{A}|$ or $|\mathrm{B}|$ might, in some cases, still be necessary.

Let $\mathcal{N} : X \in \mathscr{L}(\mathrm{A}) \mapsto \mathrm{Tr}_{\mathrm{E}}(VXV^\dagger) \in \mathscr{L}(\mathrm{B})$ be a CPTP map with isometry $V : \mathrm{A} \hookrightarrow \mathrm{B} \otimes \mathrm{E}$. Given $\rho \in \mathscr{D}(\mathrm{A})$, we consider its input entropy $S(\rho)$, its output entropy $S\left(\mathcal{N}(\rho)\right)$, and its entropy exchange $S\left(\rho, \mathcal{N}\right)$. The latter quantity is defined as follows: let $\varphi_{A'A}$ be an extension of ρ_A, $\widetilde{\varphi}_{A'BE} = \left(\mathrm{Id}_{A'} \otimes V_{A\to BE}\right) \varphi_{A'A}$, and set

$$S\left(\rho_A, \mathcal{N}_{A\to B}\right) = S\left(\mathrm{Tr}_{\mathrm{E}}\, \widetilde{\varphi}_{A'BE}\right) = S\left(\mathrm{Tr}_{\mathrm{A'B}}\, \widetilde{\varphi}_{A'BE}\right).$$

By non-negativity of the loss and the noise of a quantum channel, we then have (see [81], Section 4.5)

$$\forall\ \rho \in \mathscr{D}(\mathrm{A}),\ |S(\rho) - S(\mathcal{N}(\rho))| \leqslant S(\rho, \mathcal{N}).$$

Yet, for any $\rho \in \mathscr{D}(\mathrm{A})$, obviously $S(\rho, \mathcal{N}) \leqslant \log|\mathrm{E}|$. And hence as a consequence,

$$\log|\mathrm{E}| \geqslant \max\left\{ |S(\rho) - S(\mathcal{N}(\rho))|\ :\ \rho \in \mathscr{D}(\mathrm{A}) \right\}.$$

In particular, we may derive the two following lower bounds on $|\mathrm{E}|$, for certain CPTP maps $\mathcal{N}$,

$$\exists\ \psi_{\mathrm{A}} :\ \mathcal{N}\left(\psi_A\right) = \frac{\mathrm{Id}_B}{|\mathrm{B}|}\ \Rightarrow\ |\mathrm{E}| \geqslant |\mathrm{B}| \text{ and } \exists\ \psi_B :\ \mathcal{N}\left(\frac{\mathrm{Id}_A}{|\mathrm{A}|}\right) = \psi_B\ \Rightarrow\ |\mathrm{E}| \geqslant |\mathrm{A}|\ . \tag{5.25}$$

And this remains approximately true for an approximation of $\mathcal{N}$. Concretely, let $\widehat{\mathcal{N}} : \mathscr{L}(\mathrm{A}) \to \mathscr{L}(\mathrm{B})$ be a CPTP map such that

$$\forall\ \rho \in \mathscr{D}(\mathrm{A}),\ \left| S\big(\widehat{\mathcal{N}}(\rho)\big) - S\big(\mathcal{N}(\rho)\big) \right| \leqslant \varepsilon.$$

If $\mathcal{N}$ satisfies the first condition in equation (5.25), then

$$S\big(\widehat{\mathcal{N}}\left(\psi_A\right)\big) \geqslant S\big(\mathcal{N}\left(\psi_A\right)\big) - \varepsilon = \log|\mathrm{B}| - \varepsilon, \text{ so that } \log r_K\big(\widehat{\mathcal{N}}\big) \geqslant \log|\mathrm{B}| - \varepsilon, \text{ i.e. } r_K\big(\widehat{\mathcal{N}}\big) \geqslant e^{-\varepsilon}|\mathrm{B}|.$$

And if $\mathcal{N}$ satisfies the second condition in equation (5.25), then

$$S\left(\widehat{\mathcal{N}}\left(\frac{\mathrm{Id}_A}{|\mathrm{A}|}\right)\right) \leqslant S\left(\mathcal{N}\left(\frac{\mathrm{Id}_A}{|\mathrm{A}|}\right)\right) + \varepsilon = \varepsilon, \text{ so that } \log r_K\big(\widehat{\mathcal{N}}\big) \geqslant \log|\mathrm{A}| - \varepsilon, \text{ i.e. } r_K\big(\widehat{\mathcal{N}}\big) \geqslant e^{-\varepsilon}|\mathrm{A}|.$$

Therefore, the conclusion of this study is that, in Theorem 5.5.3, $r_K\big(\widehat{\mathcal{N}}\big) \geqslant (1-\varepsilon)\max(|\mathrm{A}|,|\mathrm{B}|)$ is for sure necessary, in general, to have the entropy approximation (5.24). It additionally tells us that there is a channel-dependent lower bound on $r_K\big(\widehat{\mathcal{N}}\big)$ so that the latter holds, namely

$$r_K\big(\widehat{\mathcal{N}}\big) \geqslant (1-\varepsilon)\max\left\{|S(\rho) - S(\mathcal{N}(\rho))| \ : \ \rho \in \mathscr{D}(\mathrm{A})\right\}.$$

Proposition 5.5.4. *Let $\mathcal{N} : \mathscr{L}(\mathrm{A}) \to \mathscr{L}(\mathrm{B})$ be a CPTP map, and assume that the CP map $\widehat{\mathcal{N}} : \mathscr{L}(\mathrm{A}) \to \mathscr{L}(\mathrm{B})$ satisfies*

$$\forall\ \rho \in \mathscr{D}(\mathrm{A}),\ (1-\varepsilon)\mathcal{N}(\rho) - \varepsilon\frac{\mathrm{Id}}{|\mathrm{B}|} \leqslant \widehat{\mathcal{N}}(\rho) \leqslant (1+\varepsilon)\mathcal{N}(\rho) + \varepsilon\frac{\mathrm{Id}}{|\mathrm{B}|}, \tag{5.26}$$

for some $0 < \varepsilon < 1/2$. Then, $\widehat{\mathcal{N}}$ is close to $\mathcal{N}$ in terms of output fidelities, in the sense that

$$\forall\ \rho \in \mathscr{D}(\mathrm{A}), \forall\ \omega \in \mathscr{D}(\mathrm{B}),\ \left|F\big(\widehat{\mathcal{N}}(\rho),\omega\big) - F\big(\mathcal{N}(\rho),\omega\big)\right| \leqslant \frac{3}{\sqrt{2}}\sqrt{\varepsilon}.$$

Proof. As noted in the proof of Proposition 5.5.1, setting $\sigma = \mathcal{N}(\rho)$, $\widehat{\sigma} = \widehat{\mathcal{N}}(\rho)$ and $\tau = \mathrm{Id}/|\mathrm{B}|$, we can re-write equation (5.26) as the two inequalities $\widehat{\sigma} \leqslant (1+\varepsilon)\sigma + \varepsilon\tau$ and $\sigma \leqslant (1+2\varepsilon)\widehat{\sigma} + 2\varepsilon\tau$. By operator monotonicity of $F(\cdot,\omega)$, and the fact that it is upper bounded by 1, these imply the two estimates

$$F(\widehat{\sigma},\omega) \leqslant \sqrt{1+\varepsilon}\,F(\sigma,\omega) + \sqrt{\varepsilon}\,F(\tau,\omega) \leqslant F(\sigma,\omega) + \frac{\varepsilon}{2} + \sqrt{\varepsilon},$$

$$F(\sigma,\omega) \leqslant \sqrt{1+2\varepsilon}\,F(\widehat{\sigma},\omega) + \sqrt{2\varepsilon}\,F\big(\tau,\omega\big) \leqslant F(\widehat{\sigma},\omega) + \varepsilon + \sqrt{2\varepsilon}.$$

Finally, just observing that $\varepsilon \leqslant \sqrt{\varepsilon/2}$ for $0 < \varepsilon < 1/2$, the conclusion of Proposition 5.5.4 directly follows. □

Theorem 5.5.5. *Fix $0 < \varepsilon < 1$ and let $\mathcal{N} : \mathscr{L}(\mathrm{A}) \to \mathscr{L}(\mathrm{B})$ be a CPTP map with Kraus rank $|\mathrm{E}| \geqslant |\mathrm{A}|, |\mathrm{B}|$. Then, there exists a CP map $\widehat{\mathcal{N}} : \mathscr{L}(\mathrm{A}) \to \mathscr{L}(\mathrm{B})$ with Kraus rank at most $C\max(|\mathrm{A}|,|\mathrm{B}|)\log(|\mathrm{E}|/\varepsilon)/\varepsilon^4$ (where $C > 0$ is a universal constant) and such that*

$$\forall\ \rho \in \mathscr{D}(\mathrm{A}), \forall\ \omega \in \mathscr{D}(\mathrm{B}),\ \left|F\big(\widehat{\mathcal{N}}(\rho),\omega\big) - F\big(\mathcal{N}(\rho),\omega\big)\right| \leqslant \varepsilon.$$

Proof. This is a direct consequence of Theorem 5.3.1, combined with Proposition 5.5.4. □

5.5.2 Destruction of correlations with few resources

It was observed in [99], Section 3, that an ε-randomizing channel (i.e. a channel which is an ε-approximation of the fully randomizing channel) approximately destroys the correlations between the system it acts on and any system the latter might be coupled to, in the following two senses: First of all, a state which is initially just classically correlated becomes almost uncorrelated (or in other words any separable state is sent close to a product state, in 1-norm distance). And second of all, whatever the initial state, the correlations present in it become almost invisible to local observers (or in other words any state is sent close to a product state, in one-way-LOCC-norm distance). Hence, having an ε-randomizing channel with few Kraus operators can be seen as having an efficient way to decouple a system of interest from its environment. Thanks to Theorem 5.3.1, we can generalize these results into Theorem 5.5.6 below. The reader is referred to Chapter 7 for a precise definition of locally restricted measurement norms (such as the one-way-LOCC-norm) and much more on the topic of data-locking and data-hiding.

Theorem 5.5.6. *Let* A, B, C *be Hilbert spaces, and assume that* $d = \max(|\mathrm{A}|, |\mathrm{B}|) < +\infty$. *For any* $0 < \varepsilon < 1$ *and* $\sigma_B^* \in \mathscr{D}(\mathrm{B})$, *there exists a CPTP map* $\widehat{\mathcal{N}} : \mathscr{L}(\mathrm{A}) \to \mathscr{L}(\mathrm{B})$ *with Kraus rank at most* $Cd\log(d/\varepsilon)/\varepsilon^2$ *(where* $C > 0$ *is a universal constant) and such that*

$$\forall\ \rho_{AC} \in \mathcal{S}(\mathrm{A{:}C}),\ \left\|\widehat{\mathcal{N}} \otimes \mathcal{Id}(\rho_{AC}) - \sigma_B^* \otimes \rho_C\right\|_1 \leqslant \varepsilon, \tag{5.27}$$

$$\forall\ \rho_{AC} \in \mathscr{D}(\mathrm{A} \otimes \mathrm{C}),\ \left\|\widehat{\mathcal{N}} \otimes \mathcal{Id}(\rho_{AC}) - \sigma_B^* \otimes \rho_C\right\|_{\mathbf{LOCC}^{\to}(\mathrm{B:C})} \leqslant \varepsilon. \tag{5.28}$$

Proof. Define the completely forgetful CPTP map $\mathcal{N} : X_A \in \mathscr{L}(\mathrm{A}) \mapsto (\mathrm{Tr}\, X_A)\sigma_B^* \in \mathscr{L}(\mathrm{B})$ (i.e. $\mathcal{N}$ sends every input state on the output state σ_B^*). By Theorem 5.3.1, there exists a CPTP map $\widehat{\mathcal{N}} : \mathscr{L}(\mathrm{A}) \to \mathscr{L}(\mathrm{B})$ with Kraus rank at most $Cd\log(d/\varepsilon)/\varepsilon^2$ such that

$$\forall\ \rho_A \in \mathscr{D}(\mathrm{A}),\ \left\|\widehat{\mathcal{N}}(\rho_A) - \mathcal{N}(\rho_A)\right\|_1 \leqslant \varepsilon \text{ i.e. } \left\|\widehat{\mathcal{N}}(\rho_A) - \sigma_B^*\right\|_1 \leqslant \varepsilon.$$

Now, following the exact same route as in the proofs of Lemmas III.1 and III.2 in [99], we get that this implies precisely equations (5.27) and (5.28), respectively. We will therefore only briefly recall the arguments here.

Concerning equation (5.27), let $\rho_{AC} \in \mathcal{S}(\mathrm{A{:}C})$, i.e. $\rho_{AC} = \sum_x p_x \rho_A^{(x)} \otimes \rho_C^{(x)}$. Then,

$$\left\|\widehat{\mathcal{N}} \otimes \mathcal{Id}(\rho_{AC}) - \sigma_B^* \otimes \rho_C\right\|_1 = \left\|\sum_x p_x \left(\widehat{\mathcal{N}}\left(\rho_A^{(x)}\right) - \sigma_B^*\right) \otimes \rho_C^{(x)}\right\|_1 \leqslant \sum_x p_x \left\|\widehat{\mathcal{N}}\left(\rho_A^{(x)}\right) - \sigma_B^*\right\|_1 \leqslant \varepsilon,$$

where the last inequality is because, by assumption, for each x, $\left\|\widehat{\mathcal{N}}\left(\rho_A^{(x)}\right) - \sigma_B^*\right\|_1 \leqslant \varepsilon$, and $\sum_x p_x = 1$.

As for equation (5.28), let $\mathrm{M} = \left(M_B^{(x)} \otimes M_C^{(x)}\right)_x \in \mathbf{LOCC}^{\to}(\mathrm{B{:}C})$, i.e. for each x, $0 \leqslant M_B^{(x)}, M_C^{(x)} \leqslant \mathrm{Id}$, and $\sum_x M_B^{(x)} = \mathrm{Id}$. Then, for any $\rho_{AC} \in \mathscr{D}(\mathrm{A} \otimes \mathrm{C})$,

$$\begin{aligned}
\left\|\widehat{\mathcal{N}} \otimes \mathcal{Id}(\rho_{AC}) - \mathcal{N} \otimes \mathcal{Id}(\rho_{AC})\right\|_{\mathrm{M}} &= \sum_x \left|\mathrm{Tr}\left[M_B^{(x)} \otimes M_C^{(x)} \left(\widehat{\mathcal{N}} \otimes \mathcal{Id}(\rho_{AC}) - \mathcal{N} \otimes \mathcal{Id}(\rho_{AC})\right)\right]\right| \\
&= \sum_x \left|\mathrm{Tr}\left[\left(\widehat{\mathcal{N}}^*\left(M_B^{(x)}\right) - \mathcal{N}^*\left(M_B^{(x)}\right)\right) \otimes M_C^{(x)} \rho_{AC}\right]\right| \\
&\leqslant \sum_x \left\|\widehat{\mathcal{N}}^*\left(M_B^{(x)}\right) - \mathcal{N}^*\left(M_B^{(x)}\right)\right\|_\infty \\
&\leqslant \varepsilon,
\end{aligned}$$

where the next-to-last inequality is because, $\|\rho_{AC}\|_1 \leqslant 1$ and for each x, $\left\|M_C^{(x)}\right\|_\infty \leqslant 1$, while the last inequality is because, by assumption, for each x, $\left\|\widehat{\mathcal{N}}^*\left(M_B^{(x)}\right) - \mathcal{N}^*\left(M_B^{(x)}\right)\right\|_\infty \leqslant \varepsilon\, \mathrm{Tr}\, M_B^{(x)}/|\mathrm{B}|$, and $\sum_x \mathrm{Tr}\, M_B^{(x)} = |\mathrm{B}|$. □

5.5.3 The case of Werner channels

An interesting case to which Theorem 5.3.3 applies is that of the so-called Werner channels. These are defined as the family of CPTP maps

$$\mathcal{W}_\lambda : X \in \mathscr{L}(\mathrm{A}) \mapsto \frac{1}{|\mathrm{A}| + 2\lambda - 1}\left[(\mathrm{Tr}\, X)\mathrm{Id} + (2\lambda - 1)X^T\right] \in \mathscr{L}(\mathrm{A}),\ 0 \leqslant \lambda \leqslant 1.$$

Denoting by π_s and π_a the symmetric and anti-symmetric states on $\mathrm{A} \otimes \mathrm{A}$, it is easy to check that, for each $0 \leqslant \lambda \leqslant 1$, the Choi-Jamiolkowski state $\tau(\mathcal{W}_\lambda)$ associated to $\mathcal{W}_\lambda$ is nothing else than the Werner state $\rho_\lambda = \lambda\pi_s + (1-\lambda)\pi_a$ (see Chapter 3, Section 3.2, for all definitions if need be). Hence, $\mathcal{W}_\lambda$ has Kraus rank $|\mathrm{A}|^2$ whenever $0 < \lambda < 1$, and $|\mathrm{A}|(|\mathrm{A}|+1)/2$, resp. $|\mathrm{A}|(|\mathrm{A}|-1)/2$, when $\lambda = 1$, resp. $\lambda = 0$, i.e. in any case full or almost full Kraus rank. These channels are thus typically of the kind that we would like to compress into more economical ones. What is more, they have the property of having only very mixed output states. Indeed,

$$\max_{\rho \in \mathscr{D}(\mathrm{A})} \|\mathcal{W}_\lambda(\rho)\|_\infty = \begin{cases} 2\lambda/(|\mathrm{A}| + 2\lambda - 1) \text{ if } \lambda \geqslant 1/2 \\ 1/(|\mathrm{A}| + 2\lambda - 1) \text{ if } \lambda < 1/2 \end{cases} \leqslant \frac{2}{|\mathrm{A}|}.$$

So by Theorem 5.3.3, we get that, for each $0 \leqslant \lambda \leqslant 1$, given $0 < \varepsilon < 1$, there exists a CP map $\widehat{\mathcal{W}_\lambda} : \mathscr{L}(\mathrm{A}) \to \mathscr{L}(\mathrm{A})$ with Kraus rank at most $C|\mathrm{A}|/\varepsilon^2$ (where $C > 0$ is a universal constant) such that

$$\forall\ \rho \in \mathscr{D}(\mathrm{A}),\ \left\|\widehat{\mathcal{W}_\lambda}(\rho) - \mathcal{W}_\lambda(\rho)\right\|_\infty \leqslant \frac{\varepsilon}{|\mathrm{A}|}.$$

In words, this means that the Werner CPTP maps can be $(\varepsilon/|\mathrm{A}|)$-approximated in $(1\to\infty)$-norm distance (hence in particular ε-approximated in $(1\to1)$-norm distance) by CP maps having Kraus rank $C|\mathrm{A}|/\varepsilon^2 \ll |\mathrm{A}|^2$.

5.6 Discussion

We have generalized in several senses the result established in [99] and [9]. First, we have shown that it holds for all quantum channels and not only for the fully randomizing one: any CPTP map from $\mathscr{L}(\mathrm{A})$ to $\mathscr{L}(\mathrm{B})$ can be ε-approximated in $(1\to1)$-norm distance by a CPTP map with Kraus rank of order $d\log(d/\varepsilon)/\varepsilon^2$, where $d = \max(|\mathrm{A}|, |\mathrm{B}|)$. Second, we have established that a stronger notion of approximation can actually be proven, namely an ε-ordering of the two CP maps, which allows to derive approximation results in terms of various output quantities (that are tighter than those induced by the rougher norm distance closeness). In the case where the channel under consideration is, as the fully randomizing channel, very noisy (meaning that all output states are very mixed), the extra $\log(d/\varepsilon)$ factor in our result can be removed. However, we do not know if this is true in general. On a related note, our study of optimality shows that there exist channels which cannot be compressed below order d Kraus operators (even to achieve the weakest notions of approximation). But what about channel-dependent lower bounds? For a given channel, would there be a more clever construction than ours (i.e. a non-universal one) that would enable its compression to a number of Kraus operators whose log would be, for instance, of order its maximum input-output entropy difference?

Finally, full or partial derandomization of our construction would be desirable. Here again the main difficulty is that most of the techniques which apply to very noisy channels may fail in general. Let us specify a bit what we mean. In [9], two approximation schemes were proposed for the fully randomizing channel $\mathscr{R} : \mathscr{L}(\mathbf{C}^d) \to \mathscr{L}(\mathbf{C}^d)$. They consisted in taking as Kraus operators $\{U_i/\sqrt{n},\ 1 \leqslant i \leqslant n\}$ with $U_1, \ldots, U_n$ sampled either from the Haar measure on $\mathscr{U}(\mathbf{C}^d)$ or from any other isotropic (aka unitary 1-design) measure on $\mathscr{U}(\mathbf{C}^d)$. It was then shown that, in order to approximate $\mathscr{R}$ up to error ε/d in $(1\to\infty)$-norm, n of order d/ε^2 was enough in the Haar-distributed case and n of order $d\log^6 d/\varepsilon^2$ was enough in the, more general, isotropically-distributed case. The advantage of the second result compared to the first one is that there exist isotropic measures which are much simpler than the Haar measure on $\mathscr{U}(\mathbf{C}^d)$, in particular discrete ones (e.g. the uniform measure over any unitary orthogonal basis of $\mathscr{L}(\mathbf{C}^d)$). Hence, from a practical point of view, generating such a measure is arguably more realistic than generating the Haar measure (the reader is e.g. referred to [36] for a more precise formulation of the claim that implementing a Haar distributed unitary is hard and an extensive discussion on how to approximate such a unitary by a more easily implementable one). Now, if $\mathcal{N} : \mathscr{L}(\mathbf{C}^d) \to \mathscr{L}(\mathbf{C}^d)$ is a channel, with environment $\mathbf{C}^s$, such that $\sup_{\rho\in\mathscr{D}(\mathbf{C}^d)} \|\mathcal{N}(\rho)\|_\infty \leqslant C/d$, then arguments of the same type apply to our construction: to approximate $\mathcal{N}$ up to error ε/d in $(1\to\infty)$-norm by sampling unit vectors in $\mathbf{C}^s$, order d/ε^2 of them is enough if they are Haar-distributed (which is the content of Theorem 5.3.1) and order $d\log^6 d/\varepsilon^2$ of them is enough if they are only assumed to be isotropically-distributed. Here as well, the gain in terms of needed amount of randomness is obvious: there exist isotropic measures which are much simpler to sample from than the Haar-measure on $S_{\mathbf{C}^s}$ (e.g. the uniform measure on any orthonormal basis of $\mathbf{C}^s$). Unfortunately, this whole reasoning (based on Dudley's upper bounding of Bernoulli averages by covering number integrals and on a sharp entropy estimate for the suprema of empirical processes in Banach spaces) fails completely for channels that have some of their outputs which are too pure.

Chapter 6

Zonoids and sparsification of quantum measurements

Based on "Zonoids and sparsification of quantum measurements", in collaboration with G. Aubrun [12].

In this chapter, we establish a connection between zonoids (a concept from classical convex geometry) and the distinguishability norms associated to quantum measurements or POVMs (Positive Operator-Valued Measures), recently introduced in quantum information theory.

This correspondence allows us to state and prove the POVM version of classical results from the local theory of Banach spaces about the approximation of zonoids by zonotopes. We show that on $\mathbf{C}^d$, the uniform POVM (the most symmetric POVM) can be sparsified, i.e. approximated by a discrete POVM, the latter having only $O(d^2)$ outcomes. We also show that similar (but weaker) approximation results actually hold for any POVM on $\mathbf{C}^d$.

By considering an appropriate notion of tensor product for zonoids, we extend our results to the multipartite setting: we show, roughly speaking, that local POVMs may be sparsified locally. In particular, the local uniform POVM on $\mathbf{C}^{d_1} \otimes \cdots \otimes \mathbf{C}^{d_k}$ can be approximated by a discrete POVM which is local and has $O(d_1^2 \times \cdots \times d_k^2)$ outcomes.

6.1 Introduction

A classical result by Lyapounov ([157], Theorem 5.5) asserts that the range of a non-atomic $\mathbf{R}^n$-valued vector measure is closed and convex. Convex sets in $\mathbf{R}^n$ obtained in this way are called zonoids. Zonoids are equivalently characterized as convex sets which can be approximated by finite sums of segments.

Here we consider a special class of vector measures: Positive Operator-Valued Measures (POVMs). In the formalism of quantum mechanics, POVMs represent the most general form of a quantum measurement. Recently, Matthews, Wehner and Winter [137] introduced the distinguishability norm associated to a POVM. This norm has an operational interpretation as the bias of the POVM for the state discrimination problem (a basic task in quantum information theory) and is closely related to the zonoid arising from Lyapounov's theorem.

A well-studied question in high-dimensional convexity is the approximation of zonoids by zonotopes. The series of papers [77, 161, 32, 170] culminates in the following result: any zonoid in $\mathbf{R}^n$ can be approximated by the sum of $O(n \log n)$ segments. The aforementioned connection between POVMs and zonoids allows us to state and prove approximation results for POVMs, which improve on previously known bounds. Precise statements appear as Theorem 6.5.3 and 6.5.4.

This chapter is organized as follows. Section 6.2 introduces POVMs and their associated distinguishability norms. Section 6.3 connects POVMs with zonoids. Section 6.4 introduces a notion of tensor product for POVMs, and the corresponding notion for zonoids. Section 6.5 pushes forward this connection to state the POVM version of approximation results for zonoids, which are proved in Sections 6.7 and 6.8. Section 6.6 provides sparsification results for local POVMs on multipartite systems.

The reader may have a look at Table 6.1, which summarizes analogies between zonoids and POVMs.

Notation

Let us recall a few standard concepts from classical convex geometry that we will need throughout our proofs. Much more on that matter is gathered in Chapter 4, Section 4.1. The support function h_K of a convex compact set $K \subset \mathbf{R}^n$ is the function defined for $x \in \mathbf{R}^n$ by $h_K(x) = \sup\{\langle x, y\rangle \ : \ y \in K\}$. Moreover, for a pair K, L of convex compact sets, the inclusion $K \subset L$ is equivalent to the inequality $h_K \leqslant h_L$. The polar of a convex set $K \subset \mathbf{R}^n$ is $K^\circ = \{x \in \mathbf{R}^n \ : \ \langle x, y\rangle \leqslant 1 \text{ whenever } y \in K\}$. The bipolar theorem (see e.g. [22]) states that $(K^\circ)^\circ$ is the closed convex hull of K and $\{0\}$. A convex body is a convex compact set with non-empty interior.

Whenever we apply tools from convex geometry in the (real) space $\mathscr{H}(\mathbf{C}^d)$ of Hermitian operators on $\mathbf{C}^d$ (e.g. polars or support functions), we use the Hilbert–Schmidt inner product $\langle A, B\rangle \mapsto \mathrm{Tr}(AB)$ to define the Euclidean structure. In that setting, in addition to the general notation specified in Chapter 1, Section 1.3, we introduce the following one: $[-\mathrm{Id}, \mathrm{Id}]$ stands for the set of $A \in \mathscr{H}(\mathbf{C}^d)$ such that $-\mathrm{Id} \leqslant A \leqslant \mathrm{Id}$. In other words $[-\mathrm{Id}, \mathrm{Id}]$ is the Hermitian part of the unit ball for $\|\cdot\|_\infty$ on $\mathscr{L}(\mathbf{C}^d)$.

We also use the following convention: whenever a formula is given for the dimension of a (sub)space, it is tacitly understood that one should take the integer part.

6.2 POVMs and distinguishability norms

In quantum mechanics, the state of a d-dimensional system is described by a positive operator on $\mathbf{C}^d$ with trace 1. The most general form of a measurement that may be performed on such a quantum system is encompassed by the formalism of Positive Operator-Valued Measures (POVMs). Given a set Ω equipped with a σ-algebra $\mathscr{F}$, a POVM on $\mathbf{C}^d$ is a map $\mathrm{M} : \mathscr{F} \to \mathscr{H}_+(\mathbf{C}^d)$ which is σ-additive and such that $\mathrm{M}(\Omega) = \mathrm{Id}$. In this definition the space $(\Omega, \mathscr{F})$ could potentially be infinite, so that the POVMs defined on it would be continuous. However, we often restrict ourselves to the subclass of discrete POVMs, and a main point of this chapter is to substantiate this "continuous to discrete" transition. The reader is referred to Chapter 2, Section 2.1, for the physical motivations behind this mathematical formalism.

A discrete POVM is a POVM in which the underlying σ-algebra $\mathscr{F}$ is required to be finite. In that case there is a finite partition $\Omega = A_1 \cup \cdots \cup A_n$ generating $\mathscr{F}$. The positive operators $M_i = \mathrm{M}(A_i)$ are often referred to as the elements of the POVM, and they satisfy the condition $M_1 + \cdots + M_n = \mathrm{Id}$. We usually identify a discrete POVM with the set of its elements by writing $\mathrm{M} = (M_i)_{1 \leqslant i \leqslant n}$. The index set $\{1, \ldots, n\}$ labels the outcomes of the measurement. The integer n is thus the number of outcomes of M and can be seen as a crude way to measure the complexity of M.

What happens when measuring with a POVM M a quantum system in a state ρ? In the case of a discrete POVM $\mathrm{M} = (M_i)_{1 \leqslant i \leqslant n}$, we know from Born's rule that the outcome i is output with probability $\mathrm{Tr}(\rho M_i)$. This simple formula can be used to quantify the efficiency of a POVM to perform the task of state discrimination. State discrimination can be described as follows: a quantum system is prepared in an unknown state which is either ρ or σ (both hypotheses being a priori equally likely), and we have to guess the unknown state. After measuring it with the discrete POVM $\mathrm{M} = (M_i)_{1 \leqslant i \leqslant n}$, the optimal strategy, based on the maximum likelihood probability, leads to a probability of wrong guess equal to [105, 102]

$$\mathbf{P}_{error} = \frac{1}{2}\left(1 - \frac{1}{2}\sum_{i=1}^{n} |\mathrm{Tr}(\rho M_i) - \mathrm{Tr}(\sigma M_i)|\right).$$

In this context, the quantity $\left(\sum_{i=1}^n |\mathrm{Tr}(\rho M_i) - \mathrm{Tr}(\sigma M_i)|\right)/2$ is therefore called the bias of the POVM M on the state pair (ρ, σ).

Following [137], we introduce a norm on $\mathscr{H}(\mathbf{C}^d)$, called the distinguishability norm associated to M, and defined for $\Delta \in \mathscr{H}(\mathbf{C}^d)$ by

$$\|\Delta\|_{\mathrm{M}} = \sum_{i=1}^{n} |\mathrm{Tr}(\Delta M_i)|. \tag{6.1}$$

It is such that $\mathbf{P}_{error} = (1 - \|\rho - \sigma\|_{\mathrm{M}}/2)/2$, and thus quantifies how powerful the POVM M is in discriminating one state from another with the smallest probability of error.

The terminology "norm" is slightly abusive since one may have $\|\Delta\|_{\mathrm{M}} = 0$ for a nonzero $\Delta \in \mathscr{H}(\mathbf{C}^d)$. The functional $\|\cdot\|_{\mathrm{M}}$ is however always a semi-norm, and it is easy to check that $\|\cdot\|_{\mathrm{M}}$ is a norm if and only if the POVM elements $(M_i)_{1 \leqslant i \leqslant n}$ span $\mathscr{H}(\mathbf{C}^d)$ as a vector space. Such POVMs are called informationally complete in the quantum information literature.

Similarly, the distinguishability norm associated to a general POVM M, defined on a set Ω equipped with a σ-algebra $\mathscr{F}$, is described for $\Delta \in \mathscr{H}(\mathbf{C}^d)$ by

$$\|\Delta\|_{\mathrm{M}} = \|\operatorname{Tr}(\Delta \mathrm{M}(\cdot))\|_{\mathrm{TV}} = \sup_{A \in \mathscr{F}} \big[\operatorname{Tr}(\Delta \mathrm{M}(A)) - \operatorname{Tr}(\Delta \mathrm{M}(\Omega \setminus A))\big] = \sup_{M \in \mathrm{M}(\mathscr{F})} \operatorname{Tr}(\Delta(2M - \mathrm{Id})). \tag{6.2}$$

Here $\|\mu\|_{\mathrm{TV}}$ denotes the total variation of a measure μ. When M is discrete, formulae (6.1) and (6.2) coincide. Note also that the inequality $\|\cdot\|_{\mathrm{M}} \leqslant \|\cdot\|_1$ holds for any POVM M, with equality on $\mathscr{H}_+(\mathbf{C}^d)$.

Given a POVM M, we denote by $B_{\mathrm{M}} = \{\|\cdot\|_{\mathrm{M}} \leqslant 1\}$ the unit ball for the distinguishability norm, and $K_{\mathrm{M}} = (B_{\mathrm{M}})^\circ$ its polar, i.e.

$$K_{\mathrm{M}} = \{A \in \mathscr{H}(\mathbf{C}^d) \ : \ \operatorname{Tr}(AB) \leqslant 1 \text{ whenever } \|B\|_{\mathrm{M}} \leqslant 1\}.$$

The set K_{M} is a compact convex set. Moreover K_{M} has nonempty interior if and only if the POVM M is informationally complete. It follows from the inequality $\|\cdot\|_{\mathrm{M}} \leqslant \|\cdot\|_1$ that K_{M} is always included in the operator interval $[-\mathrm{Id}, \mathrm{Id}]$.

On the other hand, it follows from (6.2) that $B_{\mathrm{M}} = (2\mathrm{M}(\mathscr{F}) - \mathrm{Id})^\circ$, and the bipolar theorem implies that

$$K_{\mathrm{M}} = 2\operatorname{conv}(\mathrm{M}(\mathscr{F})) - \mathrm{Id}. \tag{6.3}$$

By Lyapounov's theorem, the convex hull operation is not needed when M is non-atomic. For a discrete POVM $\mathrm{M} = (M_i)_{1 \leqslant i \leqslant n}$, equation (6.3) may be rewritten in the form

$$K_{\mathrm{M}} = \operatorname{conv}\{\pm M_1\} + \cdots + \operatorname{conv}\{\pm M_n\}, \tag{6.4}$$

where the addition of convex sets should be understood as the Minkowski sum: $A+B = \{a+b \ : \ a \in A,\ b \in B\}$.

We are going to show that POVMs can be sparsified, i.e approximated by discrete POVMs with few outcomes. The terminology "approximation" here refers to the associated distinguishability norms: a POVM M is considered to be "close" to a POVM M$'$ when their distinguishability norms satisfy inequalities of the form

$$(1-\varepsilon)\|\cdot\|_{\mathrm{M}'} \leqslant \|\cdot\|_{\mathrm{M}} \leqslant (1+\varepsilon)\|\cdot\|_{\mathrm{M}'}.$$

This notion of approximation has an operational significance: two POVMs are comparable when both lead to comparable biases when used for any state discrimination task. Let us perhaps stress that point: if one has additional information on the states to be discriminated, it may of course be used to design a POVM specifically efficient for those (one could for instance be interested in the problem of distinguishing pairs of low-rank states [164, 2]).

In this chapter, we study the distinguishability norms from a functional-analytic point of view. We are mostly interested in the asymptotic regime, when the dimension d of the underlying Hilbert space is large.

6.3 POVMs and zonoids

6.3.1 POVMs as probability measures on states

The original definition of a POVM involves an abstract measure space, and the specification of this measure space is irrelevant when considering the distinguishability norms. The following proposition, which is probably well-known, gives a more concrete look at POVMs as probability measures on the set $\mathscr{D}(\mathbf{C}^d)$ of states on $\mathbf{C}^d$.

Proposition 6.3.1. *Let* M *be a POVM on* $\mathbf{C}^d$*. There is a unique Borel probability measure* μ *on* $\mathscr{D}(\mathbf{C}^d)$ *with barycenter equal to* Id/d *and such that, for any* $\Delta \in \mathscr{H}(\mathbf{C}^d)$*,*

$$\|\Delta\|_{\mathrm{M}} = d \int_{\mathscr{D}(\mathbf{C}^d)} |\operatorname{Tr}(\Delta\rho)| \, \mathrm{d}\mu(\rho). \tag{6.5}$$

Conversely, given a Borel probability measure μ *with barycenter equal to* Id/d*, there is a POVM* M *such that* (6.5) *is satisfied.*

Proof. We use the polar decomposition for vector measures, which follows from applying the Radon–Nikodym theorem to vector measures (see [158], Theorem 6.12): a vector measure μ defined on a σ-algebra $\mathscr{F}$ on Ω and taking values in a normed space $(\mathbf{R}^n, \|\cdot\|)$ satisfies $d\mu = h d|\mu|$ for some measurable function $h : \Omega \to \mathbf{R}^n$. Moreover, one has $\|h\| = 1$ $|\mu|$-a.e. Here $|\mu|$ denotes the total variation measure of μ.

Table 6.1: A "dictionary" between zonoids and POVMs

Zonotope which is the Minkowski sum of N segments	Discrete POVM with N outcomes
Zonoid = limit of zonotopes	General POVM = limit of discrete POVMs
Tensor product of zonoids	Local POVM on a multipartite system
Euclidean unit ball B^n = most symmetric zonoid in $\mathbf{R}^n$	Uniform POVM U_d = most symmetric POVM on $\mathbf{C}^d$
"4th moment method" (explicit [156]): $cB^n \subset Z \subset CB^n$, with Z a zonotope which is the sum of $O(n^2)$ segments.	"Approximate 4-design POVM" [2]: explicit sparsification of U_d with $O(d^4)$ outcomes.
Measure concentration (non-explicit [77]): $(1-\varepsilon)B^n \subset Z \subset (1+\varepsilon)B^n$, with Z a zonotope which is the sum of $O_\varepsilon(n)$ segments.	Theorem 6.5.3: a randomly chosen POVM with $O(d^2)$ outcomes is a sparsification of U_d.
Derandomization [82, 136, 114]	?
Any zonoid in $\mathbf{R}^n$ can be approximated by a zonotope which is the sum of $O(n \log n)$ segments [170].	Theorem 6.5.4: any POVM on $\mathbf{C}^d$ can be sparsified into a sub-POVM with $O(d^2 \log d)$ outcomes.

Let M be a POVM on $\mathbf{C}^d$, defined on a σ-algebra $\mathcal{F}$ on Ω. We equip $\mathscr{H}(\mathbf{C}^d)$ with the trace norm, so that we simply have $|\mathrm{M}| = \operatorname{Tr} \mathrm{M}$ and $|\mathrm{M}|(\Omega) = d$. The polar decomposition yields a measurable function $h : \Omega \to \mathscr{H}(\mathbf{C}^d)$ such that $\|h\|_1 = 1$ $|\mathrm{M}|$-a.e. Moreover, the fact that $\mathrm{M}(\mathcal{F}) \subset \mathscr{H}_+(\mathbf{C}^d)$ implies that $h \in \mathscr{H}_+(\mathbf{C}^d)$ $|\mathrm{M}|$-a.e. Let μ be the push forward of $|\mathrm{M}|/d$ under the map h. We have

$$\mathrm{Id} = \mathrm{M}(\Omega) = \int_\Omega h \, \mathrm{d}|\mathrm{M}| = d \int_{\mathscr{H}(\mathbf{C}^d)} \rho \, \mathrm{d}\mu(\rho).$$

And since $h \in \mathscr{D}(\mathbf{C}^d)$ a.e., μ is indeed a Borel probability measure on $\mathscr{D}(\mathbf{C}^d)$, with barycenter equal to Id/d. Finally, for any $\Delta \in \mathscr{H}(\mathbf{C}^d)$,

$$\|\Delta\|_{\mathrm{M}} = \int_\Omega |\operatorname{Tr}(\Delta h)| \, \mathrm{d}|\mathrm{M}| = d \int_{\mathscr{D}(\mathbf{C}^d)} |\operatorname{Tr}(\Delta \rho)| \, \mathrm{d}\mu(\rho).$$

We postpone the proof of uniqueness to the next subsection (see after Proposition 6.3.5).

Conversely, given a Borel probability measure μ on $\mathscr{D}(\mathbf{C}^d)$ with barycenter at Id/d, consider the vector measure $\mathrm{M} : \mathscr{B} \to \mathscr{H}(\mathbf{C}^d)$, where $\mathscr{B}$ is the Borel σ-algebra on $\mathscr{D}(\mathbf{C}^d)$, defined by

$$\mathrm{M}(A) = d \int_A \rho \, \mathrm{d}\mu(\rho).$$

It is easily checked that M is a POVM and that formula (6.5) is satisfied. □

Note that in the case of a discrete POVM $\mathrm{M} = (M_i)_{1 \leqslant i \leqslant n}$, the corresponding probability measure is

$$\mu = \frac{1}{d} \sum_{i=1}^n (\operatorname{Tr} M_i) \; \delta_{M_i / \operatorname{Tr} M_i}.$$

Corollary 6.3.2. *Given a POVM* M *on* $\mathbf{C}^d$*, there is a sequence* $(\mathrm{M}_n)_n$ *of discrete POVMs such that* K_{M_n} *converges to* K_{M} *in Hausdorff distance. Moreover, if* μ *(resp.* μ_n*) denotes the probability measure on* $\mathscr{D}(\mathbf{C}^d)$ *associated to* M *(resp. to* M_n*) as in* (6.5)*, we can guarantee that the support of* μ_n *is contained into the support of* μ*.*

Proof. Let μ be the probability measure associated to M. Given n, let $(Q_k)_k$ be a finite partition of $\mathscr{D}(\mathbf{C}^d)$ into sets of diameter at most $1/n$ with respect to the trace norm. Let $\rho_k \in \mathscr{D}(\mathbf{C}^d)$ be the barycenter of the restriction of μ to Q_k (only defined when $\mu(Q_k) > 0$). The probability measure

$$\mu_n = \sum_k \mu(Q_k) \delta_{\rho_k}$$

has the same barycenter as μ, and the associated POVM M_n satisfies

$$\left|h_{K_{\mathrm{M}}}(\Delta) - h_{K_{\mathrm{M}_n}}(\Delta)\right| \leqslant d\frac{\|\Delta\|_\infty}{n},$$

and therefore $K_{\mathrm{M_n}}$ converges to K_{M}.

The condition on the supports can be enforced by changing slightly the definition of μ_n. For each k we can write $\rho_k = \sum \lambda_{k,j}\rho_{k,j}$, where $(\lambda_{k,j})_j$ is a convex combination and $(\rho_{k,j})_j$ belong to the support of μ restricted to Q_k. The measure

$$\mu'_n = \sum_k \mu(Q_k) \sum_j \lambda_{k,j}\delta_{\rho_{k,j}}$$

satisfies the same properties as μ_n, and its support is contained into the support of μ. □

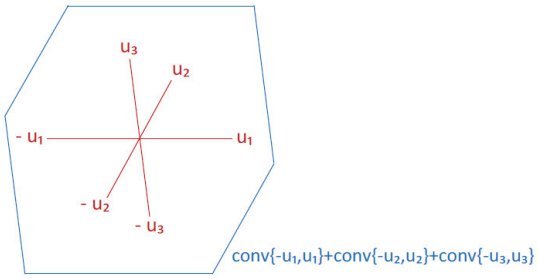
6.3.2 POVMs and zonoids

We connect here POVMs with zonoids, which form an important family of convex bodies (see [31, 163, 82] for surveys on zonoids to which we refer for all the material presented here). A zonotope $Z \subset \mathbf{R}^n$ is a closed convex set which can be written as the Minkowski sum of finitely many segments, i.e. such that there exist finite sets of vectors $(u_i)_{1\leqslant i\leqslant N}$ and $(v_i)_{1\leqslant i\leqslant N}$ in $\mathbf{R}^n$ such that

$$Z = \mathrm{conv}\{u_1, v_1\} + \cdots + \mathrm{conv}\{u_N, v_N\}. \tag{6.6}$$

A zonoid is a closed convex set which can be approximated by zonotopes (with respect to the Hausdorff distance). Every zonoid has a center of symmetry, and therefore can be translated into a (centrally) symmetric zonoid. Note that for a centrally symmetric zonotope, we can choose $v_i = -u_i$ in (6.6). An example of symmetric zonotope in $\mathbf{R}^2$ appears in Figure 6.1.

Figure 6.1: A symmetric zonotope in $\mathbf{R}^2$



Here are equivalent characterizations of zonoids.

Proposition 6.3.3. *Let $K \subset \mathbf{R}^n$ be a symmetric closed convex set. The following are equivalent.*

(i) K is a zonoid.

(ii) There is a Borel positive measure ν on the Euclidean unit sphere S^{n-1} which is even (i.e. such that $\nu(A) = \nu(-A)$ for any Borel set $A \subset S^{n-1}$) and such that, for every $x \in \mathbf{R}^n$,

$$h_K(x) = \int_{S^{n-1}} |\langle x, \theta\rangle| \, \mathrm{d}\nu(\theta). \tag{6.7}$$

(iii) There is a vector measure $\mu : (\Omega, \mathcal{F}) \to \mathbf{R}^n$ such that $K = \mu(\mathcal{F})$.

Moreover, when these conditions are satisfied, the measure ν is unique.

Remark 6.3.4. *Having the measure ν supported on the sphere and be even is only a matter of normalization and a way to enforce uniqueness: if ν is a Borel measure on $\mathbf{R}^n$ for which linear forms are integrable, there is a symmetric zonoid $K \subset \mathbf{R}^n$ such that*

$$h_K(x) = \int_{\mathbf{R}^n} |\langle x, y\rangle| \, \mathrm{d}\nu(y).$$

As an immediate consequence, we characterize which subsets of $[-\mathrm{Id}, \mathrm{Id}]$ arise as K_{M} for some POVM M.

Proposition 6.3.5. *Let $K \subset \mathscr{H}(\mathbf{C}^d)$ be a symmetric closed convex set. Then the following are equivalent.*

(i) K is a zonoid such that $K \subset [-\mathrm{Id}, \mathrm{Id}]$ and $\pm\mathrm{Id} \in K$.

(ii) There exists a POVM M on $\mathbf{C}^d$ such that $K = K_{\mathrm{M}}$.

Moreover, K is a zonotope only if the POVM M can be chosen to be discrete.

Proof. Let K be a zonoid such that $\pm\mathrm{Id} \in K \subset [-\mathrm{Id}, \mathrm{Id}]$. From Proposition 6.3.3, there is a vector measure μ defined on a σ-algebra $\mathscr{F}$ on a set Ω, whose range is K. Let $A \in \mathscr{F}$ such that $\mu(A) = -\mathrm{Id}$. The vector measure M defined for $B \in \mathscr{F}$ by

$$\mathrm{M}(B) = \frac{1}{2}\left(\mu(B \setminus A) - \mu(B \cap A)\right) = \frac{1}{2}\left(\mu(B \Delta A) + \mathrm{Id}\right)$$

is a POVM. Indeed, its range, which equals $(K + \mathrm{Id})/2$, lies inside the positive semidefinite cone, and contains Id. We get from (6.3) that $K_{\mathrm{M}} = K$.

Conversely, for any POVM M, formula (6.3) implies that $\pm\mathrm{Id} \in K \subset [-\mathrm{Id}, \mathrm{Id}]$. The fact that K is a zonoid follows, using the general fact that the convex hull of the range of a vector measure is a zonoid (see [31], Theorem 1.6).

In the case of zonotopes and discrete POVMs, these arguments have more elementary analogues which we do not repeat. □

We can now argue about the uniqueness part in Proposition 6.3.1. This is indeed a consequence of the uniqueness of the measure associated to a zonoid in Proposition 6.3.3: after rescaling and symmetrization, a measure μ on $\mathscr{D}(\mathbf{C}^d)$ satisfying (6.5) naturally induces a measure ν on the Hilbert–Schmidt sphere satisfying (6.7) for $K = K_{\mathrm{M}}$.

Another characterization of zonoids involves the Banach space $L^1 = L^1([0,1])$. A symmetric convex body K is a zonoid if and only if the normed space $(\mathbf{R}^n, h_K)$ embeds isometrically into L^1. Therefore, Proposition 6.3.5 can be restated as a characterization of distinguishability norms on $\mathscr{H}(\mathbf{C}^d)$.

Corollary 6.3.6. *Let $\|\cdot\|$ be a norm on $\mathscr{H}(\mathbf{C}^d)$. The following are equivalent*

(i) There is POVM M on $\mathbf{C}^d$ such that $\|\cdot\| = \|\cdot\|_{\mathrm{M}}$.

(i) The normed space $(\mathscr{H}(\mathbf{C}^d), \|\cdot\|)$ is isometric to a subspace of L^1, and the following inequality is satisfied for any $\Delta \in \mathscr{H}(\mathbf{C}^d)$

$$|\operatorname{Tr}\Delta| \leqslant \|\Delta\| \leqslant \operatorname{Tr}|\Delta|.$$

6.4 Local POVMs and tensor products of zonoids

6.4.1 Tensor products for zonoids

There is a natural notion of tensor product for subspaces of L^1 which appeared in the Banach space literature (see e.g. [76]).

Definition 6.4.1. *Let X, Y be two Banach spaces which can be embedded isometrically into L^1, i.e. such that there exist linear norm-preserving maps $i : X \to L^1(\mu)$ and $j : Y \to L^1(\nu)$. Then, the 1-tensor product of X and Y is defined as the completion of the algebraic tensor product $X \otimes Y$ for the norm*

$$\left\| \sum_k x_k \otimes y_k \right\|_{X \otimes^1 Y} = \int\int \left| \sum_k i(x_k)(s) j(y_k)(t) \right| \mathrm{d}\mu(s)\mathrm{d}\nu(t).$$

It can be checked that the norm above is well-defined and does not depend on the particular choice of the embeddings i, j (see e.g. [76] or Lemma 2 in [155]).

In the finite-dimensional case, subspaces of L^1 are connected to zonoids. Therefore, Definition 6.4.1 leads naturally to a notion of tensor product for (symmetric) zonoids.

Definition 6.4.2. *Let $K \subset \mathbf{R}^m$ and $L \subset \mathbf{R}^n$ be two symmetric zonoids, and suppose that ν_K and ν_L are Borel measures on S^{m-1} and S^{n-1} respectively, such that for any $x \in \mathbf{R}^m$ and $y \in \mathbf{R}^n$,*

$$h_K(x) = \int_{S^{m-1}} |\langle x, \theta \rangle| \, \mathrm{d}\nu_K(\theta) \quad \text{and} \quad h_L(y) = \int_{S^{n-1}} |\langle y, \phi \rangle| \, \mathrm{d}\nu_L(\phi).$$

The zonoid tensor product of K and L is defined as the zonoid $K \otimes^Z L \subset \mathbf{R}^n \otimes \mathbf{R}^m$ whose support function satisfies

$$h_{K \otimes^Z L}(z) = \int_{S^{m-1}} \int_{S^{n-1}} |\langle z, \theta \otimes \phi \rangle| \, \mathrm{d}\nu_K(\theta) \mathrm{d}\nu_L(\phi) \tag{6.8}$$

for any $z \in \mathbf{R}^m \otimes \mathbf{R}^n$.

As in Definition 6.4.1, this construction does not depend on the choice of the measures ν_K and ν_L. This can be seen directly: given $z \in \mathbf{R}^m \otimes \mathbf{R}^n$ and $\phi \in S^{n-1}$, set $\widetilde{z}(\phi) = (\mathrm{Id} \otimes \langle \phi |)(z)$. We have

$$h_{K \otimes^Z L}(z) = \int_{S^{n-1}} h_K(\widetilde{z}(\phi)) \, \mathrm{d}\nu_L(\phi), \tag{6.9}$$

and therefore $K \otimes^Z L$ does not depend on ν_K. The same argument applies for ν_L.

In the case of zonotopes, the zonoid tensor product takes a simpler form :

$$\left(\sum_i \mathrm{conv}\{\pm v_i\} \right) \otimes^Z \left(\sum_j \mathrm{conv}\{\pm w_j\} \right) = \sum_i \sum_j \mathrm{conv}\{\pm v_i \otimes w_j\}.$$

Here is a first simple property of the zonoid tensor product.

Lemma 6.4.3. *Given symmetric zonoids K, L and linear maps S, T, we have*

$$S(K) \otimes^Z T(L) = (S \otimes T)(K \otimes^Z L)$$

Additionally, and crucially for the applications we have in mind, the zonoid tensor product is compatible with inclusions.

Lemma 6.4.4. *Let K, K' be two symmetric zonoids in $\mathbf{R}^m$ with $K \subset K'$, and let L, L' be two symmetric zonoids in $\mathbf{R}^n$ with $L \subset L'$. Then*

$$K \otimes^Z L \subset K' \otimes^Z L'.$$

Proof. This is a special case of Lemma 2 in [155]. Here is a proof in the language of zonoids. We may assume that $L = L'$, the general case following then by arguing that $K \otimes^Z L \subset K' \otimes^Z L \subset K' \otimes^Z L'$.

In terms of support functions, we are thus reduced to showing that the inequality $h_K \leqslant h_{K'}$ implies the inequality $h_{K \otimes^Z L} \leqslant h_{K' \otimes^Z L}$, which is an easy consequence of (6.9). □

Suppose that $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$ are Banach spaces with Euclidean norms, i.e. induced by some inner products $\langle \cdot, \cdot \rangle_X$ and $\langle \cdot, \cdot \rangle_Y$. Their Euclidean tensor product $X \otimes^2 Y$ is defined (after completion) by the norm induced by the inner product on the algebraic tensor product which satisfies

$$\langle x \otimes y, x' \otimes y' \rangle = \langle x, x' \rangle_X \langle y, y' \rangle_Y.$$

It turns out that, for Euclidean norms, the tensor norms $\otimes^1$ and $\otimes^2$ are equivalent.

Proposition 6.4.5 (see [155, 28])**.** *If X and Y are two Banach spaces equipped with Euclidean norms, then*

$$\sqrt{\frac{2}{\pi}} \|\cdot\|_{X \otimes^2 Y} \leqslant \|\cdot\|_{X \otimes^1 Y} \leqslant \|\cdot\|_{X \otimes^2 Y}.$$

6.4.2 Local POVMs

In quantum mechanics, when a system is shared by several parties, the underlying global Hilbert space is the tensor product of the local Hilbert spaces corresponding to each of the subsystems. A physically relevant class of POVMs on such a multipartite system is the one of local POVMs, describing the situation where each party is only able to perform measurements on his own subsystem (cf. Chapter 7, which is entirely dedicated to this topic of POVMs satisfying locality constraints).

Definition 6.4.6. *For $i = 1, 2$, let M_i denote a POVM on $\mathbf{C}^{d_i}$, defined on a σ-algebra $\mathscr{F}_i$ on a set Ω_i. The tensor POVM $\mathrm{M}_1 \otimes \mathrm{M}_2$ is the unique map defined on the product σ-algebra $\mathscr{F}_1 \otimes \mathscr{F}_2$ on $\Omega_1 \times \Omega_2$, and such that*

$$(\mathrm{M}_1 \otimes \mathrm{M}_2)(A_1 \times A_2) = \mathrm{M}_1(A_1) \otimes \mathrm{M}_2(A_2)$$

for every $A_1 \in \mathscr{F}_1, A_2 \in \mathscr{F}_2$. By construction, $\mathrm{M}_1 \otimes \mathrm{M}_2$ is a POVM on $\mathbf{C}^{d_1} \otimes \mathbf{C}^{d_2}$.

In the discrete case, this definition becomes more transparent: if $\mathrm{M} = (M_i)_{1 \leqslant i \leqslant m}$ and $\mathrm{N} = (N_j)_{1 \leqslant j \leqslant n}$ are discrete POVMs, then $\mathrm{M} \otimes \mathrm{N}$ is also discrete, and

$$\mathrm{M} \otimes \mathrm{N} = (M_i \otimes N_j)_{1 \leqslant i \leqslant m, 1 \leqslant j \leqslant n}.$$

POVMs on $\mathbf{C}^{d_1} \otimes \mathbf{C}^{d_2}$ which can be decomposed as tensor product of two POVMs are called local POVMs. If we identify the POVMs M_1 and M_2 with measures μ_1 and μ_2 as in Proposition 6.3.1, then the measure corresponding to $\mathrm{M}_1 \otimes \mathrm{M}_2$ is the image of the product measure $\mu_1 \times \mu_2$ under the map $(\rho, \sigma) \mapsto \rho \otimes \sigma$. It thus follows that

Proposition 6.4.7. *If M and N are two POVMs, then $\| \cdot \|_{\mathrm{M} \otimes \mathrm{N}} = \| \cdot \|_{\mathrm{M}} \otimes^1 \| \cdot \|_{\mathrm{N}}$ and $K_{\mathrm{M} \otimes \mathrm{N}} = K_{\mathrm{M}} \otimes^Z K_{\mathrm{N}}$.*

These definitions and statements are given here only in the bipartite case for the sake of clarity, but can be extended to the situation where a system is shared between any number k of parties.

6.5 Sparsifying POVMs

6.5.1 The uniform POVM

It has been proved in [137] that, in several senses, the "most efficient" POVM on $\mathbf{C}^d$ is the "most symmetric" one, i.e. the uniform POVM U_d, which corresponds to the uniform measure on the set of pure states in the representation (6.5) from Proposition 6.3.5.

The corresponding norm is

$$\|\Delta\|_{\mathrm{U}_d} = d\, \mathbf{E}\, |\langle \psi | \Delta | \psi \rangle|, \tag{6.10}$$

where ψ is a random Haar-distributed unit vector on $\mathbf{C}^d$.

An important property is that the norm $\| \cdot \|_{\mathrm{U}_d}$ is equivalent to a "modified" Hilbert–Schmidt norm.

Proposition 6.5.1 ([95, 130])**.** *For every $\Delta \in \mathscr{H}(\mathbf{C}^d)$, we have*

$$\frac{1}{\sqrt{18}} \|\Delta\|_{2(1)} \leqslant \|\Delta\|_{\mathrm{U}_d} \leqslant \|\Delta\|_{2(1)}, \tag{6.11}$$

where the norm $\| \cdot \|_{2(1)}$ is defined as

$$\|\Delta\|_{2(1)} = \sqrt{\mathrm{Tr}(\Delta^2) + (\mathrm{Tr}\Delta)^2}. \tag{6.12}$$

One can check that $\|\Delta\|_{2(1)}$ equals the L^2 norm of the random variable $\langle g | \Delta | g \rangle$, where g is a standard Gaussian vector in $\mathbf{C}^d$, while the L^1 norm of this random variable is nothing else than $\|\Delta\|_{\mathrm{U}_d}$. Therefore Proposition 6.5.1 can be seen as a reverse Hölder inequality, and an interesting problem would be to find the optimal constant in that inequality (the factor $\sqrt{18}$ is presumably far from optimal).

This dimension-free lower bound on the distinguishing power of the uniform POVM is of interest in quantum information theory. One could cite as one of its applications the possibility to establish lower bounds on the dimensionality reduction of quantum states [95]. However, from a computational or algorithmic point of view, this statement involving a continuous POVM is of no practical use. There has been interest therefore in the

question of sparsifying U_d, i.e. of finding a discrete POVM, with as few outcomes as possible, which would be equivalent to U_d in terms of discriminating efficiency. Examples of such constructions arise from the theory of projective 4-designs.

Given an integer $t \geqslant 1$, an (exact) t-design is a finitely supported probability measure μ on $S_{\mathbf{C}^d}$ such that

$$\int_{S_{\mathbf{C}^d}} |\psi\rangle\langle\psi|^{\otimes t}\,\mathrm{d}\mu(\psi) = \int_{S_{\mathbf{C}^d}} |\psi\rangle\langle\psi|^{\otimes t}\,\mathrm{d}\sigma(\psi) = \binom{d+t-1}{t}^{-1} P_{\mathrm{Sym}^t(\mathbf{C}^d)}.$$

Here, σ denotes the Haar probability measure on $S_{\mathbf{C}^d}$, and $P_{\mathrm{Sym}^t(\mathbf{C}^d)}$ denotes the orthogonal projection onto the symmetric subspace $\mathrm{Sym}^t(\mathbf{C}^d) \subset (\mathbf{C}^d)^{\otimes t}$ (see Chapter 3, Section 3.1, for further details).

Note that a t-design is also a t'-design for any $t' \leqslant t$. Let μ be a 1-design. The map $\psi \mapsto |\psi\rangle\langle\psi|$ pushes forward μ into a measure $\tilde{\mu}$ on the set of (pure) states, with barycenter equal to Id/d. By Proposition 6.3.5, this measure corresponds to a POVM, and in the following we identify t-designs with the associated POVMs. For example the uniform POVM U_d is a t-design for any t.

This notion can be relaxed: define an ε-approximate t-design to be a finitely supported measure μ on $S_{\mathbf{C}^d}$ such that

$$(1-\varepsilon)\int_{S_{\mathbf{C}^d}} |\psi\rangle\langle\psi|^{\otimes t}\,\mathrm{d}\sigma(\psi) \leqslant \int_{S_{\mathbf{C}^d}} |\psi\rangle\langle\psi|^{\otimes t}\,\mathrm{d}\mu(\psi) \leqslant (1+\varepsilon)\int_{S_{\mathbf{C}^d}} |\psi\rangle\langle\psi|^{\otimes t}\,\mathrm{d}\sigma(\psi).$$

It has been proved in [2] that a 4-design (exact or approximate) supported on N points yields a POVM M with N outcomes such that

$$C^{-1}\|\cdot\|_{\mathrm{U}_d} \leqslant \|\cdot\|_{\mathrm{M}} \leqslant C\|\cdot\|_{\mathrm{U}_d} \tag{6.13}$$

for some constant C. The proof is based on the fourth moment method, which is used to control the first absolute moment of a random variable by its second and fourth moments.

Now, what is the minimal cardinality of a 4-design? The support of any exact or ε-approximate (provided $\varepsilon < 1$) 4-design must contain at least $\dim(\mathrm{Sym}^4(\mathbf{C}^d)) = \binom{d+3}{4} = \Omega(d^4)$ points. Conversely, an argument based on Carathéodory's theorem shows that there exist exact 4-designs with $O(d^8)$ points. Starting from such an exact 4-design, the sparsification procedure from [23] gives a deterministic and efficient algorithm which outputs an ε-approximate 4-design supported by $O(d^4/\varepsilon^2)$ points.

However, this approach has two drawbacks: the constant C from (6.13) cannot be taken close to 1, and the number of outcomes has to be $\Omega(d^4)$. We are going to remove both inconveniences in our Theorem 6.5.3.

6.5.2 Euclidean subspaces

How do these ideas translate into the framework of zonoids? The analogue of $\mathrm{U_d}$ is the most symmetric zonoid, namely the Euclidean ball $B^n \subset \mathbf{R}^n$. To connect with literature from functional analysis, it is worth emphasizing that approximating B^n by a zonotope which is the sum of N segments is equivalent to embedding the space $\ell_2^n = (\mathbf{R}^n, \|\cdot\|_2)$ into the space $\ell_1^N = (\mathbf{R}^N, \|\cdot\|_1)$. Indeed, assume that $x_1, \ldots, x_N$ are points in $\mathbf{R}^n$ such that, for some constants c, C,

$$cZ \subset B^n \subset CZ,$$

where $Z = \mathrm{conv}\{\pm x_1\} + \cdots + \mathrm{conv}\{\pm x_N\}$. Then the map $u : \mathbf{R}^n \to \mathbf{R}^N$ defined by

$$u(x) = \Big(\langle x, x_1\rangle, \cdots, \langle x, x_N\rangle\Big)$$

satisfies $c\|u(x)\|_1 \leqslant \|x\|_2 \leqslant C\|u(x)\|_1$ for any $x \in \mathbf{R}^n$. In this context, the ratio C/c is often called the distortion of the embedding.

An early result by Rudin [156] shows an explicit embedding of ℓ_2^n into $\ell_1^{O(n^2)}$ with distortion $\sqrt{3}$. This is proved by the fourth moment method and can be seen as the analogue of the constructions based on 4-designs. The following theorem (a variation on Dvoretzky's theorem, recalled as Lemma 4.2.4 in Chapter 4, Section 4.2) has been a major improvement on Rudin's result, showing that ℓ_1^N has almost Euclidean sections of proportional dimension.

Theorem 6.5.2 ([77]). *For every $0 < \varepsilon < 1$, there exists a subspace $E \subset \mathbf{R}^N$ of dimension $n = c(\varepsilon)N$ such that for any $x \in E$,*

$$(1-\varepsilon)M\|x\|_2 \leqslant \|x\|_1 \leqslant (1+\varepsilon)M\|x\|_2, \tag{6.14}$$

where M denotes the average of the 1-norm over the Euclidean unit sphere S^{N-1}.

Theorem 6.5.2 was first proved in [77]. As explained in greater depth in Chapter 4, Section 4.2, the reasoning makes a seminal use of individual measure concentration in the form of Lévy's lemma (see Lemma 4.2.1) and a discretization via nets (see Lemma 4.2.3) to yield global concentration of measure of Dvoretzky-type (see Lemma 4.2.4). The argument shows that a generic subspace E (i.e. picked uniformly at random amongst all $c(\varepsilon)N$-dimensional subspaces of $\mathbf{R}^N$) satisfies the conclusion of the theorem with high probability whenever $c(\varepsilon) = O\left(\varepsilon^2 |\log \varepsilon|^{-1}\right)$. This was later improved in [83] to $c(\varepsilon) = O\left(\varepsilon^2\right)$.

6.5.3 Sparsification of the uniform POVM

Translated in the language of zonotopes, Theorem 6.5.2 states that the sum of $O(n)$ randomly chosen segments in $\mathbf{R}^n$ is close to the Euclidean ball B^n. More precisely, for any $0 < \varepsilon < 1$, if $N = c(\varepsilon)^{-1} n$ and $x_1, \ldots, x_N$ are randomly chosen points in $\mathbf{R}^n$, the zonotope $Z = \mathrm{conv}\{\pm x_1\} + \cdots + \mathrm{conv}\{\pm x_N\}$ is ε-close to the Euclidean ball B^n, in the sense that $(1-\varepsilon)Z \subset B^n \subset (1+\varepsilon)Z$.

By analogy, we expect a POVM constructed from $O(d^2)$ randomly chosen elements to be close to the uniform POVM. This random construction can be achieved as follows: let $(\psi_i)_{1 \leqslant i \leqslant n}$ be independent random vectors, uniformly chosen on the unit sphere of $\mathbf{C}^d$. Set $P_i = |\psi_i\rangle\langle\psi_i|$, $1 \leqslant i \leqslant n$, and $S = P_1 + \cdots + P_n$. When $n \geqslant d$, S is almost surely invertible, and we may consider the random POVM

$$\mathrm{M} = (S^{-1/2} P_i S^{-1/2})_{1 \leqslant i \leqslant n}. \tag{6.15}$$

Theorem 6.5.3. *Let* M *be a random POVM on* $\mathbf{C}^d$ *with* n *outcomes, defined as in* (6.15)*, and let* $0 < \varepsilon < 1$*. If* $n \geqslant C\varepsilon^{-2} |\log \varepsilon| d^2$*, then with high probability the POVM* M *satisfies the inequalities*

$$(1-\varepsilon)\|\Delta\|_{\mathrm{U}_d} \leqslant \|\Delta\|_{\mathrm{M}} \leqslant (1+\varepsilon)\|\Delta\|_{\mathrm{U}_d}$$

for every $\Delta \in \mathscr{H}(\mathbf{C}^d)$.

By "with high probability" we mean that the probability that the conclusion fails is less than $\exp(-c(\varepsilon)d)$ for some constant $c(\varepsilon)$. Theorem 6.5.3 is proved in Section 6.7, the proof being based on a careful use of ε-nets and deviation inequalities. It does not seem possible to deduce formally Theorem 6.5.3 from the existing Banach space literature.

Theorem 6.5.3 shows that the uniform POVM on $\mathbf{C}^d$ can be ε-approximated (in the sense of closeness of distinguishability norms) by a POVM with $n = O(\varepsilon^{-2} |\log \varepsilon| d^2)$ outcomes. Note that the dependence of n with respect to d is optimal: since a POVM on $\mathbf{C}^d$ must have at least d^2 outcomes to be informationally complete, one cannot hope for a tighter dimensional dependence. The dependence with respect to ε is less clear: the factor $|\log \varepsilon|$ can probably be removed but we do not pursue this direction.

Our construction is random and a natural question is whether deterministic constructions yielding comparable properties exist. A lot of effort has been put in derandomizing Theorem 6.5.2. We refer to [114] for bibliography and mention two of the latest results. Given any $0 < \gamma < 1$, it is shown in [114] how to construct, from cn^γ random bits (i.e. an amount of randomness sub-linear in n) a subspace of ℓ_1^N satisfying (6.14) with $N \leqslant (\gamma\varepsilon)^{-C\gamma} n$. A completely explicit construction appears in [113], with $N \leqslant n 2^{C(\varepsilon)(\log\log n)^2} = n^{1+C(\varepsilon)o(n)}$. It is not obvious how to adapt these constructions to obtain sparsifications of the uniform POVM using few or no randomness.

6.5.4 Sparsification of any POVM

Theorem 6.5.2 initiated intensive research in the late 80's [161, 32, 170] on the theme of "approximation of zonoids by zonotopes", trying to extend the result for the Euclidean ball (the most symmetric zonoid) to an arbitrary zonoid. This culminated in Talagrand's proof [170] that for any zonoid $Y \subset \mathbf{R}^n$ and any $0 < \varepsilon < 1$, there exists a zonotope $Z \subset \mathbf{R}^n$ which is the sum of $O(\varepsilon^{-2} n \log n)$ segments and such that $(1-\varepsilon)Y \subset Z \subset (1+\varepsilon)Y$. A more precise version is stated in Section 6.8. Whether the $\log n$ factor can be removed is still an open problem.

This result easily implies a similar result for POVMs, provided we consider the larger class of sub-POVMs. A discrete sub-POVM with n outcomes is a finite family $\mathrm{M} = (M_i)_{1 \leqslant i \leqslant n}$ of n positive operators such that $S = \sum_{i=1}^n M_i \leqslant \mathrm{Id}$. As for POVMs, the norm associated to a sub-POVM M is defined for $\Delta \in \mathscr{H}(\mathbf{C}^d)$ by

$$\|\Delta\|_{\mathrm{M}} = \sum_{i=1}^n |\operatorname{Tr}(\Delta M_i)|.$$

We prove the following result in Section 6.8.

Theorem 6.5.4. *Given any POVM* M *on* $\mathbf{C}^d$ *and any* $0 < \varepsilon < 1$*, there is a sub-POVM* $\mathrm{M}' = (M'_i)_{1 \leqslant i \leqslant n}$*, with* $n \leqslant C\varepsilon^{-2} d^2 \log(d)$ *such that, for any* $\Delta \in \mathscr{H}(\mathbf{C}^d)$*,*

$$(1-\varepsilon)\|\Delta\|_{\mathrm{M}} \leqslant \|\Delta\|_{\mathrm{M}'} \leqslant \|\Delta\|_{\mathrm{M}}.$$

Moreover, we can guarantee that the states $M'_i / \operatorname{Tr}(M'_i)$ *belong to the support of the measure* μ *associated to* M*.*

We do not know whether Theorem 6.5.4 still holds if we want M' to be a POVM. Given a sub-POVM $(M_i)_{1 \leqslant i \leqslant n}$, there are at least two natural ways to modify it into a POVM. A solution is to add an extra outcome corresponding to the operator $\mathrm{Id} - S$, and another one is to substitute $S^{-1/2} M_i S^{-1/2}$ in place of M_i, as we proceeded in (6.15). However for a general POVM, the error terms arising from this renormalization step may exceed the quantity to be approximated.

6.6 Sparsifying local POVMs

Proposition 6.6.1 below is an immediate corollary of Lemma 6.4.4 and Proposition 6.4.7. In words, it shows that, on a multipartite system, a local POVM can be sparsified by tensorizing sparsifications of each of its factors.

Proposition 6.6.1. *Let* $0 < \varepsilon < 1$*. Let* $\mathrm{M}_1, \ldots, \mathrm{M}_k$ *be POVMs and* $\mathrm{M}'_1, \ldots, \mathrm{M}'_k$ *be (sub-)POVMs, on* $\mathbf{C}^{d_1}, \ldots, \mathbf{C}^{d_k}$ *respectively, satisfying, for all* $1 \leqslant i \leqslant k$*, and for all* $\Delta \in \mathscr{H}(\mathbf{C}^{d_i})$*,*

$$(1-\varepsilon)\|\Delta\|_{\mathrm{M}_i} \leqslant \|\Delta\|_{\mathrm{M}'_i} \leqslant (1+\varepsilon)\|\Delta\|_{\mathrm{M}_i}.$$

Then, for any $\Delta \in \mathscr{H}(\mathbf{C}^{d_1} \otimes \cdots \otimes \mathbf{C}^{d_k})$*,*

$$(1-\varepsilon)^k \|\Delta\|_{\mathrm{M}_1 \otimes \cdots \otimes \mathrm{M}_k} \leqslant \|\Delta\|_{\mathrm{M}'_1 \otimes \cdots \otimes \mathrm{M}'_k} \leqslant (1+\varepsilon)^k \|\Delta\|_{\mathrm{M}_1 \otimes \cdots \otimes \mathrm{M}_k}.$$

Let us give a concrete application of Proposition 6.6.1. We consider k finite-dimensional Hilbert spaces $\mathbf{C}^{d_1}, \ldots, \mathbf{C}^{d_k}$ and define the local uniform POVM on the k-partite Hilbert space $\mathbf{C}^{d_1} \otimes \cdots \otimes \mathbf{C}^{d_k}$ as the tensor product of the k uniform POVMs $\mathrm{U}_{d_1}, \ldots, \mathrm{U}_{d_k}$. We will denote it by LU. The corresponding distinguishability norm can be described, for any $\Delta \in \mathscr{H}(\mathbf{C}^{d_1} \otimes \cdots \otimes \mathbf{C}^{d_k})$, as

$$\|\Delta\|_{\mathrm{LU}} = d\, \mathbf{E} \left| \langle \psi_1 \otimes \cdots \otimes \psi_k | \Delta | \psi_1 \otimes \cdots \otimes \psi_k \rangle \right|,$$

where $d = d_1 \times \cdots \times d_k$ is the dimension of the global Hilbert space, and where the random unit vectors $\psi_1, \ldots, \psi_k$ are independent and Haar-distributed in $\mathbf{C}^{d_1}, \ldots, \mathbf{C}^{d_k}$ respectively.

The following multipartite generalization of Proposition 6.5.1 shows that the norm $\|\cdot\|_{\mathrm{LU}}$, in analogy to the norm $\|\cdot\|_{\mathrm{U}}$, is equivalent to a "modified" Hilbert–Schmidt norm.

Proposition 6.6.2 ([130])**.** *For every* $\Delta \in \mathscr{H}(\mathbf{C}^{d_1} \otimes \cdots \otimes \mathbf{C}^{d_k})$*, we have*

$$\frac{1}{18^{k/2}} \|\Delta\|_{2(k)} \leqslant \|\Delta\|_{\mathrm{LU}} \leqslant \|\Delta\|_{2(k)}, \tag{6.16}$$

where the norm $\|\cdot\|_{2(k)}$ *is defined as*

$$\|\Delta\|_{2(k)} = \sqrt{\sum_{I \subset \{1,\ldots,k\}} \operatorname{Tr}\left[\left(\operatorname{Tr}_I \Delta \right)^2 \right]}. \tag{6.17}$$

Here Tr_I *denotes the partial trace over all parties* $I \subset \{1, \ldots, k\}$*.*

Proof of Proposition 6.6.2. A direct proof appears in [130], but we find interesting to show that it can be deduced (with a worst constant) from Proposition 6.5.1. If we denote by $\langle \cdot, \cdot \rangle_H$ the inner product inducing a Euclidean norm $\|\cdot\|_H$, we have

$$\langle A_1 \otimes \cdots \otimes A_k, B_1 \otimes \cdots \otimes B_k \rangle_{2(k)} = \langle A_1, B_1 \rangle_{2(1)} \times \cdots \times \langle A_k, B_k \rangle_{2(1)}$$

which is equivalent to saying that

$$\|\cdot\|_{2(k)} = \|\cdot\|_{2(1)} \otimes^2 \cdots \otimes^2 \|\cdot\|_{2(1)}.$$

We thus get by Proposition 6.4.5 that

$$c_0^{k-1}\|\cdot\|_{2(k)} \leqslant \|\cdot\|_{2(1)} \otimes^1 \cdots \otimes^1 \|\cdot\|_{2(1)} \leqslant \|\cdot\|_{2(k)}$$

with $c_0 = \sqrt{2/\pi}$. Now, we just have to observe that, we know by Proposition 6.4.7 that, on $\mathscr{H}(\mathbf{C}^{d_1} \otimes \cdots \otimes \mathbf{C}^{d_k})$, $\|\cdot\|_{\mathrm{LU}} = \|\cdot\|_{\mathrm{U}_{d_1}} \otimes^1 \cdots \otimes^1 \|\cdot\|_{\mathrm{U}_{d_k}}$, and by Proposition 6.5.1 that $c\|\cdot\|_{2(1)} \leqslant \|\cdot\|_{\mathrm{U}_d} \leqslant \|\cdot\|_{2(1)}$ for some constant c (e.g. $c = 1/\sqrt{18}$ works). So by Lemma 6.4.4,

$$c^k \|\cdot\|_{2(1)} \otimes^1 \cdots \otimes^1 \|\cdot\|_{2(1)} \leqslant \|\cdot\|_{\mathrm{LU}} \leqslant \|\cdot\|_{2(1)} \otimes^1 \cdots \otimes^1 \|\cdot\|_{2(1)},$$

and therefore

$$c_0^{k-1}c^k\|\cdot\|_{2(k)} \leqslant \|\cdot\|_{\mathrm{LU}} \leqslant \|\cdot\|_{2(k)}.$$ □

Remarkably, local dimensions do not appear in equation (6.16). This striking fact that local POVMs can have asymptotically non-vanishing distinguishing power can be used to construct an algorithm that solves the Weak Membership Problem for separability in quasi-polynomial time (see [34] for a description of the latter algorithm in the bipartite case, which is based on symmetric extension search, the central topic of Chapter 9). Hence the importance of being able to sparsify the local uniform POVM by a POVM for which the locality property is preserved and which has a number of outcomes that optimally scales as the square of the global dimension. We state the corresponding multipartite version of Theorem 6.5.3, which is straightforwardly obtained by combining the unipartite version with Proposition 6.6.1.

Theorem 6.6.3. *Let $0 < \varepsilon < 1$. For all $1 \leqslant i \leqslant k$, let M_i be a random POVM on $\mathbf{C}^{d_i}$ with $n_i \geqslant C\varepsilon^{-2}|\log\varepsilon|d_i^2$ outcomes, defined as in (6.15). Then, with high probability, the local POVM $\mathrm{M}_1 \otimes \cdots \otimes \mathrm{M}_k$ on $\mathbf{C}^{d_1} \otimes \cdots \otimes \mathbf{C}^{d_k}$ is such that, for any $\Delta \in \mathscr{H}(\mathbf{C}^{d_1} \otimes \cdots \otimes \mathbf{C}^{d_k})$,*

$$(1-\varepsilon)^k\|\Delta\|_{\mathrm{LU}} \leqslant \|\Delta\|_{\mathrm{M}_1\otimes\cdots\otimes\mathrm{M}_k} \leqslant (1+\varepsilon)^k\|\Delta\|_{\mathrm{LU}}.$$

Let us put in words the content of Theorem 6.6.3: the local uniform POVM on $\mathbf{C}^{d_1} \otimes \cdots \otimes \mathbf{C}^{d_k}$ can be $k\varepsilon$-approximated (in terms of distinguishability norms) by a POVM which is also local and has a total number of outcomes $n = O(C^k\varepsilon^{-2k}|\log\varepsilon|^k d^2)$, where $d = d_1 \times \cdots \times d_k$. Note that the dimensional dependence of n is optimal. On the contrary, the dependence of n on ε deteriorates as k grows. The high-dimensional situation our result applies to is thus really the one of a "small" number of "large" subsystems (i.e. k fixed and $d_1, \ldots, d_k \to +\infty$), and not of a "large" number of "small" subsystems.

6.7 Proof of the main theorem concerning the sparsification of the uniform POVM

6.7.1 Proof of Theorem 6.5.3

In this subsection we prove Theorem 6.5.3. Let $n \in \mathbf{N}$ and $(\psi_i)_{1\leqslant i\leqslant n}$ be independent random unit vectors, uniformly distributed on the unit sphere of $\mathbf{C}^d$. Our main technical estimates are a couple of probabilistic inequalities. Proposition 6.7.1 is an immediate consequence of Theorem 1 in [9]. Proposition 6.7.2 is a consequence of a Bernstein-type inequality, recalled as Theorem 4.2.5 in Chapter 4, Section 4.2. However, its proof requires some careful estimates which we postpone to Subsection 6.7.2.

Proposition 6.7.1. *If $(\psi_i)_{1\leqslant i\leqslant n}$ are independent random vectors, uniformly distributed on the unit sphere of $\mathbf{C}^d$, then for every $0 < \eta < 1$*

$$\mathbf{P}\left((1-\eta)\frac{\mathrm{Id}}{d} \leqslant \frac{1}{n}\sum_{i=1}^n |\psi_i\rangle\langle\psi_i| \leqslant (1+\eta)\frac{\mathrm{Id}}{d}\right) \geqslant 1 - C^d\exp(-cn\eta^2).$$

Proposition 6.7.2. *Fix $\Delta \in \mathscr{H}(\mathbf{C}^d)$, and let $(\psi_i)_{1\leqslant i\leqslant n}$ be independent random vectors, uniformly distributed on the unit sphere of $\mathbf{C}^d$. For each $1 \leqslant i \leqslant n$, consider next the random variables $X_i = d|\langle\psi_i|\Delta|\psi_i\rangle|$ and $Y_i = X_i - \mathbf{E}\,X_i = X_i - \|\Delta\|_{\mathrm{U}_d}$. Then, for any $t > 0$,*

$$\mathbf{P}\left(\left|\frac{1}{n}\sum_{i=1}^n Y_i\right| \geqslant t\|\Delta\|_{\mathrm{U}_d}\right) \leqslant 2\exp(-c_0'n\min(t,t^2)).$$

We now show how to derive Theorem 6.5.3 from the estimates in Propositions 6.7.1 and 6.7.2. For each $1 \leqslant i \leqslant n$, set $P_i = |\psi_i\rangle\langle\psi_i|$, and introduce the (random) norm defined for any $\Delta \in \mathscr{H}(\mathbf{C}^d)$ as

$$|||\Delta||| = \frac{d}{n}\sum_{i=1}^{n} |\operatorname{Tr}(\Delta P_i)|.$$

We will now prove that $|||\cdot|||$ is, with probability close to 1, a good approximation to $\|\cdot\|_{\mathrm{U}_d}$. First, using Proposition 6.7.2, we obtain that for any $0 < \varepsilon < 1$ and any $\Delta \in \mathscr{H}(\mathbf{C}^d)$

$$\mathbf{P}\left((1-\varepsilon)\|\Delta\|_{\mathrm{U}_d} \leqslant |||\Delta||| \leqslant (1+\varepsilon)\|\Delta\|_{\mathrm{U}_d}\right) \geqslant 1 - 2\exp(-c_0' n\varepsilon^2). \tag{6.18}$$

We next use a net argument. Fix $0 < \varepsilon < 1/3$ and a ε-net $\mathscr{A}$ inside the unit ball for the norm $\|\cdot\|_{\mathrm{U}_d}$, with respect to the distance induced by $\|\cdot\|_{\mathrm{U}_d}$. A standard volumetric argument (see Lemma 4.2.3 in Chapter 4, Section 4.2) shows that we may assume $|\mathscr{A}| \leqslant (1+2/\varepsilon)^{d^2} \leqslant (3/\varepsilon)^{d^2}$. Introduce the quantities

$$A := \sup\{|||\Delta||| \ : \ \|\Delta\|_{\mathrm{U}_d} \leqslant 1\},$$

$$A' := \sup\{|||\Delta||| \ : \ \Delta \in \mathscr{A}\}.$$

Given Δ such that $\|\Delta\|_{\mathrm{U}_d} \leqslant 1$, there is $\Delta_0 \in \mathscr{A}$ with $\|\Delta - \Delta_0\|_{\mathrm{U}_d} \leqslant \varepsilon$. By the triangle inequality, we have $|||\Delta||| \leqslant A' + |||\Delta - \Delta_0||| \leqslant A' + \varepsilon A$. Taking supremum over Δ yields $A \leqslant A' + \varepsilon A$ i.e. $A \leqslant A'/(1-\varepsilon)$.

If we introduce $B := \inf\{|||\Delta||| \ : \ \|\Delta\|_{\mathrm{U}_d} = 1\}$ and $B' := \inf\{|||\Delta||| \ : \ \Delta \in \mathscr{A}\}$, a similar argument shows that $B \geqslant B' - \varepsilon A$, so that in fact $B \geqslant B' - \varepsilon A'/(1-\varepsilon)$. We therefore have the implications

$$1-\varepsilon \leqslant B' \leqslant A' \leqslant 1+\varepsilon \quad \Rightarrow \quad 1 - \varepsilon - \frac{\varepsilon(1+\varepsilon)}{1-\varepsilon} \leqslant B \leqslant A \leqslant \frac{1+\varepsilon}{1-\varepsilon} \quad \Rightarrow \quad 1 - 3\varepsilon \leqslant B \leqslant A \leqslant 1 + 3\varepsilon. \tag{6.19}$$

By the union bound, we get from (6.18) that $\mathbf{P}(1-\varepsilon \leqslant B' \leqslant A' \leqslant 1+\varepsilon) \geqslant 1 - 2|\mathscr{A}|\exp(-c_0' n\varepsilon^2)$. Combined with (6.19), and using homogeneity of norms, this yields

$$\mathbf{P}\Big((1-3\varepsilon)\|\cdot\|_{\mathrm{U}_d} \leqslant |||\cdot||| \leqslant (1+3\varepsilon)\|\cdot\|_{\mathrm{U}_d}\Big) \geqslant 1 - 2\left(\frac{3}{\varepsilon}\right)^{d^2}\exp(-c_0' n\varepsilon^2). \tag{6.20}$$

This probability estimate is non-trivial, and can be made close to 1, provided $n \gtrsim d^2\varepsilon^{-2}|\log\varepsilon|$.

Whenever $n \geqslant d$, the vectors $(\psi_i)_{1\leqslant i\leqslant n}$ generically span $\mathbf{C}^d$, and therefore the operator $S = P_1 + \cdots + P_n$ is invertible. We may then define $\widetilde{P}_i = S^{-1/2} P_i S^{-1/2}$ so that $\mathrm{M} = (\widetilde{P}_i)_{1\leqslant i\leqslant n}$ is a POVM. The norm associated to M is, for any $\Delta \in \mathscr{H}(\mathbf{C}^d)$,

$$\|\Delta\|_{\mathrm{M}} = \sum_{i=1}^{n} |\operatorname{Tr}(\Delta\widetilde{P}_i)|.$$

We now argue that the norms $|||\cdot|||$ and $\|\cdot\|_{\mathrm{M}}$ are similar enough (modulo normalization), because the modified operators $\widetilde{P}_i$ are close enough to the initial ones P_i. This is achieved by showing that $T := \sqrt{n/d}\, S^{-1/2}$ is close to Id (in operator-norm distance). We use Proposition 6.7.1 for $\eta = \varepsilon\|\Delta\|_{\mathrm{U}_d}/\|\Delta\|_1$. By Proposition 6.5.1, we have $\eta \geqslant \varepsilon/\sqrt{18d}$. Proposition 6.7.1 implies that

$$\mathbf{P}(\|T - \mathrm{Id}\|_\infty \geqslant \eta) \leqslant \mathbf{P}(\|T^{-2} - \mathrm{Id}\|_\infty \geqslant \eta) \leqslant C^d \exp(-c' n\varepsilon^2/d). \tag{6.21}$$

This upper bound is much smaller than 1 provided $n \geqslant C_1\varepsilon^{-2}d^2$. Also, note that the event $\|T - \mathrm{Id}\|_\infty \leqslant \eta$ implies that

$$\|\Delta - T\Delta T\|_{\mathrm{M}} \leqslant \|\Delta - T\Delta T\|_1 \leqslant \|\Delta\|_1\|\mathrm{Id} - T\|_\infty\,(1 + \|T\|_\infty) \leqslant 2\eta\|\Delta\|_1 = 2\varepsilon\|\Delta\|_{\mathrm{U}_d}.$$

Using the cyclic property of the trace, we check that $\|T\Delta T\|_{\mathrm{M}} = |||\Delta|||$. Now, choose n larger than both $C_0\varepsilon^{-2}|\log\varepsilon|d^2$ and $C_1\varepsilon^{-2}d^2$. With high probability, the events from equations (6.20) and (6.21) both hold. We then obtain for every $\Delta \in \mathscr{H}(\mathbf{C}^d)$,

$$\|\Delta\|_{\mathrm{M}} \leqslant \|T\Delta T\|_{\mathrm{M}} + \|\Delta - T\Delta T\|_{\mathrm{M}} \leqslant |||\Delta||| + 2\varepsilon\|\Delta\|_{\mathrm{U}_d} \leqslant (1+5\varepsilon)\|\Delta\|_{\mathrm{U}_d}$$

and similarly $\|\Delta\|_{\mathrm{M}} \geqslant (1-5\varepsilon)\|\Delta\|_{\mathrm{U}_d}$. This is precisely the result from Theorem 6.5.3 with 5ε instead of ε, which of course can be absorbed by renaming the constants appropriately.

6.7.2 Proof of Proposition 6.7.2

The proof is a direct application of a large deviation inequality for sums of independent ψ_1 (aka sub-exponential) random variables. More details on that topic are gathered in Chapter 4, Section 4.2.

For $\Delta \in \mathscr{H}(\mathbf{C}^d)$, consider for each $1 \leqslant i \leqslant n$ the random variables $X_i = d|\operatorname{Tr}(\Delta P_i)|$ with $P_i = |\psi_i\rangle\langle\psi_i|$, and $Y_i = X_i - \mathbf{E}\, X_i = d|\operatorname{Tr}(\Delta P_i)| - \|\Delta\|_{\mathrm{U}_d}$. The random variables Y_i, $1 \leqslant i \leqslant n$, are independent and have mean zero. The key lemma is a bound on their ψ_1-norm.

Lemma 6.7.3. *Let $\Delta \in \mathscr{H}(\mathbf{C}^d)$ and consider the random variable $X := d|\operatorname{Tr}(\Delta P)|$, where $P = |\psi\rangle\langle\psi|$ with ψ uniformly distributed on the unit sphere of $\mathbf{C}^d$. Then, first of all $\|X\|_{\psi_1} \leqslant \|\Delta\|_{2(1)}$, and as a consequence $\|X - \mathbf{E}\, X\|_{\psi_1} \leqslant 3\|\Delta\|_{2(1)} \leqslant 3\sqrt{18}\|\Delta\|_{\mathrm{U}_d}$.*

Therefore, we may apply Bernstein's inequality, recalled as Theorem 4.2.5 in Chapter 4, Section 4.2, with $M = \sigma \leqslant 3\sqrt{18}\|\Delta\|_{\mathrm{U}_d}$, yielding Proposition 6.7.2.

Proof of Lemma 6.7.3. For each integer q, we compute

$$\mathbf{E}\,[\operatorname{Tr}(\Delta P)]^{2q} = \mathbf{E}\operatorname{Tr}\left(\Delta^{\otimes 2q} P^{\otimes 2q}\right) = \operatorname{Tr}\left(\Delta^{\otimes 2q}\left[\mathbf{E}\, P^{\otimes 2q}\right]\right).$$

In order to go further, we simply observe that

$$\mathbf{E}\, P^{\otimes 2q} = \frac{(2q)!}{(d+2q-1)\times\cdots\times d} P_{\mathrm{Sym}^{2q}(\mathbf{C}^d)} = \frac{1}{(d+2q-1)\times\cdots\times d}\sum_{\pi\in\mathfrak{S}(2q)} U(\pi),$$

where $P_{\mathrm{Sym}^{2q}(\mathbf{C}^d)}$ denotes the orthogonal projection onto the symmetric subspace $\mathrm{Sym}^{2q}(\mathbf{C}^d) \subset (\mathbf{C}^d)^{\otimes 2q}$, and for each permutation $\pi \in \mathfrak{S}(2q)$, $U(\pi)$ denotes the associated permutation unitary on $(\mathbf{C}^d)^{\otimes 2q}$ (see e.g. [93] and Chapter 3, Section 3.1, of this manuscript for much more on that matter). This yields

$$\mathbf{E}\,[\operatorname{Tr}(\Delta P)]^{2q} = \frac{1}{(d+2q-1)\times\cdots\times d}\sum_{\pi\in\mathfrak{S}(2q)} \operatorname{Tr}\left(\Delta^{\otimes 2q} U(\pi)\right).$$

If $\ell_1, \ldots, \ell_k$ denote the lengths of the cycles appearing in the cycle decomposition of a permutation $\pi \in \mathfrak{S}(2q)$, we have $\ell_1 + \cdots + \ell_k = 2q$ and

$$\operatorname{Tr}\left(\Delta^{\otimes 2q} U(\pi)\right) = \prod_{i=1}^{k} \operatorname{Tr}(\Delta^{\ell_i}).$$

Now, for any integer $\ell \geqslant 2$, we have $|\operatorname{Tr}(\Delta^\ell)| \leqslant [\operatorname{Tr}(\Delta^2)]^{\ell/2} \leqslant \|\Delta\|_{2(1)}^{\ell}$. The inequality $|\operatorname{Tr}(\Delta^\ell)| \leqslant \|\Delta\|_{2(1)}^{\ell}$ is also (trivially) true for $\ell = 1$. Therefore $\left|\operatorname{Tr}\left(\Delta^{\otimes 2q} U(\pi)\right)\right| \leqslant \|\Delta\|_{2(1)}^{2q}$. It follows that

$$\mathbf{E}\,[\operatorname{Tr}(\Delta P)]^{2q} \leqslant \frac{(2q)!}{d^{2q}}\|\Delta\|_{2(1)}^{2q} \leqslant \left(\frac{2q\|\Delta\|_{2(1)}}{d}\right)^{2q},$$

so that $\left(\mathbf{E}\, X^{2q}\right)^{1/2q} \leqslant 2q\|\Delta\|_{2(1)}$, and thus $\|X\|_{\psi_1} \leqslant \|\Delta\|_{2(1)}$. The last part of the Lemma follows from the triangle inequality, since $\|\mathbf{E}\, X\|_{\psi_1} = |\mathbf{E}\, X| \leqslant 2\|X\|_{\psi_1}$, and from the equivalence (6.11) between the norms $\|\cdot\|_{\mathrm{U}_d}$ and $\|\cdot\|_{2(1)}$. □

6.8 Proof of the main theorem concerning the sparsification of any POVM

In this section we prove Theorem 6.5.4. Here is a version of Talagrand's theorem which is suitable for our purposes.

Theorem 6.8.1 ([170])**.** *Let $Z \subset \mathbf{R}^n$ be a symmetric zonotope, with*

$$Z = \sum_{i\in I} \operatorname{conv}\{\pm u_i\}$$

for a finite family of vectors $(u_i)_{i\in I}$. Then for every $\varepsilon > 0$ there exists a subset $J \subset I$ with $|J| \leqslant Cn\log n/\varepsilon^2$, and positive numbers $(\lambda_i)_{i\in J}$ such that the zonotope

$$Z' = \sum_{i\in J} \operatorname{conv}\{\pm\lambda_i u_i\}$$

satisfies $Z' \subset Z \subset (1+\varepsilon)Z'$.

Theorem 6.5.4 is a very simple consequence of Theorem 6.8.1. Let M be a POVM to be sparsified. Using Corollary 6.3.2, we may assume that $\mathrm{M} = (M_i)_{i\in I}$ is discrete. Applying Theorem 6.8.1 to the zonotope $K_{\mathrm{M}} = \sum_{i\in I} \operatorname{conv}\{\pm M_i\}$ (which lives in a d^2-dimensional space), we obtain a zonotope $Z' = \sum_{i\in J} \operatorname{conv}\{\pm\lambda_i M_i\}$ with $|J| \leqslant Cd^2 \log d/\varepsilon^2$ such that $Z' \subset K_{\mathrm{M}} \subset (1+\varepsilon)Z'$. It remains to show that $\mathrm{M}' = (\lambda_i M_i)_{i\in J}$ is a sub-POVM. We know that $h_{Z'} \leqslant h_{K_{\mathrm{M}}}$. Therefore, given a unit vector $x \in \mathbf{C}^d$, the inequality $h_{Z'}(\Delta) \leqslant h_{K_{\mathrm{M}}}(\Delta)$ applied with $\Delta = |x\rangle\langle x|$ shows that

$$\sum_{i\in J} \lambda_i \, |\langle x|M_i|x\rangle| \leqslant \||x\rangle\langle x|\|_{\mathrm{M}} \leqslant \||x\rangle\langle x|\|_1 = 1,$$

and therefore $\sum_{i\in J} \lambda_i M_i \leqslant \mathrm{Id}$, as required. Since the inclusions $Z' \subset K_{\mathrm{M}} \subset (1+\varepsilon)Z'$ are equivalent to the inequalities $\|\cdot\|_{\mathrm{M}'} \leqslant \|\cdot\|_{\mathrm{M}} \leqslant (1+\varepsilon)\|\cdot\|_{\mathrm{M}'}$, Theorem 6.5.4 follows.

Part III

Some aspects of generic entanglement: data-hiding and relaxations of separability in high dimensions

When talking about entanglement in multipartite quantum systems, the picture changes drastically depending on the system's size. In small dimensions, relaxing the notion of separability to one which is easier to handle is usually a quite fruitful approach. Oppositely, as the dimensions grow, any too simple necessary condition for separability is doomed to be very rough. These very general and hand-waving assertions actually apply to many more specific settings and can be made much more precise by using tools from high-dimensional convex geometry. Such line of study was arguably initiated by Hayden, Leung and Winter in [100]. In this seminal paper, the typical value of various correlation measures was estimated for random high-dimensional multipartite states. Among others, it was already observed there that, in bipartite quantum systems, features such as large entanglement of formation and small distillable entanglement are the rule rather than the exception when the dimensions of the two subsystems are large. It may have appeared surprising at first that having such bound-entangled like properties is in fact a common trait. Indeed, exhibiting explicit examples of states being so is usually hard, because they may be rare (or even nonexisting) in small dimensions while computations rapidly become intractable as dimensions grow.

Chapter 7 is in keeping with Chapter 6 in the previous part, where the functional-analytical study of distinguishability norms associated to POVMs was initiated. And actually, the first question we look at could just as well have been put in Part II, since it consists in looking at finite sub-families of the infinite family **ALL** of all POVMs, and asking: how many POVMs do they have to contain to achieve near to the maximum discrimination efficiency attained by **ALL**? However, the focus turns next to multipartite systems, where POVMs satisfying certain locality constraints are considered. The goal is then to try and quantify how such restrictions might affect the ability to distinguish global states. More concretely, we look at several classes of locally restricted POVMs on multipartite quantum systems, and ask the following: as the dimensions of the underlying local spaces increase, does there exist an unbounded gap between them, and if so, is it an exceptional or a typical feature? Let us briefly summarize the obtained results, concentrating on the bipartite case for the sake of clarity. In the hierarchy between the families of local POVMs **LO**, local POVMs with one-way classical communication $\mathbf{LOCC}^{\rightarrow}$, local POVMs with two-way classical communication **LOCC**, separable POVMs **SEP**, positive under partial transposition POVMs **PPT** and all POVMs **ALL**, there exist unbounded gaps between the corresponding distinguishability norms at each step. Furthermore, this unbounded gap is generic in **SEP** vs **PPT**, but not generic in $\mathbf{LOCC}^{\rightarrow}$ vs **LOCC**, **LOCC** vs **SEP**, and **PPT** vs **ALL** (concerning the unbounded gap **LO** vs $\mathbf{LOCC}^{\rightarrow}$, we are unable to conclude about its typicality). These results translate nicely in terms of data-hiding considerations: this phenomenon, consisting of two multipartite states which are very different (hence very well distinguishable by some global measurement) but nevertheless look almost the same to observers that can only perform measurements on their subsystem and communicate classically, is in fact exhibited by most multipartite states. Let us also say just one word concerning the main ingredients in the proofs. In order to estimate what is the value of $\|\cdot\|_{\mathbf{M}}$ to be expected, for **M** being one of the above-mentioned families of POVMs, two steps are required: one must first determine the average value of $\|\cdot\|_{\mathbf{M}}$ by estimating certain size parameters of (the polar of) its unit ball, and second make use of random matrix theory and concentration of measure to argue that this average behaviour occurs with overwhelming probability in high dimension.

Chapter 8 is dedicated to the study of separability and entanglement in multipartite quantum systems. Certifying that a bipartite state is not separable can always be done by exhibiting an entanglement witness constructed from a positive map. In the multipartite case, the picture becomes more intricate. Indeed, even asserting that a state is not biseparable (i.e. a convex combination of states which are separable across a given bipartition) may be a delicate task. We show however that it is always possible to construct such a genuine multipartite entanglement witness by lifting entanglement witnesses which only reveal bipartite entanglement. In small dimensions, this approach is quite versatile since it allows for a formulation of the problem as a semidefinite program, and can therefore be solved efficiently. Nevertheless, as one could have expected, any state-independent construction is condemned to become weaker and weaker as the dimensions grow. We back up this affirmation by focussing on one specific positive map relaxation of separability, namely positivity under partial transposition. We thus prove that, on high dimensional multipartite systems, the set of states which have a positive partial transpose across every cut is much bigger than the set of biseparable states. We additionally construct a whole class of random multipartite states which have the property of being, with overwhelming probability as the dimensions increase, fully positive under partial transposition and nevertheless (robustly) genuinely multipartite entangled. These two arguments substantiate that our universal schemes, even though efficient and effective to detect genuine multipartite entanglement in, for instance, three-qutrit systems, will miss most genuinely multipartite entangled states on higher dimensional systems.

Finally, in Chapter 9, we take a closer look at one particular hierarchy of separability tests, known as the

hierarchy of k-extendibility tests, indexed by $k \in \mathbf{N}$. It consists of a sequence of increasingly constraining necessary conditions for separability (expressible as semidefinite programs of increasing size) which is asymptotically also sufficient. In real life however, only a finite number of checks can be performed, so it makes sense to ask, for a fixed $k \in \mathbf{N}$, what is the strength of the kth test. Now, it is known that there exist states which are far from separable (in either standard or operational distance measures) even though highly extendible. But these worst case scenarios do not exclude the possibility of making stronger statements about average or typical behaviors. This is precisely the question we tackle in Chapter 9, being especially interested in the high-dimensional regime. There are at least two distinct ways of answering it in a quantitative manner. The first approach consists in estimating a certain size parameter of the set of k-extendible states, and compare the obtained value with the known corresponding estimate for the set of separable states to see how the sizes of these two sets of states scale with one another. The second approach consists in looking at *random-induced states* (i.e. random mixed states which are obtained by partial tracing over an ancilla space a uniformly distributed pure state) and characterizing when these are with high probability k-extendible or not, so that again, comparing the obtained result with the known one for separability provides some information on how powerful the k-extendibility test typically is to detect entanglement. These two routes lead to the same conclusion, namely that if $k \in \mathbf{N}$ is a fixed parameter, then k-extendibility becomes a very loose relaxation of separability as the dimensions of the underlying local spaces grow. Nevertheless, whatever other well-studied separability criterion is defeated by the k-extendibility criterion above a certain dimension independent value of k (from either one or the other point of view). Furthermore, it is possible to partially extend these results to the case were k is not fixed but instead grows with the local dimensions as well. We are thus able to see that, for some growth rate of k, the set of k-extendible states lies strictly in-between the set of separable states and the set of all states.

Similarly in Chapters 7, 8 and 9, the tools that we use to estimate the sizes of the convex bodies under consideration (either sets of POVMs or sets of states) are both geometric and probabilistic. Indeed, there are two main size parameters that we consider, the *volume radius* and the *mean width*. The volume radius, as the name indicates, is a volumetric quantity. Understanding how it behaves under intersection, Minkowski sum, symmetrization, tensor product etc. is the very essence of classical convex geometry. The mean width on the other hand, again in line with the denomination, is a probabilistic quantity: estimating it can be phrased as estimating the average of the supremum of a certain Gaussian process. Powerful union-bound type arguments exist for that (yielding sharp bounds at their highest level of sophistication and, for our purposes, already pretty good ones in their simpler forms). Then, again in all three chapters, making statements about typical behaviours for random states requires, in addition to a mean value estimate, one extra ingredient, namely concentration of Lipschitz functions (in spherical or Gaussian random variables) around their average in high dimension. Note that in order to get, as a starting point, the average case behaviour, there are basically two routes that we may follow: either we relate it to some (already computed) mean width (appealing to the heuristic that unitary invariant random matrix ensembles having comparable spectra have comparable norms), or we compute it directly using standard tools from random matrix theory (e.g. moment method).

Part III – Table of contents

Chapter 7 Locally restricted measurements on multipartite quantum systems 81

7.1 Introduction 81
7.2 Distinguishing quantum states: survey of our results 83
7.3 On the complexity of the class of all POVMs 86
7.4 Existing unbounded gap between **LO** and **LOCC** 88
7.5 Generic unbounded gap between **SEP** and **PPT** 92
7.6 Applications to quantum data hiding 97
7.7 Miscellaneous remarks and questions 99
7.8 Appendix: Volume estimates for some Schatten norm unit balls and related convex bodies . . 100

Chapter 8 Relaxations of separability in multipartite quantum systems 103

8.1 Introduction 103
8.2 Characterizing relaxations of separability with semidefinite programs 104
8.3 Constructing multiparticle witnesses from bipartite witnesses 105
8.4 Relaxations of separability beyond positive maps 108
8.5 Estimating the performance of PPT relaxations in high dimensions 109
8.6 Conclusion 114

Chapter 9 k-extendibility of high-dimensional bipartite quantum states 117

9.1 Introduction 117
9.2 Mean width of the set of k-extendible states for "small" k 120
9.3 Discussion and comparison with the mean width of the set of PPT states 124
9.4 Adding the PPT constraint on the extension 125
9.5 Preliminaries on random-induced states and witnesses 126
9.6 Non k-extendibility of random-induced states for "small" k 128
9.7 Discussion and comparison with other separability criteria 132
9.8 The unbalanced case 132
9.9 Miscellaneous questions 134
9.10 Appendix A: Combinatorics of permutations and partitions: short summary of standard facts 137
9.11 Appendix B: Computing moments of random matrices: Wick formula and genus expansion . . 138
9.12 Appendix C: A needed combinatorial fact: relating the number of cycles in some specific permutations on either $[p] \times [k]$ or $[p]$ 139
9.13 Appendix D: Moments of "modified" GUE matrices (proof) 141
9.14 Appendix E: Moments of partially transposed "modified" GUE matrices (proof) 142
9.15 Appendix F: Moments of "modified" Wishart matrices (proof) 142
9.16 Appendix G: Counting geodesics vs non-geodesics pairings and permutations 144
9.17 Appendix H: Extra remarks on the convergence of the studied random matrix ensembles . . . 149

Chapter 7

Locally restricted measurements on multipartite quantum systems

Based on "Locally restricted measurements on a multipartite quantum system: data hiding is generic", in collaboration with G. Aubrun [11].

We study the distinguishability norms associated to families of locally restricted POVMs on multipartite systems. These norms (introduced by Matthews, Wehner and Winter) quantify how quantum measurements, subject to locality constraints, perform in the task of discriminating two multipartite quantum states. We mainly address the following question regarding the behaviour of these distinguishability norms in the high-dimensional regime: On a bipartite space, what are the relative strengths of standard classes of locally restricted measurements? We show that the class of PPT measurements typically performs almost as well as the class of all measurements whereas restricting to local measurements and classical communication, or even just to separable measurements, implies a substantial loss. We also provide examples of state pairs which can be perfectly distinguished by local measurements if (one-way) classical communication is allowed between the parties, but very poorly without it. Finally, we study how many POVMs are needed to distinguish almost perfectly any pair of states on $\mathbf{C}^d$, showing that the answer is $\exp(\Theta(d^2))$.

7.1 Introduction

How quantum measurements can help us make decisions? We consider a basic problem, the task of distinguishing two quantum states, where this question has a neat answer. Given a POVM (Positive Operator-Valued Measure) M on $\mathbf{C}^d$, Matthews, Wehner and Winter [137] introduced its distinguishability norm $\|\cdot\|_{\mathrm{M}}$, which has the property that given a pair (ρ, σ) of quantum states, $\|\rho - \sigma\|_{\mathrm{M}}$ is the bias observed when the POVM M is used optimally to distinguish ρ from σ (the larger is the norm, the more efficient is the POVM). More generally, we can associate to a family of POVMs $\mathbf{M}$ the norm $\|\cdot\|_{\mathbf{M}} = \sup\{\|\cdot\|_{\mathrm{M}} \ : \ \mathrm{M} \in \mathbf{M}\}$ which corresponds to the bias achieved by the best POVM from the family.

Here, in the line of Chapter 6, we study these norms from a functional-analytic point of view and are mostly interested in the asymptotic regime, when the dimension of the underlying Hilbert space tends to infinity.

How many essentially distinct POVMs are there?

The (infinite) family **ALL** of all POVMs on $\mathbf{C}^d$ achieves maximal efficiency in the distinguishability task, and in some sense gives us perfect information. It was indeed one of the seminal observations by Holevo [105] and Helstrom [102] that $\|\cdot\|_{\mathbf{ALL}} = \|\cdot\|_1$, so that two orthogonal quantum states could be perfectly distinguished (i.e. with a zero probability of error) by a suitable measurement. But how "complex" is the class **ALL**? What about finite subfamilies? How many POVMs are needed to obtain near-to-optimal efficiency? We show (Theorem 7.2.1) that $\exp(\Theta(d^2))$ different POVMs are necessary (and sufficient) to obtain approximation within a constant factor. The concept of mean width (from convex geometry) plays an important role in our proof, which is detailed in Section 7.3.

Locally restricted POVMs on a multipartite quantum system

On a multipartite quantum system, experimenters usually cannot implement any global observable. For instance, they may be only able to perform quantum measurements on their own subsystem (and then perhaps to communicate the results classically). A natural question in such situation is thus to quantify the relative strengths of several classes of measurements, restricted by these locality constraints, such as LOCC, separable or PPT measurements (precise definitions appear in Section 7.2.3).

Let us summarize the main result in this chapter (restricting here to the bipartite case for the sake of clarity). We consider typical discrimination tasks, in the following sense. Let ρ and σ be states chosen independently and uniformly at random within the set of all states on $\mathbf{C}^d \otimes \mathbf{C}^d$. We show that our ability to distinguish ρ from σ depends in an essential way on the class of the allowed measurements. Indeed, with high probability, $\|\rho-\sigma\|_{\mathbf{PPT}}$ is of order 1 (as $\|\rho-\sigma\|_{\mathbf{ALL}}$) while $\|\rho-\sigma\|_{\mathbf{SEP}}$, $\|\rho-\sigma\|_{\mathbf{LOCC}}$ and $\|\rho-\sigma\|_{\mathbf{LOCC}^{\rightarrow}}$ are of order $1/\sqrt{d}$. This shows that data hiding is generic: typically, high-dimensional quantum states cannot be distinguished locally even though they look different globally. These results appear as Theorem 7.2.2 in Section 7.2.4. The proofs are detailed in Section 7.5. They rely, as a first essential step, on estimates on the volume radius and the mean width of the (polar of) the unit balls associated to the norms $\|\cdot\|_{\mathbf{PPT}}$, $\|\cdot\|_{\mathbf{SEP}}$ and $\|\cdot\|_{\mathbf{LOCC}}$ (Theorem 7.5.1). The use of concentration of measure and random matrix theory (Proposition 7.5.3) then allows to pass from these global estimates to the estimates in a typical direction quoted above. In Section 7.6 corollaries on quantum data hiding are derived and detailed, both in the bipartite and in the generalized multipartite case.

We also provide examples of random bipartite states ρ, σ on $\mathbf{C}^d \otimes \mathbf{C}^d$ which are such that $\|\rho-\sigma\|_{\mathbf{LOCC}^{\rightarrow}} = 2$ while, with high probability, $\|\rho - \sigma\|_{\mathbf{LO}}$ is of order $1/\sqrt{d}$. The precise result appears in Theorem 7.2.4 and is proved in Section 7.4. Following the same proof technique, one can then construct, more generally, random bipartite states ρ, σ such that, either $\|\rho - \sigma\|_{\mathbf{LOCC}^{\rightarrow}} \gg \|\rho - \sigma\|_{\mathbf{LO}}$ or $\|\rho - \sigma\|_{\mathbf{LOCC}} \gg \|\rho - \sigma\|_{\mathbf{LOCC}^{\rightarrow}}$.

Table 7.1 summarizes our various conclusions.

Table 7.1: Unbounded gaps between locally restricted distinguishability norms

Norm hierarchy	$\|\cdot\|_{\mathbf{LO}}$	$\leqslant$	$\|\cdot\|_{\mathbf{LOCC}^{\rightarrow}}$	$\leqslant$	$\|\cdot\|_{\mathbf{LOCC}}$	$\leqslant$	$\|\cdot\|_{\mathbf{SEP}}$	$\leqslant$	$\|\cdot\|_{\mathbf{PPT}}$	$\leqslant$	$\|\cdot\|_{\mathbf{ALL}}$
Existing unbounded gap?			yes		yes		?		yes		yes
Generic unbounded gap?			?		no		no		yes		no

Notation

In addition to the general notation specified in Chapter 1, Section 1.3, we introduce the following one: When A, B are Hermitian matrices, we denote by $[A, B]$ the order interval, i.e. the set of Hermitian matrices C such that both $C - A$ and $B - C$ are positive semidefinite matrices. In particular, $[-\mathrm{Id}, \mathrm{Id}]$ is the Hermitian part of the unit ball for $\|\cdot\|_\infty$ on the set of all complex matrices.

When A and B are quantities depending on the dimension, the notation $A \lesssim B$ means that there is a constant C such that $A \leqslant CB$. The notation $A \simeq B$ means both $A \lesssim B$ and $B \lesssim A$, and $A \sim B$ means that the ratio A/B tends to 1 when the dimension tends to infinity.

Extra notation, concepts and results from convex geometry, which are needed throughout our proofs, are gathered in Chapter 4, Section 4.1.

7.2 Distinguishing quantum states: survey of our results

7.2.1 General setting

In this section, we gather some basic information about norms associated to POVMs, and refer e.g. to [137] for more details and proofs. A POVM (Positive Operator-Valued Measure) on $\mathbf{C}^d$ is a finite family $\mathrm{M} = (M_i)_{i\in I}$ of positive operators on $\mathbf{C}^d$ such that

$$\sum_{i\in I} M_i = \mathrm{Id}.$$

One could consider also continuous POVMs, where the finite sum is replaced by an integral. However this is not necessary, since continuous POVMs appear as limit cases of discrete POVMs which we consider here (see Chapter 6).

Given a POVM $\mathrm{M} = (M_i)_{i\in I}$ on $\mathbf{C}^d$, and denoting by $\{|i\rangle,\ i \in I\}$ an orthonormal basis of $\mathbf{C}^{|I|}$, we may associate to M the CPTP (Completely Positive and Trace-Preserving) map

$$\mathscr{M} : \Delta \in \mathscr{H}(\mathbf{C}^d) \mapsto \sum_{i\in I} \mathrm{Tr}\left(M_i\Delta\right)|i\rangle\langle i| \in \mathscr{H}\left(\mathbf{C}^{|I|}\right).$$

The reader is referred to Chapter 2, Section 2.3, for further comments. The measurement (semi-)norm associated to M is then defined, for any $\Delta \in \mathscr{H}(\mathbf{C}^d)$, as

$$\|\Delta\|_{\mathrm{M}} := \|\mathscr{M}(\Delta)\|_1 = \sum_{i\in I} |\,\mathrm{Tr}\left(M_i\Delta\right)|.$$

Note that for any $\Delta \in \mathscr{H}(\mathbf{C}^d)$, $\|\Delta\|_{\mathrm{M}} \leqslant \|\Delta\|_1$, with equality if $\Delta \in \mathscr{H}_+(\mathbf{C}^d)$.

In general, $\|\cdot\|_{\mathrm{M}}$ is a semi-norm, and may vanish on non-zero Hermitians. A necessary and sufficient condition for $\|\cdot\|_{\mathrm{M}}$ to be a norm is that the POVM $\mathrm{M} = (M_i)_{i\in I}$ is informationally complete, i.e. that the family of operators $(M_i)_{i\in I}$ spans $\mathscr{H}(\mathbf{C}^d)$ as a linear space. This especially implies that M has a total number of outcomes satisfying $|I| \geqslant d^2 = \dim \mathscr{H}(\mathbf{C}^d)$.

We denote by $B_{\|\cdot\|_{\mathrm{M}}}$ the unit ball associated to $\|\cdot\|_{\mathrm{M}}$, and by K_{M} the polar of $B_{\|\cdot\|_{\mathrm{M}}}$ (i.e. the unit ball associated to the norm dual to $\|\cdot\|_{\mathrm{M}}$). In other words, this means that the support function of K_{M} is defined, for any $\Delta \in \mathscr{H}(\mathbf{C}^d)$, as

$$h_{K_{\mathrm{M}}}(\Delta) = \|\Delta\|_{\mathrm{M}}. \tag{7.1}$$

Precise definitions of these concepts are given in Chapter 4, Section 4.1.

More generally, one can define the "measurement" or "distinguishability" norm associated to a whole set $\mathbf{M}$ of POVMs on $\mathbf{C}^d$ as

$$\|\cdot\|_{\mathbf{M}} := \sup_{\mathrm{M}\in\mathbf{M}} \|\cdot\|_{\mathrm{M}}.$$

The corresponding unit ball, and its polar, are

$$B_{\|\cdot\|_{\mathbf{M}}} = \bigcap_{\mathrm{M}\in\mathbf{M}} B_{\|\cdot\|_{\mathrm{M}}},$$

$$K_{\mathbf{M}} = \mathrm{conv}\left(\bigcup_{\mathrm{M}\in\mathbf{M}} K_{\mathrm{M}}\right).$$

As mentioned earlier on in the introduction, these measurement norms are related to the task of distinguishing quantum states. Let us consider the situation where a system (with associated Hilbert space $\mathbf{C}^d$) can be either in state ρ or in state σ, with equal prior probabilities 1/2. It is known [105, 102] that a decision process based on the maximum likelihood rule after performing the POVM M on the system yields a probability of error

$$\mathbf{P}_{error} = \frac{1}{2}\left(1 - \left\|\frac{1}{2}\rho - \frac{1}{2}\sigma\right\|_{\mathrm{M}}\right).$$

In this context, the operational interpretation of the quantity $\|\rho-\sigma\|_{\mathrm{M}}$ is thus clear (and actually justifies the terminology of "distinguishability norm"): up to a factor 1/2, it is nothing else than the bias of the POVM M on the state pair (ρ, σ).

Something that is worth pointing at is that, for any set $\mathbf{M}$ of POVMs on $\mathbf{C}^d$, there exists a set $\widetilde{\mathbf{M}}$ of 2-outcome POVMs on $\mathbf{C}^d$ which is such that $\|\cdot\|_{\mathbf{M}} = \|\cdot\|_{\widetilde{\mathbf{M}}}$. It may be explicitly defined as

$$\widetilde{\mathbf{M}} := \left\{ \big(M, \mathrm{Id} - M\big),\ \exists\ (M_i)_{i\in I} \in \mathbf{M},\ \exists\ \widetilde{I} \subset I :\ M = \sum_{i\in\widetilde{I}} M_i \right\}.$$

Note then that

$$K_{\mathbf{M}} = \mathrm{conv}\left\{2M - \mathrm{Id},\ (M, \mathrm{Id} - M) \in \widetilde{\mathbf{M}}\right\}.$$

7.2.2 On the complexity of the class of all POVMs

Denote by **ALL** the family of all POVMs on $\mathbf{C}^d$. As we already noticed, $\|\cdot\|_{\mathbf{ALL}} = \|\cdot\|_1$ and therefore $K_{\mathbf{ALL}}$ equals $[-\mathrm{Id}, \mathrm{Id}]$, which is the unit ball in $\mathscr{H}(\mathbf{C}^d)$ for the operator norm.

The family **ALL** is obviously infinite. Since real-life situations can involve only finitely many apparatuses, it makes sense to ask what must be the cardinality of a finite family of POVMs $\mathbf{M}$ which achieves close to perfect discrimination, i.e. such that the inequality $\|\cdot\|_{\mathbf{M}} \geqslant \lambda\|\cdot\|_{\mathbf{ALL}}$ holds for some $0 < \lambda < 1$. We show that the answer is exponential in d^2. More precisely, we have the theorem below.

Theorem 7.2.1. *There are universal constants $c, C > 0$ such that the following holds:*

(i) For any dimension d and any $0 < \varepsilon < 1$, there is a family $\mathbf{M}$ consisting of at most $\exp(C|\log\varepsilon|d^2)$ POVMs on $\mathbf{C}^d$ such that $\|\cdot\|_{\mathbf{M}} \geqslant (1-\varepsilon)\|\cdot\|_{\mathbf{ALL}}$.

(ii) For any $\varepsilon > C/\sqrt{d}$, any family $\mathbf{M}$ of POVMs on $\mathbf{C}^d$ such that $\|\cdot\|_{\mathbf{M}} \geqslant \varepsilon\|\cdot\|_{\mathbf{ALL}}$ contains at least $\exp(c\varepsilon^2 d^2)$ POVMs.

Theorem 7.2.1 is proved in Section 7.3. It is clear that the conclusion of (ii) fails for $\varepsilon \lesssim 1/\sqrt{d}$, since a single POVM M (e.g. the uniform POVM, see [137] or Chapter 6 of the present manuscript) may satisfy $\|\cdot\|_{\mathrm{M}} \gtrsim \|\cdot\|_1/\sqrt{d}$.

7.2.3 Locally restricted measurements on a bipartite quantum system

We now study the class of locally restricted POVMs. We assume that the underlying global Hilbert space is the tensor product of several local Hilbert spaces. However, for simplicity, we focus on the case of a bipartite system in which both parts play the same role and consider the Hilbert space $\mathrm{H} = \mathbf{C}^d \otimes \mathbf{C}^d$. Several classes of POVMs can be defined on H due to various levels of locality restrictions (consult [137] or [130] for further information).

The most restricted class of POVMs on H is the one of local measurements, whose elements are tensor products of measurements on each of the sub-systems:

$$\mathbf{LO} := \left\{ (M_i \otimes N_j)_{i\in I, j\in J} \ :\ M_i \geqslant 0,\ N_j \geqslant 0,\ \sum_{i\in I} M_i = \mathrm{Id}_{\mathbf{C}^d},\ \sum_{j\in J} N_j = \mathrm{Id}_{\mathbf{C}^d} \right\}.$$

This corresponds to the situation where parties are not allowed to communicate.

Then, we consider the class of separable measurements, whose elements are the measurements on H made of tensor operators

$$\mathbf{SEP} := \left\{ (M_j \otimes N_j)_{j\in J} \ :\ M_j \geqslant 0,\ N_j \geqslant 0,\ \sum_{j\in J} M_j \otimes N_j = \mathrm{Id}_{\mathbf{C}^d\otimes\mathbf{C}^d} \right\}.$$

An important subclass of **SEP** is the class **LOCC** (Local Operations and Classical Communication) of measurements that can be implemented by a finite sequence of local operations on the sub-systems followed by classical communication between the parties. This class can be described recursively as the smallest subclass of **SEP** which contains **LO** and is stable under the following operation: given a POVM $\mathrm{M} = (M_i)_{i\in I}$ on $\mathbf{C}^d$, and for each $i \in I$ a **LOCC** POVM $\big(R(i)_j \otimes S(i)_j\big)_{j\in J_i}$, the POVMs

$$\Big(M_i^{1/2} R(i)_j M_i^{1/2} \otimes S(i)_j\Big)_{i\in I, j\in J_i} \quad \text{and} \quad \Big(R(i)_j \otimes M_i^{1/2} S(i)_j M_i^{1/2}\Big)_{i\in I, j\in J_i}$$

are in **LOCC**. A subclass of **LOCC** is the class $\mathbf{LOCC}^{\rightarrow}$ of one-way LOCC POVMs, which has a simpler description

$$\mathbf{LOCC}^{\rightarrow} := \left\{ (M_i \otimes N_{i,j})_{i\in I, j\in J_i} \ : \ M_i \geqslant 0,\ N_{i,j} \geqslant 0,\ \sum_{i\in I} M_i = \mathrm{Id}_{\mathbf{C}^d},\ \sum_{j\in J_i} N_{i,j} = \mathrm{Id}_{\mathbf{C}^d} \right\}.$$

Finally, we consider the class of positive under partial transpose (PPT) measurements, whose elements are the measurements on H made of operators that remain positive when partially transposed on one sub-system:

$$\mathbf{PPT} := \left\{ (M_j)_{j\in J} \ : \ M_j \geqslant 0,\ M_j^{\Gamma} \geqslant 0,\ \sum_{j\in J} M_j = \mathrm{Id}_{\mathbf{C}^d\otimes\mathbf{C}^d} \right\}.$$

The partial transposition Γ is defined by its action on tensor operators on H: $(M\otimes N)^{\Gamma} := M^T \otimes N$, M^T denoting the usual transpose of M. Let us point out that, even though the expression of a matrix transpose depends on the chosen basis, its eigenvalues on the contrary are intrinsic. Therefore the PPT notion is basis-independent.

It is clear from the definitions that we have the chain of inclusions

$$\mathbf{LO} \subset \mathbf{LOCC}^{\rightarrow} \subset \mathbf{LOCC} \subset \mathbf{SEP} \subset \mathbf{PPT} \subset \mathbf{ALL}$$

and consequently the chain of norm inequalities

$$\|\cdot\|_{\mathbf{LO}} \leqslant \|\cdot\|_{\mathbf{LOCC}^{\rightarrow}} \leqslant \|\cdot\|_{\mathbf{LOCC}} \leqslant \|\cdot\|_{\mathbf{SEP}} \leqslant \|\cdot\|_{\mathbf{PPT}} \leqslant \|\cdot\|_{\mathbf{ALL}}. \tag{7.2}$$

All the inequalities in (7.2) are known to be strict provided $d > 2$. Note though that the difference between the norms $\|\cdot\|_{\mathbf{LOCC}^{\rightarrow}}$ and $\|\cdot\|_{\mathbf{LOCC}}$, as well as between $\|\cdot\|_{\mathbf{LOCC}}$ and $\|\cdot\|_{\mathbf{SEP}}$, has been established only very recently [45].

Here, we are interested in the high-dimensional behaviour of these norms, and the general question we investigate is whether or not the various gaps in the hierarchy are bounded (independently of the dimension of the subsystems). It is already known that the gap between **PPT** and **ALL** is unbounded. An important example is provided by the symmetric state π_s and the antisymmetric state π_a on $\mathbf{C}^d \otimes \mathbf{C}^d$ (see Chapter 3, Section 3.2, for precise definitions and further comments), which satisfy (see e.g. [64])

$$\|\pi_s - \pi_a\|_{\mathbf{ALL}} = 2 \quad \text{while} \quad \|\pi_s - \pi_a\|_{\mathbf{PPT}} = \frac{4}{d+1}.$$

We show however (see Theorem 7.2.2) that such feature is not generic. This is in contrast with the gap between **SEP** and **PPT** which we prove to be generically unbounded (see Theorem 7.2.2). We also provide examples of unbounded gap between **LO** and $\mathbf{LOCC}^{\rightarrow}$ (see Theorems 7.2.4 and 7.4.4), as well as between $\mathbf{LOCC}^{\rightarrow}$ and **LOCC** (see Theorem 7.4.5). But we do not know if this situation is typical. Regarding the gap between **LOCC** and **SEP**, determining whether it is bounded is still an open problem.

Note also that for states of low rank, the gaps between these norms remain bounded. It follows from the results of [130] that, for $\Delta \in \mathscr{H}(\mathbf{C}^d \otimes \mathbf{C}^d)$ of rank r, we have

$$\|\Delta\|_{\mathrm{LO}} \geqslant \frac{1}{18\sqrt{r}} \|\Delta\|_{\mathbf{ALL}}.$$

7.2.4 Discriminating power of the different classes of locally restricted measurements

Our main result compares the efficiency of the classes $\mathbf{LOCC}^{\rightarrow}$, **LOCC**, **SEP**, **PPT** and **ALL** to perform a typical discrimination task. Here "typical" means the following: we consider the problem of distinguishing ρ from σ, where ρ and σ are random states, chosen independently at random with respect to the uniform measure (i.e. the Lebesgue measure induced by the Hilbert–Schmidt distance) on the set of all states. It turns out that the PPT constraint on the allowed measurements is not very restrictive, affecting typically the performance by only a constant factor, while the separability one implies a more substantial loss. This shows that generic bipartite states are data hiding: separable measurements (and even more so local measurements followed by classical communication) can poorly distinguish them (see [99] for another instance of this phenomenon and Section 7.6 for a more detailed discussion on that topic).

Theorem 7.2.2. *There are universal constants $C, c > 0$ such that the following holds. Given a dimension d, let ρ and σ be random states, independent and uniformly distributed on the set of states on $\mathbf{C}^d \otimes \mathbf{C}^d$. Then, with high probability,*

$$c \leqslant \|\rho - \sigma\|_{\mathbf{PPT}} \leqslant \|\rho - \sigma\|_{\mathbf{ALL}} \leqslant C,$$

$$\frac{c}{\sqrt{d}} \leqslant \|\rho - \sigma\|_{\mathbf{LOCC}^{\rightarrow}} \leqslant \|\rho - \sigma\|_{\mathbf{LOCC}} \leqslant \|\rho - \sigma\|_{\mathbf{SEP}} \leqslant \frac{C}{\sqrt{d}}.$$

Here, "with high probability" means that the probability that one of the conclusions fails is less than $\exp(-c_0 d^3)$ for some constant $c_0 > 0$.

An immediate consequence of the high probability estimates is that one can find in $\mathbf{C}^d \otimes \mathbf{C}^d$ exponentially many states which are pairwise data hiding.

Corollary 7.2.3. *There are constants $C, c > 0$ such that, if $\mathcal{A}$ denotes a set of $\exp(cd^3)$ independent random states uniformly distributed on the set of states on $\mathbf{C}^d \otimes \mathbf{C}^d$, with high probability any pair of distinct states $\rho, \sigma \in \mathcal{A}$ satisfies the conclusions of Theorem 7.2.2.*

We deduce Theorem 7.2.2 from estimates on the mean width and the volume of the unit balls $K_{\mathbf{LOCC}^{\rightarrow}}$, $K_{\mathbf{SEP}}$ and $K_{\mathbf{PPT}}$. The use of concentration of measure allows to pass from these global estimates to the estimates in a typical direction that appear in Theorem 7.2.2. We include all this material in Section 7.5.

We also show that even the smallest amount of communication has a huge influence: we give examples of states which are perfectly distinguishable under local measurements and one-way classical communication but very poorly distinguishable under local measurements with no communication between the parties.

Theorem 7.2.4. *There is a universal constant $C > 0$ such that the following holds: for any d, there exists states ρ and σ on $\mathbf{C}^d \otimes \mathbf{C}^d$ such that*

$$\|\rho - \sigma\|_{\mathbf{LOCC}^{\rightarrow}} = 2,$$

and

$$\|\rho - \sigma\|_{\mathbf{LO}} \leqslant \frac{C}{\sqrt{d}}. \tag{7.3}$$

These states are constructed as follows: assuming without loss of generality that d is even, let E be a fixed $d/2$-dimensional subspace of $\mathbf{C}^d$, let $U_1, \ldots, U_d$ be random independent Haar-distributed unitaries on $\mathbf{C}^d$, and define the random states $\rho_i = U_i P_E U_i^{\dagger}/(d/2)$ and $\sigma_i = U_i P_{E^{\perp}} U_i^{\dagger}/(d/2)$, $1 \leqslant i \leqslant d$, on $\mathbf{C}^d$ (where P_E and $P_{E^{\perp}}$ denote the orthogonal projections onto E and $E^{\perp}$ respectively). Then, denoting by $\{|1\rangle, \ldots, |d\rangle\}$ an orthonormal basis of $\mathbf{C}^d$, define

$$\rho = \frac{1}{d} \sum_{i=1}^{d} |i\rangle\langle i| \otimes \rho_i \quad \text{and} \quad \sigma = \frac{1}{d} \sum_{i=1}^{d} |i\rangle\langle i| \otimes \sigma_i.$$

The pair (ρ, σ) satisfies (7.3) with high probability.

Theorem 7.2.4 is proved in Section 7.4. It is built on the idea that, typically, a single POVM cannot succeed simultaneously in several "sufficiently different" discrimination tasks. The above construction is also generalized there to show the existence of an unbounded gap between $\mathbf{LOCC}^{\rightarrow}$ and $\mathbf{LOCC}$.

7.3 On the complexity of the class of all POVMs

In this section, we determine how many distinct POVMs a set $\mathbf{M}$ of POVMs on $\mathbf{C}^d$ must contain in order to approximate the set $\mathbf{ALL}$ of all POVMs on $\mathbf{C}^d$ (in the sense that $\lambda\|\cdot\|_{\mathbf{ALL}} \leqslant \|\cdot\|_{\mathbf{M}} \leqslant \|\cdot\|_{\mathbf{ALL}}$ for some $0 < \lambda < 1$).

The reason for the $\exp(d^2)$ scaling in the first part of Theorem 7.2.1 is that these POVMs should be able to discriminate any two states within the family of states $\{P_E/\dim E\}$, where E varies among all subspaces of $\mathbf{C}^d$, and P_E denotes the orthogonal projection onto E. The set of k-dimensional subspaces of $\mathbf{C}^d$ has dimension $k(d-k)$, which is of order d^2 when k is proportional to d.

The second part of Theorem 7.2.1 requires an extra ingredient, since a single POVM may be able to discriminate exponentially many pairs of subspaces. The concept of mean width (see Chapter 4, Section 4.1) provides a neat answer to this problem.

To begin with, we prove the first part of Theorem 7.2.1. Note that the condition $\|\cdot\|_{\mathbf{M}} \geqslant (1-\varepsilon)\|\cdot\|_{\mathbf{ALL}}$ is equivalent to $K_{\mathbf{M}} \supset (1-\varepsilon)[-\mathrm{Id}, \mathrm{Id}]$, the set K_{M} being defined in (7.1). We thus only have to make use of the well-known lemma below.

Lemma 7.3.1 (Approximation of convex bodies by polytopes)**.** *Given a symmetric convex body $K \subset \mathbf{R}^n$ and $0 < \varepsilon < 1$, there is a finite family $(x_i)_{i\in I}$ such that $|I| \leqslant (3/\varepsilon)^n$ and*

$$(1-\varepsilon)K \subset \mathrm{conv}\{\pm x_i \ : \ i \in I\} \subset K.$$

Proof. Let $\mathscr{A}$ be an ε-net in K, with respect to $\|\cdot\|_K$ (the gauge of K, as defined in Chapter 4, Section 4.1). A standard volumetric argument (see Lemma 4.2.3 in Chapter 4, Section 4.2) shows that we may ensure that $|\mathscr{A}| \leqslant (3/\varepsilon)^n$. Let $P := \mathrm{conv}(\pm\mathscr{A}) \subset K$. Given any $x \in K$, there exists $x' \in \mathscr{A}$ such that $\|x - x'\|_K \leqslant \varepsilon$. Therefore

$$\|x\|_P \leqslant \|x'\|_P + \|x - x'\|_P \leqslant 1 + \varepsilon A,$$

where $A := \sup\{\|y\|_P \ : \ y \in K\}$. Taking supremum over $x \in K$, we obtain $A \leqslant 1 + \varepsilon A$ and therefore (A is easily seen to be finite) $A \leqslant (1-\varepsilon)^{-1}$. We thus proved the inequality $\|\cdot\|_P \leqslant (1-\varepsilon)^{-1}\|\cdot\|_K$, which is equivalent to the inclusion $(1-\varepsilon)K \subset P$. □

When applied to the d^2-dimensional convex body $K_{\mathbf{ALL}} = [-\mathrm{Id}, \mathrm{Id}]$, Lemma 7.3.1 implies that there is a finite family $(A_i)_{i\in I} \subset [-\mathrm{Id}, \mathrm{Id}]$ with $|I| \leqslant (3/\varepsilon)^{d^2}$ and $\mathrm{conv}\{\pm A_i \ : \ i \in I\} \supset (1-\varepsilon)[-\mathrm{Id}, \mathrm{Id}]$. For every $i \in I$, we may consider the POVM

$$\mathrm{M}_i := \left(\frac{\mathrm{Id} + A_i}{2}, \frac{\mathrm{Id} - A_i}{2}\right).$$

If we denote $\mathbf{M} := \{\mathrm{M}_i \ : \ i \in I\}$, then for any $i \in I$, $\pm A_i \in K_{\mathrm{M}_i}$ and therefore $(1-\varepsilon)[-\mathrm{Id}, \mathrm{Id}] \subset K_{\mathbf{M}}$, which is precisely what we wanted to prove.

We now show the second part of Theorem 7.2.1. The key observation is the following lemma, where we denote by α_n the mean width of a segment $[-x, x]$ for x a unit vector in $\mathbf{R}^n$, so that $\alpha_n \sim \sqrt{2/\pi n}$ (see Chapter 4, Section 4.1).

Lemma 7.3.2. *Let* M *be a POVM on $\mathbf{C}^d$. Then the mean width of the set K_{M}, defined by equation* (7.1)*, satisfies $w(K_{\mathrm{M}}) \leqslant d\alpha_{d^2}$, with equality if* M *is a rank-*1 *POVM (note that $d\alpha_{d^2}$ is of order* 1*).*

It may be pointed out that the assertion of Lemma 7.3.2 implies that, as far as the mean width is concerned, all rank-1 POVMs are comparable.

Proof. Given any POVM M, there is a rank-1 POVM M′ such that $K_{\mathrm{M}} \subset K_{\mathrm{M}'}$ (this is easily seen by splitting the POVM elements from M as a sum of rank-1 operators). Therefore, it suffices to show that $w(K_{\mathrm{M}}) = d\alpha_{d^2}$ for any rank-1 POVM. Let $\mathrm{M} = (p_i|\psi_i\rangle\langle\psi_i|)_{i\in I}$ be a rank-1 POVM, where $(p_i)_{i\in I}$ are positive numbers and $(\psi_i)_{i\in I}$ are unit vectors such that

$$\sum_{i\in I} p_i|\psi_i\rangle\langle\psi_i| = \mathrm{Id}.$$

By taking the trace, we check that the total mass of $\{p_i \ : \ i \in I\}$ equals d. We then have, for any $\Delta \in \mathscr{H}(\mathbf{C}^d)$,

$$h_{K_{\mathrm{M}}}(\Delta) = \sum_{i\in I} p_i|\langle\psi_i|\Delta|\psi_i\rangle|.$$

Hence, denoting by $S_{HS}(\mathbf{C}^d)$ the Hilbert–Schmidt unit sphere of $\mathscr{H}(\mathbf{C}^d)$ (which has dimension d^2-1) equipped with the uniform measure σ, the mean width of K_{M} can be computed as

$$w(K_{\mathrm{M}}) = \int_{S_{HS}(\mathbf{C}^d)} h_{K_{\mathrm{M}}}(\Delta)\,\mathrm{d}\sigma(\Delta) = \sum_{i\in I} p_i \left(\int_{S_{HS}(\mathbf{C}^d)} |\langle\psi_i|\Delta|\psi_i\rangle|\,\mathrm{d}\sigma(\Delta)\right) = \sum_{i\in I} p_i\alpha_{d^2} = d\alpha_{d^2}.$$ □

Assume that $\mathbf{M}$ is a family of N POVMs such that $\|\Delta\|_{\mathbf{M}} \geqslant \varepsilon\|\Delta\|_1$ for any $\Delta \in \mathscr{H}(\mathbf{C}^d)$. This implies that $K_{\mathbf{M}} \supset \varepsilon[-\mathrm{Id}, \mathrm{Id}]$ and therefore that

$$w(K_{\mathbf{M}}) \geqslant \varepsilon w([-\mathrm{Id}, \mathrm{Id}]) \simeq \varepsilon\sqrt{d}, \tag{7.4}$$

where we used last the estimate on the mean width of $[-\mathrm{Id}, \mathrm{Id}]$ from Theorem 7.8.1 in Appendix 7.8. On the other hand, we have

$$K_{\mathbf{M}} = \mathrm{conv}\left(\bigcup_{\mathrm{M}\in\mathbf{M}} K_{\mathrm{M}}\right), \tag{7.5}$$

so that $K_{\mathbf{M}}$ is the convex hull of N sets, each of them of mean width bounded by an absolute constant (by Lemma 7.3.2). We may apply Lemma 4.1.5 in Chapter 4, Section 4.1, with $\lambda = \sqrt{d}$ since $[-\mathrm{Id}, \mathrm{Id}]$ is contained in the Hilbert–Schmidt ball of radius $\sqrt{d}$. Recalling that the ambient dimension is $n = d^2$, we get

$$w(K_{\mathbf{M}}) \leqslant C\left(1 + \frac{\sqrt{\log N}}{\sqrt{d}}\right). \tag{7.6}$$

A comparison of the bounds (7.4) and (7.6) immediately yields $\log N \gtrsim \varepsilon^2 d^2$, as required.

7.4 Existing unbounded gap between LO and LOCC

7.4.1 Unbounded gap LO/LOCC$^{\rightarrow}$

In this section we give a proof of Theorem 7.2.4. Let $\{|1\rangle, \ldots, |d\rangle\}$ be an orthonormal basis of $\mathbf{C}^d$. For d even, we consider a fixed $d/2$-dimensional subspace $E \subset \mathbf{C}^d$, and denote $\Delta_0 = 2P_E - \mathrm{Id}$. We then pick $U_1, \ldots, U_d$ random independent Haar-distributed unitaries on $\mathbf{C}^d$, and for $1 \leqslant i \leqslant d$ we consider the random operators $\Delta_i = U_i \Delta_0 U_i^{\dagger}$. We finally introduce

$$\Delta = \sum_{i=1}^{d} |i\rangle\langle i| \otimes \Delta_i. \tag{7.7}$$

For each $1 \leqslant i \leqslant d$, let $\mathrm{M}_i = (M_i, \mathrm{Id} - M_i)$ be a POVM on $\mathbf{C}^d$ such that $\|\Delta_i\|_{\mathrm{M}_i} = \|\Delta_i\|_1$. Then,

$$\mathrm{M} = (|i\rangle\langle i| \otimes M_i, |i\rangle\langle i| \otimes (\mathrm{Id} - M_i))_{1\leqslant i\leqslant d}$$

is a POVM on $\mathbf{C}^d \otimes \mathbf{C}^d$ which is in $\mathbf{LOCC}^{\rightarrow}$. Therefore,

$$\|\Delta\|_{\mathbf{LOCC}^{\rightarrow}} \geqslant \|\Delta\|_{\mathrm{M}} = \sum_{i=1}^{d} \|\Delta_i\|_1 = d^2,$$

and there is actually equality in the inequality above since we also have

$$\|\Delta\|_{\mathbf{LOCC}^{\rightarrow}} \leqslant \|\Delta\|_1 = \sum_{i=1}^{d} \|\Delta_i\|_1 = d^2.$$

Theorem 7.2.4 will follow (with ρ and σ being the positive and negative parts of Δ, after renormalization) if we prove that $\|\Delta\|_{\mathbf{LO}} \lesssim Cd^{3/2}$ with high probability.

Proposition 7.4.1. *For $\Delta \in \mathscr{H}(\mathbf{C}^d \otimes \mathbf{C}^d)$ defined as in (7.7), we have*

$$\|\Delta\|_{\mathbf{LO}} = \sup\left\{\sum_{i=1}^{d} \|\Delta_i\|_{\mathrm{N}} \ : \ \mathrm{N} \text{ POVM on } \mathbf{C}^d\right\}. \tag{7.8}$$

This quantity can be upper bounded as follows, where $\mathscr{A}$ denotes a $1/16$-net in $S_{\mathbf{C}^d}$

$$\|\Delta\|_{\mathbf{LO}} \leqslant d \sup_{x\in S_{\mathbf{C}^d}} \sum_{i=1}^{d} |\langle x|\Delta_i|x\rangle| \tag{7.9}$$

$$\leqslant 2d \sup_{x\in\mathscr{A}} \sum_{i=1}^{d} |\langle x|\Delta_i|x\rangle|. \tag{7.10}$$

Proof. The inequality $\geqslant$ in (7.8) follows by considering the **LO** POVM $(|i\rangle\langle i|)_{1\leqslant i\leqslant d} \otimes \mathrm{N}$. Conversely, given POVMs $\mathrm{M} = (M_j)_{j\in J}$ and $\mathrm{N} = (N_k)_{k\in K}$ on $\mathbf{C}^d$, we have

$$\begin{aligned}\|\Delta\|_{\mathrm{M}\otimes\mathrm{N}} &= \sum_{j\in J,k\in K} \left| \sum_{i=1}^{d} \mathrm{Tr}\left(\left(|i\rangle\langle i| \otimes \Delta_i\right)(M_j \otimes N_k)\right)\right| \\ &\leqslant \sum_{i=1}^{d} \left(\sum_{j\in J} |\langle i|M_j|i\rangle|\right)\left(\sum_{k\in K} |\mathrm{Tr}\left(\Delta_i N_k\right)|\right) \\ &\leqslant \sum_{i=1}^{d} \|\Delta_i\|_{\mathrm{N}},\end{aligned}$$

the last inequality being because, for each $1 \leqslant i \leqslant d$, $\sum_{j\in J} |\langle i|M_j|i\rangle| = \sum_{j\in J}\langle i|M_j|i\rangle = \langle i|i\rangle = 1$. Taking the supremum over M and N gives the inequality $\leqslant$ in (7.8).

The supremum in (7.8) is unchanged when restricting to the supremum on POVMs whose elements have rank 1, since splitting the POVM elements as sum of rank 1 operators does not decrease the distinguishability norm. If N is such a POVM, its elements can be written as $(\alpha_k|x_k\rangle\langle x_k|)_{k\in K}$, where $(x_k)_{k\in K}$ are unit vectors and $(\alpha_k)_{k\in K}$ positive numbers satisfying $\sum_{k\in K}\alpha_k = d$. We thus have in that case

$$\sum_{i=1}^{d} \|\Delta_i\|_{\mathrm{N}} = \sum_{i=1}^{d}\sum_{k\in K} |\,\mathrm{Tr}(\Delta_i \cdot \alpha_k|x_k\rangle\langle x_k|)| \leqslant d \sup_{x\in S_{\mathbf{C}^d}} \sum_{i=1}^{d} |\langle x|\Delta_i|x\rangle|\,,$$

proving (7.9).

To prove (7.10), we introduce the function g defined for $x, y \in \mathbf{C}^d$ by $g(x,y) = \sum_{i=1}^{d} |\langle x|\Delta_i|y\rangle|$, and the function f defined for $x \in \mathbf{C}^d$ by $f(x) = g(x,x)$. Denote by G the supremum of g over $S_{\mathbf{C}^d} \times S_{\mathbf{C}^d}$, by F the supremum of f over $S_{\mathbf{C}^d}$ and by F' the supremum of f over a δ-net $\mathscr{A}$. For any $x, y \in \mathbf{C}^d$, we have by the polarisation identity

$$\langle x|\Delta_i|y\rangle = \frac{1}{4}\left(\langle x+y|\Delta_i|x+y\rangle + i\langle x+iy|\Delta_i|x+iy\rangle - \langle x-y|\Delta_i|x-y\rangle - i\langle x-iy|\Delta_i|x-iy\rangle\right),$$

so that $g(x,y) \leqslant \left(f(x+y) + f(x+iy) + f(x-y) + f(x-iy)\right)/4$ and therefore $G \leqslant 4F$.

Given $x \in S_{\mathbf{C}^d}$, there exists $x' \in \mathscr{A}$ such that $\|x - x'\|_2 \leqslant \delta$, and by the triangle inequality

$$|\langle x|\Delta_i|x\rangle| \leqslant |\langle x|\Delta_i|x - x'\rangle| + |\langle x - x'|\Delta_i|x'\rangle| + |\langle x'|\Delta_i|x'\rangle|\,.$$

Summing over i and taking supremum over $x \in S_{\mathbf{C}^d}$ gives

$$F \leqslant 2\delta G + F' \leqslant 8\delta F + F'.$$

For $\delta = 1/16$, we obtain $F \leqslant 2F'$, and therefore (7.10) follows from (7.9). □

To bound $\|\Delta\|_{\mathbf{LO}}$, we combine Proposition 7.4.1 with the following result.

Proposition 7.4.2. *Let x be a fixed unit vector in $\mathbf{C}^d$, E be a fixed $d/2$-dimensional subspace of $\mathbf{C}^d$ and $\Delta_0 = 2P_E - \mathrm{Id}$, $(U_i)_{1\leqslant i\leqslant n}$ be Haar-distributed independent random unitaries on $\mathbf{C}^d$, and for each $1 \leqslant i \leqslant n$, set $\Delta_i = U_i\Delta_0 U_i^\dagger$. Then, for any $t > 1$,*

$$\mathbf{P}\left(\sum_{i=1}^{n} |\langle x|\Delta_i|x\rangle| \geqslant (1+t)n\,\mathbf{E}\,|\langle x|\Delta_1|x\rangle|\right) \leqslant e^{-c_0 nt},$$

$c_0 > 0$ being a universal constant.

Proof. Proposition 7.4.2 is a consequence of Proposition 6.7.2 in Chapter 6, which is itself a variation on Bernstein inequalities (recalled as Theorem 4.2.5 in Chapter 4, Section 4.2). The quantity $\mathbf{E}\,|\langle x|\Delta_1|x\rangle|$ is equal to the so-called "uniform norm" of Δ_1 (see [137] and Chapter 6 of the present manuscript) and we use the bound from [130]

$$\mathbf{E}\,|\langle x|\Delta_1|x\rangle| \leqslant \frac{1}{d}\|\Delta_1\|_2 = \frac{1}{\sqrt{d}}.$$

□

We now complete the proof of Theorem 7.2.4. Let $\mathscr{A}$ be a minimal 1/16-net in $S_{\mathbf{C}^d}$, so that $|\mathscr{A}| \leqslant 48^{2d}$ (see Lemma 4.2.3 in Chapter 4, Section 4.2). Using Propositions 7.4.1 and 7.4.2 (for $n = d$), and the union bound, we obtain that for any $t > 1$

$$\mathbf{P}\left(\|\Delta\|_{\mathbf{LO}} \geqslant 2(1+t)d^{3/2}\right) \leqslant \mathbf{P}\left(\exists\, x \in \mathscr{A} \;:\; \sum_{i=1}^{d} |\langle x|\Delta_i|x\rangle| \geqslant (1+t)\sqrt{d}\right) \leqslant 48^{2d} e^{-c_0 dt}.$$

This estimate is less than 1 when t is larger than some number t_0. This shows that $\|\Delta\|_{\mathbf{LO}} \leqslant 2(1+t_0)d^{3/2}$ with high probability while $\|\Delta\|_{\mathbf{LOCC}^{\rightarrow}} = d^2$, and Theorem 7.2.4 follows.

Remark 7.4.3. *The operator Δ defined by equation (7.7) can be rewritten as $\Delta = d^2(\rho' - \mathrm{Id}/d^2)$, with*

$$\rho' = \frac{2}{d^2}\sum_{i=1}^{d} |i\rangle\langle i| \otimes U_i P_E U_i^{\dagger}.$$

Hence by Theorem 7.2.4, $\|\rho' - \mathrm{Id}/d^2\|_{\mathbf{LO}} \leqslant C/\sqrt{d}$ with high probability, while $\|\rho' - \mathrm{Id}/d^2\|_{\mathbf{LOCC}^{\rightarrow}} = 1$. This property is characteristic of data locking states. These are states whose accessible mutual information (i.e. the maximum classical mutual information that can be achieved by local measurements) drastically underestimates their quantum mutual information (see [62] for the original description of this phenomenon, and Chapter 2, Section 2.5, in this manuscript for a reminder of the definition of mutual information). Now, following [68] and [71], data locking may also be defined in terms of distinguishability from the maximally mixed state by local measurements: informally, a state ρ on $\mathbf{C}^d \otimes \mathbf{C}^d$ which is such that $\|\rho - \mathrm{Id}/d^2\|_{\mathbf{LO}} \ll \|\rho - \mathrm{Id}/d^2\|_{\mathbf{LOCC}^{\rightarrow}}$ may be used for information locking.

7.4.2 Generalization: unbounded gaps LO/LOCC$^{\rightarrow}$ and LOCC$^{\rightarrow}$/LOCC

For each $r \in \mathbf{N}$, we define $\mathbf{LOCC}^{(\mathbf{r})}$ as the class of local POVMs with r rounds of classical communication between the parties. In particular, making the link with the notation that were previously introduced, we have $\mathbf{LOCC}^{\rightarrow} = \mathbf{LOCC}^{(\mathbf{1})}$ while $\mathbf{LOCC} = \lim_{r\to+\infty} \mathbf{LOCC}^{(\mathbf{r})}$. As before, we let E be a $d/2$-dimensional subspace of $\mathbf{C}^d$, and denote by P the orthogonal projector onto E, $P^{\perp}$ the orthogonal projector onto $E^{\perp}$. Then, define the two orthogonal states τ_+, τ_- on $\mathbf{C}^d$ by

$$\tau_+ = \frac{P}{d/2} \quad \text{and} \quad \tau_- = \frac{P^{\perp}}{d/2}.$$

Given $m \in \mathbf{N}$, let $\{|1\rangle, \ldots, |m\rangle\}$ be an orthonormal basis of $\mathbf{C}^m$, and $U_1, \ldots, U_m$ be unitaries on $\mathbf{C}^d$. Then, define the two bipartite states $\rho_+^{(1)}, \rho_-^{(1)}$ on $\mathrm{A} \otimes \mathrm{B}$, with $\mathrm{A} \equiv \mathbf{C}^m$, $\mathrm{B} \equiv \mathbf{C}^d$, as

$$\left[\rho_{+/-}^{(1)}\right]_{\mathrm{AB}} = \frac{1}{m}\sum_{i=1}^{m} |i\rangle\langle i|_{\mathrm{A}} \otimes \left[U_i \tau_{+/-} U_i^{\dagger}\right]_{\mathrm{B}}. \tag{7.11}$$

Given $n \in \mathbf{N}$, let $\{|1\rangle, \ldots, |n\rangle\}$ be an orthonormal basis of $\mathbf{C}^n$, and $V_1, \ldots, V_n$ be unitaries on $\mathbf{C}^m$. Then, define the two bipartite states $\rho_+^{(2)}, \rho_-^{(2)}$ on $\mathrm{A} \otimes \mathrm{B}$, with $\mathrm{A} \equiv \mathbf{C}^n \otimes \mathbf{C}^d$, $\mathrm{B} \equiv \mathbf{C}^m$, as

$$\left[\rho_{+/-}^{(2)}\right]_{\mathrm{AB}} = \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n} \left[|j\rangle\langle j| \otimes U_i \tau_{+/-} U_i^{\dagger}\right]_{\mathrm{A}} \otimes \left[V_j |i\rangle\langle i| V_j^{\dagger}\right]_{\mathrm{B}}. \tag{7.12}$$

Theorem 7.4.4. *Let $0 \leqslant \delta < 1/2$. Consider the states $\rho_+^{(1)}$ and $\rho_-^{(1)}$ as defined by equation (7.11) and the Hermitian $\Delta^{(1)} = \rho_+^{(1)} - \rho_-^{(1)}$ on $\mathrm{A} \otimes \mathrm{B} \equiv \mathbf{C}^m \otimes \mathbf{C}^d$. For $m \simeq d^{1-\delta}$, there exist $U_1, \ldots, U_m$ such that*

$$\|\Delta^{(1)}\|_{\mathbf{LOCC}^{(1)}} = 2 \quad \textit{and} \quad \|\Delta^{(1)}\|_{\mathbf{LO}} \lesssim \frac{1}{d^{1/2-\delta}}.$$

In words: if the hypotheses of Theorem 7.4.4 are satisfied, the states $\rho_+^{(1)}$ and $\rho_-^{(1)}$ are perfectly distinguishable by local POVMs and one round of classical communication, but very poorly by local POVMs without any classical communication.

Proof. The $\mathbf{LOCC^{(1)}}$ POVM which enables perfect discrimination of $\rho_+^{(1)}$ and $\rho_-^{(1)}$ is

$$\left(|i\rangle\langle i|_{\mathrm{A}} \otimes \left[U_i P U_i^\dagger\right]_{\mathrm{B}}, |i\rangle\langle i|_{\mathrm{A}} \otimes \left[U_i P^\perp U_i^\dagger\right]_{\mathrm{B}}\right)_{1\leqslant i\leqslant m}.$$

So the first equality is clear.

The second inequality is just a slight generalization of Theorem 7.2.4. Indeed, it was just observed in its proof that, setting $\Delta_i = U_i\tau_+U_i^\dagger - U_i\tau_-U_i^\dagger$ for each $1 \leqslant i \leqslant m$, we have

$$\|\Delta^{(1)}\|_{\mathbf{LO}} = \sup\left\{\frac{1}{m}\sum_{i=1}^m \|\Delta_i\|_{\mathrm{N}} \ : \ \mathrm{N\ POVM\ on\ } \mathbf{C}^d\right\}.$$

It was then established that, for $m = Cd^{1-\delta}$ and $U_1, \ldots, U_m$ independent Haar distributed unitaries on $\mathbf{C}^d$, the latter quantity is, with probability greater than $1/2$, smaller than $C'/d^{1/2-\delta}$ (where $C, C' > 0$ are universal constants). □

Theorem 7.4.5. *Let $0 \leqslant \delta < 1/2$. Consider the states $\rho_+^{(2)}$ and $\rho_-^{(2)}$ as defined by equation (7.12) and the Hermitian $\Delta^{(2)} = \rho_+^{(2)} - \rho_-^{(2)}$ on $\mathrm{A}\otimes\mathrm{B} \equiv \left(\mathbf{C}^n \otimes \mathbf{C}^d\right) \otimes \mathbf{C}^m$. For $n \simeq m \simeq d^{1-\delta}$, there exist $U_1, \ldots, U_m$ and $V_1, \ldots, V_n$ such that*

$$\|\Delta^{(2)}\|_{\mathbf{LOCC^{(2)}}} = 2 \quad \textit{and} \quad \|\Delta^{(2)}\|_{\mathbf{LOCC^{(1)}}} \lesssim \frac{1}{d^{1/2-\delta}}.$$

In words: if the hypotheses of Theorem 7.4.5 are satisfied, the states $\rho_+^{(2)}$ and $\rho_-^{(2)}$ are perfectly distinguishable by local POVMs and two rounds of classical communication, but very poorly by local POVMs and one round of classical communication only.

Proof. The $\mathbf{LOCC^{(2)}}$ POVM which enables perfect discrimination of $\rho_+^{(2)}$ and $\rho_-^{(2)}$ is

$$\left(|j\rangle\langle j|_{\mathrm{A}_1} \otimes \left[U_i P U_i^\dagger\right]_{\mathrm{A}_2} \otimes \left[V_j|i\rangle\langle i|V_j^\dagger\right]_{\mathrm{B}}, |j\rangle\langle j|_{\mathrm{A}_1} \otimes \left[U_i P^\perp U_i^\dagger\right]_{\mathrm{A}_2} \otimes \left[V_j|i\rangle\langle i|V_j^\dagger\right]_{\mathrm{B}}\right)_{1\leqslant i\leqslant m,\ 1\leqslant j\leqslant n}.$$

So the first equality is clear.

Let us turn to showing the second inequality. Setting $\Delta_i = U_i\tau_+U_i^\dagger - U_i\tau_-U_i^\dagger$ for each $1 \leqslant i \leqslant m$, we have by definition

$$\|\Delta^{(2)}\|_{\mathbf{LOCC^{(1)}}} = \sup \sum_\alpha \frac{1}{mn}\sum_{i=1}^m\sum_{j=1}^n \|\Delta_i \otimes |j\rangle\langle j|\|_{\mathrm{N}_\alpha} \operatorname{Tr}\left(V_j|i\rangle\langle i|V_j^\dagger M_\alpha\right)$$
$$\text{s.t. } \mathrm{M} = (M_\alpha)_\alpha \text{ POVM on } \mathbf{C}^m \text{ and } \mathrm{N}_\alpha \text{ POVMs on } \mathbf{C}^d \otimes \mathbf{C}^n.$$

Yet, if $\mathrm{N} = (N_\beta)_\beta$ is a POVM on $\mathbf{C}^d \otimes \mathbf{C}^n$, then for each $1 \leqslant j \leqslant n$, $\mathrm{N}(j) = \left(\operatorname{Tr}_n\left[\mathrm{Id}_d \otimes |j\rangle\langle j|_n N_\beta\right]\right)_\beta$ is a POVM on $\mathbf{C}^d$ which is such that, for any Hermitian Δ on $\mathbf{C}^d$, $\|\Delta \otimes |j\rangle\langle j|\|_{\mathrm{N}} = \|\Delta\|_{\mathrm{N}(j)}$. Therefore,

$$\|\Delta^{(2)}\|_{\mathbf{LOCC^{(1)}}} \leqslant \sup \sum_\alpha \frac{1}{m}\sum_{i=1}^m \left(\frac{1}{n}\sum_{j=1}^n \operatorname{Tr}\left(V_j|i\rangle\langle i|V_j^\dagger M_\alpha\right)\right)\|\Delta_i\|_{\mathrm{N}_\alpha}$$
$$\text{s.t. } \mathrm{M} = (M_\alpha)_\alpha \text{ POVM on } \mathbf{C}^m \text{ and } \mathrm{N}_\alpha \text{ POVMs on } \mathbf{C}^d.$$

Now choose unitaries $V_1, \ldots, V_n$ on $\mathbf{C}^m$ which are 1-randomizing, i.e. such that, for any unit vector $x \in \mathbf{C}^m$,

$$\left\|\frac{1}{n}\sum_{j=1}^n V_j|x\rangle\langle x|V_j^\dagger - \frac{\mathrm{Id}}{m}\right\|_\infty \leqslant \frac{1}{m}. \tag{7.13}$$

This is achieved with probability greater than $1/2$ by independent Haar distributed unitaries on $\mathbf{C}^m$ if $n = Cm$, where $C > 0$ is a universal constant (see [99]). In that case, for any POVM $\mathrm{M} = (M_\alpha)_\alpha$ on $\mathbf{C}^m$, we have that for each α and $1 \leqslant i \leqslant m$,

$$\frac{1}{n}\sum_{j=1}^n \operatorname{Tr}\left(V_j|i\rangle\langle i|V_j^\dagger M_\alpha\right) \leqslant \left\|\frac{1}{n}\sum_{j=1}^n V_j|x\rangle\langle x|V_j^\dagger\right\|_\infty \|M_\alpha\|_1 \leqslant \frac{2}{m}\operatorname{Tr} M_\alpha,$$

the last inequality being by assumption (7.13), and because $M_\alpha \geqslant 0$. Since additionally $\sum_\alpha M_\alpha = \mathrm{Id}$, so that $\sum_\alpha \mathrm{Tr}\, M_\alpha/m = 1$, we finally get

$$\begin{aligned}\|\Delta^{(2)}\|_{\mathbf{LOCC}^{(1)}} &\leqslant \sup\left\{\frac{2}{m}\sum_{i=1}^{m}\sum_{\alpha}\lambda_\alpha\|\Delta_i\|_{\mathrm{N}_\alpha} \;:\; \lambda = (\lambda_\alpha)_\alpha \text{ p.d., } \mathrm{N}_\alpha \text{ POVMs on } \mathbf{C}^d\right\}\\ &= \sup\left\{\frac{2}{m}\sum_{i=1}^{m}\|\Delta_i\|_{\mathrm{N}} \;:\; \mathrm{N} \text{ POVM on } \mathbf{C}^d\right\}.\end{aligned}$$

As explained in the proof of Theorem 7.4.4, we know that there exist universal constants $C, C' > 0$ such that, for $m = Cd^{1-\delta}$ and $U_1, \ldots, U_m$ independent Haar distributed unitaries on $\mathbf{C}^d$, the latter quantity is, with probability greater than $1/2$, smaller than $C'/d^{1/2-\delta}$. □

7.5 Generic unbounded gap between SEP and PPT

7.5.1 Volume and mean width estimates

The first step towards Theorem 7.2.2 is to estimate globally the size of the (dual) unit balls $K_{\mathbf{PPT}}$, $K_{\mathbf{SEP}}$ and $K_{\mathbf{LOCC}^{\rightarrow}}$ associated to the measurement norms $\|\cdot\|_{\mathbf{PPT}}$, $\|\cdot\|_{\mathbf{SEP}}$ and $\|\cdot\|_{\mathbf{LOCC}^{\rightarrow}}$. Classical invariants which are generally useful to quantify the size of convex bodies include the volume radius and the mean width, which are defined in Chapter 4, Section 4.1.

Recall that whenever we use tools from convex geometry in the space $\mathscr{H}(\mathbf{C}^d \otimes \mathbf{C}^d)$ (which has real dimension d^4) it is tacitly understood that we use the Euclidean structure induced by the Hilbert–Schmidt inner product. The definitions of the volume radius and the mean width of $K_{\mathbf{M}}$ thus become

$$\mathrm{vrad}(K_{\mathbf{M}}) = \left(\frac{\mathrm{vol}\, K_{\mathbf{M}}}{\mathrm{vol}\, B_{HS}(\mathbf{C}^d \otimes \mathbf{C}^d)}\right)^{1/d^4} \text{ and } w(K_{\mathbf{M}}) = \int_{S_{HS}(\mathbf{C}^d\otimes\mathbf{C}^d)} \|\Delta\|_{\mathbf{M}}\, \mathrm{d}\sigma(\Delta),$$

where $B_{HS}(\mathbf{C}^d \otimes \mathbf{C}^d)$ denotes the Hilbert–Schmidt unit ball of $\mathscr{H}(\mathbf{C}^d \otimes \mathbf{C}^d)$ and $S_{HS}(\mathbf{C}^d \otimes \mathbf{C}^d)$ its Hilbert–Schmidt unit sphere equipped with the uniform measure σ. Here are the estimates on the volume radius and the mean width of $K_{\mathbf{PPT}}$, $K_{\mathbf{SEP}}$ and $K_{\mathbf{LOCC}^{\rightarrow}}$. As a reference, recall that (on $\mathbf{C}^d \otimes \mathbf{C}^d$)

$$\mathrm{vrad}(K_{\mathbf{ALL}}) \simeq w(K_{\mathbf{ALL}}) \simeq d.$$

This follows from Theorem 7.8.1 in Appendix 7.8 once we have in mind that $K_{\mathbf{ALL}} = [-\mathrm{Id}, \mathrm{Id}]$.

Theorem 7.5.1. *In $\mathbf{C}^d \otimes \mathbf{C}^d$, one has*

$$\mathrm{vrad}\,(K_{\mathbf{PPT}}) \simeq w\,(K_{\mathbf{PPT}}) \simeq d$$

and

$$\mathrm{vrad}\,(K_{\mathbf{LOCC}^{\rightarrow}}) \simeq w\,(K_{\mathbf{LOCC}^{\rightarrow}}) \simeq \mathrm{vrad}\,(K_{\mathbf{LOCC}}) \simeq w\,(K_{\mathbf{LOCC}}) \simeq \mathrm{vrad}\,(K_{\mathbf{SEP}}) \simeq w\,(K_{\mathbf{SEP}}) \simeq \sqrt{d}.$$

To prove these results, we will make essential use of the Urysohn inequality (Theorem 4.1.3 in Chapter 4, Section 4.1): for any convex body $K \subset \mathbf{R}^n$, we have $\mathrm{vrad}(K) \leqslant w(K)$. In particular, Theorem 7.5.1 follows from the following four inequalities: (a) $w(K_{\mathbf{PPT}}) \lesssim d$, (b) $\mathrm{vrad}(K_{\mathbf{PPT}}) \gtrsim d$, (c) $w(K_{\mathbf{SEP}}) \lesssim \sqrt{d}$, and (d) $\mathrm{vrad}(K_{\mathbf{LOCC}^{\rightarrow}}) \gtrsim \sqrt{d}$.

(a) Proof that $w(K_{\mathbf{PPT}}) \lesssim d$

This follows from the inclusion $K_{\mathbf{PPT}} \subset [-\mathrm{Id}, \mathrm{Id}]$, together with the estimate on the mean width of $[-\mathrm{Id}, \mathrm{Id}]$ from Theorem 7.8.1 in Appendix 7.8.

(b) Proof that $\mathrm{vrad}(K_{\mathbf{PPT}}) \gtrsim d$

We start by noticing that

$$K_{\mathbf{PPT}} = [-\mathrm{Id}, \mathrm{Id}] \cap [-\mathrm{Id}, \mathrm{Id}]^{\Gamma}.$$

We apply the Milman–Pajor inequality (Corollary 4.1.7 in Chapter 4, Section 4.1) to the convex body $[-\mathrm{Id}, \mathrm{Id}]$ (which indeed has the origin as center of mass) and to the orthogonal transformation Γ (the partial transposition). This yields

$$\mathrm{vrad}\,(K_{\mathbf{PPT}}) \geqslant \frac{1}{2}\frac{\mathrm{vrad}\,([-\mathrm{Id}, \mathrm{Id}])^2}{w\,([-\mathrm{Id}, \mathrm{Id}])} \simeq d,$$

where we used the estimates on the volume radius and the mean width of $[-\mathrm{Id}, \mathrm{Id}]$ from Theorem 7.8.1 in Appendix 7.8.

(c) Proof that $w(K_{\mathbf{SEP}}) \lesssim \sqrt{d}$

We are going to relate $K_{\mathbf{SEP}}$ with the set $\mathcal{S}$ of separable states on $\mathbf{C}^d \otimes \mathbf{C}^d$. In fact, denoting the cone with base $\mathcal{S}$ by

$$\mathbf{R}^+\mathcal{S} := \{\lambda\rho \ : \ \lambda \in \mathbf{R}^+, \ \rho \in \mathcal{S}\},$$

we have $K_{\mathbf{SEP}} = L \cap (-L)$, where

$$L := 2\left(\mathbf{R}^+\mathcal{S} \cap [0, \mathrm{Id}]\right) - \mathrm{Id}.$$

This gives immediately an upper bound on the mean width of $K_{\mathbf{SEP}}$

$$w(K_{\mathbf{SEP}}) \leqslant w(L) \leqslant 2w(\mathbf{R}^+\mathcal{S} \cap [0, \mathrm{Id}]) \leqslant 2w(\{\lambda\rho \ : \ \lambda \in [0, d^2], \ \rho \in \mathcal{S}\}) = 2d^2 w(\mathrm{conv}(\{0\}, \mathcal{S})).$$

Now, if K, K' are two convex sets such that $K \cap K' \neq \emptyset$, then $w(\mathrm{conv}(K, K')) \leqslant w(K) + w(K')$. So, denoting by α_n the mean width of a segment $[-x, x]$ for x a unit vector in $\mathbf{R}^n$, we have

$$w(\mathrm{conv}(\{0\}, \mathcal{S})) \leqslant w(\mathrm{conv}\{0, \mathrm{Id}/d^2\}) + w(\mathcal{S}) \lesssim \frac{\alpha_{d^4}}{d} + \frac{1}{d^{3/2}} \lesssim \frac{1}{d^{3/2}},$$

where we used the estimate $w(\mathcal{S}) \simeq d^{-3/2}$ from Theorem 7.8.4 in Appendix 7.8, and the fact that $\alpha_n \simeq n^{-1/2}$ (see Chapter 4, Section 4.1).

(d) Proof that $\mathrm{vrad}(K_{\mathbf{LOCC}^\rightarrow}) \gtrsim \sqrt{d}$

We consider the following set of states on $\mathbf{C}^d \otimes \mathbf{C}^d$

$$T = \mathrm{conv}\left\{|\psi\rangle\langle\psi| \otimes \sigma \ : \ \psi \in S_{\mathbf{C}^d}, \ \sigma \text{ a state on } \mathbf{C}^d \text{ such that } \|\sigma\|_\infty \leqslant 3/d\right\}.$$

A connection between T and $\mathbf{LOCC}^\rightarrow$ is given by the following lemma.

Lemma 7.5.2. *Let ρ and ρ' be operators in T such that $\rho + \rho' = 2\mathrm{Id}/d^2$. Then, the operators $d^2\rho/6$ and $d^2\rho'/6$ belong to $K_{\mathbf{LOCC}^\rightarrow}$.*

Proof. There exist convex combinations $(\alpha_i)_{i\in I}, (\alpha'_j)_{j\in J}$, unit vectors $(\psi_i)_{i\in I}, (\psi'_j)_{j\in J}$ and states $(\sigma_i)_{i\in I}, (\sigma'_j)_{j\in J}$ satisfying $\|\sigma_i\|_\infty \leqslant 3/d$, $\|\sigma'_j\|_\infty \leqslant 3/d$, such that

$$\rho = \sum_{i\in I} \alpha_i |\psi_i\rangle\langle\psi_i| \otimes \sigma_i \quad \text{and} \quad \rho' = \sum_{j\in J} \alpha'_j |\psi'_j\rangle\langle\psi'_j| \otimes \sigma'_j.$$

Define states $(\tau_i)_{i\in I}$ and $(\tau'_j)_{j\in J}$ by the relations $\sigma_i + 2\tau_i = \sigma'_j + 2\tau'_j = 3\mathrm{Id}/d$. It can then be checked that the following POVM is in $\mathbf{LOCC}^\rightarrow$

$$\mathrm{M} = \left(\frac{d^2}{6}\alpha_i |\psi_i\rangle\langle\psi_i| \otimes \sigma_i, \ \frac{d^2}{6}\alpha_i |\psi_i\rangle\langle\psi_i| \otimes 2\tau_i, \ \frac{d^2}{6}\alpha'_j |\psi'_j\rangle\langle\psi'_j| \otimes \sigma'_j, \ \frac{d^2}{6}\alpha'_j |\psi'_j\rangle\langle\psi'_j| \otimes 2\tau'_j\right)_{i\in I, \, j\in J}.$$

Hence, the operators $d^2\rho/6$ and $d^2\rho'/6$ belong to K_{M} and therefore to $K_{\mathbf{LOCC}^\rightarrow}$. □

Let $\widetilde{T}$ be the symmetrization of T defined as $\widetilde{T} = T \cap \left\{2\mathrm{Id}/d^2 - T\right\}$. By Lemma 7.5.2 and the fact that $K_{\mathbf{LOCC}^\rightarrow}$ is centrally symmetric, we have

$$\frac{d^2}{6}\,\mathrm{conv}(\widetilde{T}, -\widetilde{T}) \subset K_{\mathbf{LOCC}^\rightarrow}.$$

We are going to give a lower bound on the volume radius of $\widetilde{T}$. The center of mass of the set T equals the maximally mixed state Id/d^2 (indeed, the center of mass commutes with local unitaries). By the Milman–Pajor inequality (Corollary 4.1.7 in Chapter 4, Section 4.1) this implies that $\mathrm{vrad}(\widetilde{T}) \geqslant \mathrm{vrad}(T)/2$. On the other hand, one has (see definitions in Appendix 7.8)

$$\mathrm{conv}(T, -T) \supset \frac{1}{d} \cdot B_1(\mathbf{C}^d) \hat{\otimes} B_\infty(\mathbf{C}^d). \tag{7.14}$$

Let us check (7.14). An extreme point of $\frac{1}{d} \cdot B_1(\mathbf{C}^d) \hat{\otimes} B_\infty(\mathbf{C}^d)$ has the form $\pm |\psi\rangle\langle\psi| \otimes A$ for $\psi \in S_{\mathbf{C}^d}$ and $A \in \mathscr{H}(\mathbf{C}^d)$ such that $\|A\|_\infty \leqslant 1/d$. Let $\varepsilon = 2 - \|A\|_1 \geqslant 1$ and let A^+, A^- be the positive and negative parts of A. Set $\lambda^\pm = \varepsilon/4 + \mathrm{Tr}\, A^\pm/2$ (so that $\lambda^+ + \lambda^- = 1$), and consider the states $\rho^\pm = (\varepsilon/4 \cdot \mathrm{Id}/d + A^\pm/2)/\lambda^\pm$. We have

$$\|\rho^\pm\|_\infty \leqslant \frac{\varepsilon/4d + 1/2d}{\varepsilon/4} \leqslant \frac{3}{d}$$

and therefore $\rho^\pm \in T$. Since $A = \lambda^+ \rho^+ - \lambda^- \rho^-$, this shows (7.14). Using Theorem 7.8.2 in Appendix 7.8, it follows that

$$\mathrm{vrad}(\mathrm{conv}(T, -T)) \gtrsim d^{-3/2}.$$

And therefore,

$$\mathrm{vrad}(\mathrm{conv}(\widetilde{T}, -\widetilde{T})) \gtrsim \mathrm{vrad}(\widetilde{T}) \gtrsim \mathrm{vrad}(T) \gtrsim \mathrm{vrad}(\mathrm{conv}(T, -T)) \gtrsim d^{-3/2},$$

the first and third inequalities being due to the Rogers–Shephard inequality (Theorem 4.1.8 in Chapter 4, Section 4.1). We eventually get

$$\mathrm{vrad}(K_{\mathbf{LOCC}^\rightarrow}) \gtrsim \sqrt{d}.$$

7.5.2 Discriminating between two generic states

Let $\mathbf{M}$ be a family of POVMs on $\mathbf{C}^d$ (possibly reduced to a single POVM). We relate the mean width $w(K_{\mathbf{M}})$ to the typical performance of $\mathbf{M}$ for discriminating two random states, chosen independently and uniformly from the set $\mathscr{D}(\mathbf{C}^d)$ of all states on $\mathbf{C}^d$.

Proposition 7.5.3. *Let $\mathbf{M}$ be a family of POVMs on $\mathbf{C}^d$, and denote $\omega := w(P_{H_0} K_{\mathbf{M}})$, where P_{H_0} stands for the orthogonal projection onto the hyperplane $H_0 \subset \mathscr{H}(\mathbf{C}^d)$ of trace 0 Hermitian operators on $\mathbf{C}^d$. Let ρ and σ be two random states, chosen independently with respect to the uniform measure on $\mathscr{D}(\mathbf{C}^d)$. Then,*

$$\mathbf{E} := \mathbf{E}\, \|\rho - \sigma\|_{\mathbf{M}} \simeq \frac{\omega}{\sqrt{d}}. \tag{7.15}$$

Moreover, we have the concentration estimate

$$\forall\ t > 0,\ \mathbf{P}\left(\left|\|\rho - \sigma\|_{\mathbf{M}} - \mathbf{E}\right| > t\right) \leqslant 2\exp(-cd^2t^2), \tag{7.16}$$

$c > 0$ being a universal constant.

We first deduce Theorem 7.2.2 from Theorem 7.5.1 and Proposition 7.5.3 (we warn the reader that we apply the latter on the space $\mathbf{C}^d \otimes \mathbf{C}^d$, and therefore the ambient dimension is d^2 instead of d).

Proof of Theorem 7.2.2. Let $\mathbf{M} \in \{\mathbf{LOCC}, \mathbf{LOCC}^\rightarrow, \mathbf{SEP}, \mathbf{PPT}\}$. While we computed $w(K_{\mathbf{M}})$ in Theorem 7.5.1, the relevant quantity here is $w(P_{H_0} K_{\mathbf{M}})$. We show that both are comparable. We first have the upper bound (see (4.1) from Chapter 4, Section 4.1)

$$w(P_{H_0} K_{\mathbf{M}}) \lesssim w(K_{\mathbf{M}}).$$

To get the reverse bound, we consider the volume radius rather than the mean width. If we denote more generally by H_t the hyperplane of trace t operators on $\mathbf{C}^d$, we have by Fubini's theorem

$$\mathrm{vol}_{d^4}(K_{\mathbf{M}}) = \frac{1}{d} \int_{-d^2}^{d^2} \mathrm{vol}_{d^4-1}(K_{\mathbf{M}} \cap H_t)\, \mathrm{d}t.$$

By the Brunn–Minkowski inequality, the function under the integral is maximal when $t = 0$, and therefore

$$\mathrm{vol}_{d^4}(K_{\mathbf{M}}) \leqslant 2d\,\mathrm{vol}_{d^4-1}(K_{\mathbf{M}} \cap H_0).$$

It follows easily that $w(P_{H_0}K_{\mathbf{M}}) \geqslant \mathrm{vrad}(P_{H_0}K_{\mathbf{M}}) \geqslant \mathrm{vrad}(K_{\mathbf{M}} \cap H_0) \gtrsim \mathrm{vrad}(K_{\mathbf{M}}) \simeq w(K_{\mathbf{M}})$, the first inequality being the Urysohn inequality (Theorem 4.1.3 in Chapter 4, Section 4.1) and the last estimate being by Theorem 7.5.1. Once this is known, Theorem 7.2.2 is immediate from Proposition 7.5.3. □

Proof of Proposition 7.5.3. We first show the concentration estimate (7.16), using the following representation due to Życzkowski and Sommers [166]: ρ has the same distribution as $MM^\dagger$, where M is uniformly distributed on the Hilbert–Schmidt unit sphere of $d \times d$ complex matrices. We estimate the Lipschitz constant of the function $(M, N) \mapsto \|MM^\dagger - NN^\dagger\|_{\mathbf{M}}$, defined on the product of two such unit spheres, as follows:

$$\begin{aligned}\|M_1M_1^\dagger - N_1N_1^\dagger\|_{\mathbf{M}} - \|M_2M_2^\dagger - N_2N_2^\dagger\|_{\mathbf{M}} &\leqslant \|M_1M_1^\dagger - M_2M_2^\dagger\|_{\mathbf{M}} + \|N_1N_1^\dagger - N_2N_2^\dagger\|_{\mathbf{M}} \\ &\leqslant \|M_1M_1^\dagger - M_2M_2^\dagger\|_1 + \|N_1N_1^\dagger - N_2N_2^\dagger\|_1 \\ &\leqslant 2\left(\|M_1 - M_2\|_2 + \|N_1 - N_2\|_2\right).\end{aligned}$$

The second inequality is simply because $\|\cdot\|_{\mathbf{M}} \leqslant \|\cdot\|_1$, while the third inequality follows from Cauchy–Schwarz inequality (and the fact that M_1, M_2, N_1, N_2 have unit Hilbert–Schmidt norm), after noticing that $||AA^\dagger - BB^\dagger||_1 \leqslant ||(A-B)B^\dagger||_1 + ||A(A-B)^\dagger||_1$. We obtain as a consequence of Lemma 7.5.4 below, a variation on Lévy's lemma (recalled as Lemma 4.2.1 in Chapter 4, Section 4.2), the desired estimate

$$\mathbf{P}\left(|\|\rho - \sigma\|_{\mathbf{M}} - \mathbf{E}| > t\right) \leqslant 2\exp(-cd^2t^2).$$

In our application of Lemma 7.5.4, we identify the set of complex $d \times d$ matrices with $\mathbf{R}^n$ for $n = 2d^2$, and use $L = 2$.

Lemma 7.5.4. *Let S^{n-1} be the Euclidean unit sphere in $\mathbf{R}^n$, and equip $S^{n-1} \times S^{n-1}$ with the measure $\mu \otimes \mu$, where μ is the uniform probability measure on S^{n-1}, and with the metric $d((x,y),(x',y')) := \|x-x'\| + \|y-y'\|$. For any L-Lipschitz function $f : S^{n-1} \times S^{n-1} \to \mathbf{R}$ and any $t > 0$,*

$$\mathbf{P}(|f - \mathbf{E}\,f| > t) \leqslant 2\exp(-cnt^2/L^2),$$

$c > 0$ being a universal constant.

Lemma 7.5.4 can be deduced quickly from the usual Lévy lemma (see Lemma 4.2.1 in Chapter 4, Section 4.2) which quantifies the phenomenon of concentration of measure on the sphere. If we define, for any fixed $x \in S^{n-1}$, $E_x := \int_{S^{n-1}} f(x,y)\,\mathrm{d}\mu(y)$, we may apply Lévy's lemma to show that the function $y \in S^{n-1} \mapsto f(x,y)$ concentrates around its expectation E_x, and again Lévy's lemma to show that the function $x \in S^{n-1} \mapsto E_x$ (which is L-Lipschitz, as an average of L-Lipschitz functions) is also well-concentrated.

We now prove the first part of Proposition 7.5.3. Let Δ be a random matrix uniformly chosen from the Hilbert–Schmidt sphere in the hyperplane H_0, and ρ, σ be independent random states with uniform distribution. We claim that, from a very rough perspective, the spectra of $\rho - \sigma$ and $\Delta/\sqrt{d}$ look similar. More precisely, we have the following lemma.

Lemma 7.5.5. *Let ρ, σ be independent random states uniformly chosen from $\mathcal{D}(\mathbf{C}^d)$, and Δ be a random matrix uniformly chosen from the Hilbert–Schmidt sphere in the hyperplane H_0. Then with large probability, on the one hand*

$$\|\Delta\|_1 \simeq \sqrt{d},\ \|\Delta\|_2 = 1 \text{ and } \|\Delta\|_\infty \simeq 1/\sqrt{d},$$

while on the other hand

$$\|\rho - \sigma\|_1 \simeq 1,\ \|\rho - \sigma\|_2 \simeq 1/\sqrt{d} \text{ and } \|\rho - \sigma\|_\infty \simeq 1/d.$$

Moreover these statements hold in expectation: for example $\mathbf{E}\,\|\Delta\|_\infty \simeq 1/\sqrt{d}$ and $\mathbf{E}\,\|\rho - \sigma\|_\infty \simeq 1/d$.

In order to compare $\rho - \sigma$ with Δ, we rely on the following lemma. For $x = (x_1, \ldots, x_n) \in \mathbf{R}^n$, we denote $\|x\|_\infty = \max\{|x_i| \ :\ 1 \leqslant i \leqslant n\}$ and $\|x\|_1 = \sum_{i=1}^n |x_i|$.

Lemma 7.5.6. *Let $E = \{x \in \mathbf{R}^n \ : \ \sum_{i=1}^n x_i = 0\}$ and let $|||\cdot|||$ be a norm on E which is invariant under permutation of coordinates. Then, for any nonzero vectors $x, y \in E$, we have*

$$|||x||| \leqslant 2n \frac{\|x\|_\infty}{\|y\|_1} |||y|||. \tag{7.17}$$

Assuming both lemmas, we now complete the proof of Proposition 7.5.3. On the hyperplane $E \subset \mathbf{R}^d$ of vectors whose sum of coordinates is zero, we define a norm by

$$|||x||| := \int_{\mathcal{U}(d)} \|U \operatorname{diag}(x) U^\dagger\|_{\mathbf{M}} \, \mathrm{d}U,$$

where the integral is taken with respect to the Haar measure on the unitary group, and $\operatorname{diag}(x)$ denotes the diagonal matrix on $\mathbf{C}^d$ with diagonal elements equal to the coordinates of x. Note that $|||\cdot|||$ is obviously invariant under permutation of coordinates. Also, Δ has the same distribution as $U \operatorname{diag}(\operatorname{spec}(\Delta)) U^\dagger$, where U is a Haar-distributed unitary matrix independent from Δ and $\operatorname{spec}(A) \in \mathbf{R}^d$ denotes the spectrum of $A \in \mathscr{H}(\mathbf{C}^d)$ (the ordering of eigenvalues being irrelevant). The same holds for $\rho - \sigma$ instead of Δ, and it follows that

$$\mathbf{E}\, |||\operatorname{spec}(\Delta)||| = \mathbf{E}\, ||\Delta||_{\mathbf{M}} \quad \text{and} \quad \mathbf{E}\, |||\operatorname{spec}(\rho - \sigma)||| = \mathbf{E}\, ||\rho - \sigma||_{\mathbf{M}}.$$

Let us show that

$$\mathbf{E}\, ||\rho - \sigma||_{\mathbf{M}} \simeq \mathbf{E}\, \frac{1}{\sqrt{d}} ||\Delta||_{\mathbf{M}}. \tag{7.18}$$

We first prove the inequality $\lesssim$. Say that a vector $y \in E$ satisfies the condition $(\star)$ if $\|y\|_1 \geqslant c\sqrt{d}$, where we may choose the constant c such that the random vector $\operatorname{spec}(\Delta)$ satisfies the condition $(\star)$ with probability larger than $1/2$ (this is possible, as we check using Lemma 7.5.5). Now, by Lemma 7.5.6, for any $y \in E$ satisfying condition $(\star)$ and any $x \in E$, we have

$$|||x||| \lesssim \sqrt{d} \|x\|_\infty \cdot |||y|||.$$

We apply this inequality with $x = \operatorname{spec}(\rho - \sigma)$ and take expectation. This gives (using the statement about expectations in Lemma 7.5.5)

$$\mathbf{E}\, |||\rho - \sigma|||_{\mathbf{M}} \lesssim \frac{1}{\sqrt{d}} |||y|||.$$

This inequality is true for any $y \in E$ satisfying condition $(\star)$. Therefore,

$$\mathbf{E}\, ||\Delta||_{\mathbf{M}} = \mathbf{E}\, |||\operatorname{spec}(\Delta)||| \gtrsim \sqrt{d} \cdot \mathbf{P}\big(\operatorname{spec}(\Delta) \text{ satisfies condition } (\star)\big)\, \mathbf{E}\, \|\rho - \sigma\|_{\mathbf{M}} \simeq \sqrt{d}\, \mathbf{E}\, \|\rho - \sigma\|_{\mathbf{M}},$$

as needed. This proves one half of (7.18), and the reverse inequality is proved along the exact same lines. Finally, we note that

$$\mathbf{E}\, ||\Delta||_{\mathbf{M}} = w\left(P_{H_0} K_{\mathbf{M}}\right),$$

which, together with (7.18), shows (7.15), and concludes the proof. □

Proof of Lemma 7.5.5. This is folklore in random matrix theory, in fact much more precise results are known (for example, $\simeq$ can be replaced with $\sim$, with specific constants implicit in that notation). However, most of the literature focusses on slightly different random setups. Accordingly, we sketch an essentially self-contained elementary argument for completeness.

First of all, we observe that it is enough to prove the upper estimate for $\|\cdot\|_\infty$ and the lower estimate for $\|\cdot\|_2$. Indeed, the remaining upper estimates and the lower estimate for $\|\cdot\|_\infty$ follow then from the generally valid inequalities $\|\cdot\|_1 \leqslant \sqrt{d}\|\cdot\|_2 \leqslant d\|\cdot\|_\infty$, while the lower bound for $\|\cdot\|_1$ follows from $\|\cdot\|_2 \leqslant \|\cdot\|_1^{1/2} \|\cdot\|_\infty^{1/2}$.

The upper bound on $\|\cdot\|_\infty$ can be proved by a standard net argument. The lower bound on $\|\Delta\|_2$ is trivial, while for $\|\rho - \sigma\|_2$ we may proceed as follows. First, using concentration of measure in the form of Lemma 7.5.4, $\mathbf{E}\, \|\rho - \sigma\|_2$ is comparable to $\left(\mathbf{E}\, \|\rho - \sigma\|_2^2\right)^{1/2}$. Next, by Jensen inequality,

$$\mathbf{E}\, \|\rho - \sigma\|_2^2 \geqslant \mathbf{E}\, \|\rho - \operatorname{Id}/d\|_2^2.$$

Recalling that ρ can be represented as $MM^\dagger$, with M uniformly distributed on the Hilbert–Schmidt unit sphere of $d \times d$ complex matrices, the last quantity can be expanded as

$$\mathbf{E} \left\| \rho - \frac{\mathrm{Id}}{d} \right\|_2^2 = \mathbf{E} \operatorname{Tr} |M|^4 - \frac{1}{d}$$

and it can be checked by moments expansion that $\mathbf{E} \operatorname{Tr} |M|^4 \sim 2/d$. □

Proof of Lemma 7.5.6. Define $\alpha = 2n\|x\|_\infty / \|y\|_1$. By elementary properties of majorization (see e.g. Chapter II in [30]) it is enough to show that x is majorized by αy, i.e. that for every $1 \leqslant k \leqslant n$,

$$\sum_{i=1}^k x_i^\downarrow \leqslant \alpha \sum_{i=1}^k y_i^\downarrow,$$

where $(x_i^\downarrow)_{1 \leqslant i \leqslant n}, (y_i^\downarrow)_{1 \leqslant i \leqslant n}$ denote the non-increasing rearrangement of x, y. This follows from the inequalities

$$\frac{1}{\|x\|_\infty} \sum_{i=1}^k x_i^\downarrow \leqslant \min(k, n-k) \leqslant \frac{2n}{\|y\|_1} \sum_{i=1}^k y_i^\downarrow. \tag{7.19}$$

The left-hand inequality in (7.19) follows from the triangle inequality, once we have in mind that

$$x_1^\downarrow + \cdots + x_k^\downarrow = -(x_{k+1}^\downarrow + \cdots + x_n^\downarrow).$$

To prove the right-hand inequality in (7.19), note that the sum of positive coordinates of y and the sum of negative coordinates of y both equal $\|y\|_1/2$. Let ℓ be the number of positive coordinates of y. If $k \leqslant \ell$, then

$$y_1^\downarrow + \cdots + y_k^\downarrow \geqslant \frac{k}{\ell} \frac{\|y\|_1}{2} \geqslant \frac{k}{2n} \|y\|_1,$$

while if $k > \ell$, then

$$y_1^\downarrow + \cdots + y_k^\downarrow = -(y_{k+1}^\downarrow + \cdots + y_n^\downarrow) \geqslant \frac{n-k}{n-\ell} \frac{\|y\|_1}{2} \geqslant \frac{n-k}{2n} \|y\|_1.$$

And the proof of Lemma 7.5.6 is thus complete. □

7.6 Applications to quantum data hiding

7.6.1 Bipartite data hiding

As already mentioned, what Theorem 7.2.2 establishes is that generic bipartite states are data hiding for separable measurements but not for PPT measurements. This fact somehow counterbalances the usually cited constructions of data hiding schemes using Werner states (see e.g. [63, 64, 69] and [137, 130]). Werner states are indeed data hiding in the exact same way for both separable and PPT measurements.

Besides, results in the same vein as those from Theorem 7.2.2 but more specifically orientated towards applications to quantum data hiding may be quite directly written down. In fact, one often thinks of data hiding states as being orthogonal states, hence perfectly distinguishable by the suitable global measurement, that are nevertheless barely distinguishable by any local measurement. The following theorem provides a statement in that direction.

Theorem 7.6.1. *There are universal constants $C, c > 0$ such that the following holds. Given a dimension d, let E be a $d^2/2$-dimensional subspace of $\mathbf{C}^d \otimes \mathbf{C}^d$ (we assume without loss of generality that d is even). Let also $\rho = UP_E U^\dagger/(d^2/2)$ and $\sigma = UP_{E^\perp}U^\dagger/(d^2/2)$, where U is a Haar-distributed random unitary on $\mathbf{C}^d \otimes \mathbf{C}^d$. Then,*

$$\|\rho - \sigma\|_{\mathbf{ALL}} = 2,$$

whereas with high probability,

$$c \leqslant \|\rho - \sigma\|_{\mathbf{PPT}} \leqslant C \text{ and } \frac{c}{\sqrt{d}} \leqslant \|\rho - \sigma\|_{\mathbf{SEP}} \leqslant \frac{C}{\sqrt{d}}.$$

Here, "with high probability" means that the probability that one of the conclusions fails is less than $\exp(-c_0 d^3)$ for some constant $c_0 > 0$.

Proof. The first part of Theorem 7.6.1 is clear: the random states ρ and σ are orthogonal by construction, so that $\|\rho-\sigma\|_{\mathbf{ALL}} = \|\rho-\sigma\|_1 = 2$.

To prove the second part of Theorem 7.6.1, the only thing we have to show is that Proposition 7.5.3 also holds for the random states ρ and σ considered here.

Now, for any family $\mathbf{M}$ of POVMs on $\mathbf{C}^d\otimes\mathbf{C}^d$, the function $f : U \in \mathcal{U}(\mathbf{C}^d\otimes\mathbf{C}^d) \mapsto \left\|U(P_E - P_{E^\perp})U^\dagger/(d^2/2)\right\|_{\mathbf{M}}$ is $8/d$-Lipschitz. Indeed, by the same arguments as in the proof of (7.16),

$$\begin{aligned} f(U_1) - f(U_2) &\leqslant \frac{2}{d^2}\left(\|U_1P_EU_1^\dagger - U_2P_EU_2^\dagger\|_{\mathbf{M}} + \|U_1P_{E^\perp}U_1^\dagger - U_2P_{E^\perp}U_2^\dagger\|_{\mathbf{M}}\right) \\ &\leqslant \frac{2}{d}\left(\|U_1P_EU_1^\dagger - U_2P_EU_2^\dagger\|_2 + \|U_1P_{E^\perp}U_1^\dagger - U_2P_{E^\perp}U_2^\dagger\|_2\right) \\ &\leqslant \frac{4}{d}\left(\|U_1P_E - U_2P_E\|_2 + \|U_1P_{E^\perp} - U_2P_{E^\perp}\|_2\right) \\ &\leqslant \frac{8}{d}\|U_1 - U_2\|_2. \end{aligned}$$

And any L-Lipschitz function $g : \mathcal{U}(\mathbf{C}^n) \to \mathbf{R}$ satisfies the concentration estimate (see the Appendix in [138])

$$\forall\, t > 0,\ \mathbf{P}(|g - \mathbf{E}\, g| > t) \leqslant 2\exp(-cnt^2/L^2),\quad c \text{ being a universal constant.}$$

The function f thus satisfies $\mathbf{P}(|f - \mathbf{E}\, f| > t) \leqslant 2\exp(-cd^4t^2)$. So the concentration estimate (7.16) in Proposition 7.5.3 is in fact still true for the random states under consideration.

What is more, the results from Lemma 7.5.5 remain valid too because we here even have the equalities

$$\left\|\frac{2}{d^2}U(P_E - P_{E^\perp})U^\dagger\right\|_1 = 2,\quad \left\|\frac{2}{d^2}U(P_E - P_{E^\perp})U^\dagger\right\|_2 = \frac{2}{d}\quad\text{and}\quad \left\|\frac{2}{d^2}U(P_E - P_{E^\perp})U^\dagger\right\|_\infty = \frac{2}{d^2}.$$

So since the random Hermitians $U(P_E - P_{E^\perp})U^\dagger$ and $V\,\mathrm{diag}\left(\mathrm{spec}\left(U(P_E - P_{E^\perp})U^\dagger\right)\right)V^\dagger$ have the same distribution, for $U, V \in \mathcal{U}(\mathbf{C}^d\otimes\mathbf{C}^d)$ independent and Haar-distributed, one may apply Lemma 7.5.6 to conclude that the expectation estimate (7.15) in Proposition 7.5.3 is in fact still true too for the random states under consideration. □

In words, Theorem 7.6.1 stipulates the following. Picking a subspace E uniformly at random from the set of $d^2/2$-dimensional subspaces of $\mathbf{C}^d\otimes\mathbf{C}^d$, and then considering the states $\rho = P_E/(d^2/2)$ and $\sigma = P_{E^\perp}/(d^2/2)$, one gets examples of states which are perfectly distinguishable by some global measurement and which are with high probability data-hiding for separable measurements but not data-hiding for PPT measurements.

Remark 7.6.2. *Let us come back on the example of the symmetric state π_s and the antisymmetric state π_a on $\mathbf{C}^d\otimes\mathbf{C}^d$. They satisfy (see e.g. [64])*

$$\|\pi_s - \pi_a\|_{\mathbf{SEP}} = \|\pi_s - \pi_a\|_{\mathbf{PPT}} = \frac{4}{d+1} = \frac{2}{d+1}\|\pi_s - \pi_a\|_{\mathbf{ALL}}. \tag{7.20}$$

They are consequently "exceptional" data hiding states for two reasons. First, as mentioned before, because they are equally PPT and SEP data hiding. And second because they are "more" data hiding than generic states: their SEP norm is of order $1/d \ll 1/\sqrt{d}$, hence almost reaching the known lower-bound valid for any states ρ,σ on $\mathbf{C}^d\otimes\mathbf{C}^d$ (see e.g. [137]) namely $\|\rho-\sigma\|_{\mathbf{SEP}} \geqslant 2\|\rho-\sigma\|_{\mathbf{ALL}}/d$.

7.6.2 Multipartite vs bipartite data hiding

In Theorem 7.5.1, we focussed on the bipartite case $\mathrm{H} = (\mathbf{C}^d)^{\otimes 2}$ for the sake of clarity. However, generalizations to the general k-partite case $\mathrm{H} = (\mathbf{C}^d)^{\otimes k}$ are quite straightforward, at least in the situation where the high-dimensional composite system of interest is made of a "small" number of "large" subsystems (i.e. k is fixed and d tends to infinity).

Let us denote by $\mathbf{PPT}_{d,k}$ and $\mathbf{SEP}_{d,k}$ the sets of respectively k-PPT and k-separable POVMs on $(\mathbf{C}^d)^{\otimes k}$. On the one hand, an iteration of the Milman–Pajor inequality (Corollary 4.1.7 in Chapter 4, Section 4.1) leads to the estimate

$$c^{2^k}d^{k/2} \leqslant \mathrm{vrad}(K_{\mathbf{PPT}_{d,k}}) \leqslant w(K_{\mathbf{PPT}_{d,k}}) \leqslant Cd^{k/2},$$

for some constants $c, C > 0$ depending neither on k nor on d.

On the other hand, the generalization of Theorem 7.8.4 in Appendix 7.8 to the set $\mathcal{S}_{d,k}$ of k-separable states on $(\mathbf{C}^d)^{\otimes k}$ is known, namely (see [16], Theorem 1)

$$\frac{c^k}{d^{k-1/2}} \leqslant \mathrm{vrad}(\mathcal{S}_{d,k}) \leqslant w(\mathcal{S}_{d,k}) \leqslant C\frac{\sqrt{k\log k}}{d^{k-1/2}},$$

and implies that

$$c^k d^{1/2} \leqslant \mathrm{vrad}(K_{\mathbf{SEP}_{d,k}}) \leqslant w(K_{\mathbf{SEP}_{d,k}}) \leqslant C\sqrt{k\log k}d^{1/2},$$

for some constants $c, C > 0$ depending neither on k nor on d.

A multipartite analogue of Theorem 7.2.2 can then be derived, following the exact same lines of proof.

Theorem 7.6.3. *There exist constants $c_k, C_k > 0$, depending only on k, such that the following holds. Given a dimension d, let ρ and σ be random states, independent and uniformly distributed on the set of states on $(\mathbf{C}^d)^{\otimes k}$. Then, with high probability,*

$$c_k \leqslant \|\rho-\sigma\|_{\mathbf{PPT}_{d,k}} \leqslant \|\rho-\sigma\|_{\mathbf{ALL}} \leqslant C_k,$$

$$\frac{c_k}{\sqrt{d^{k-1}}} \leqslant \|\rho-\sigma\|_{\mathbf{SEP}_{d,k}} \leqslant \frac{C_k}{\sqrt{d^{k-1}}}.$$

This means that, forgetting about the dependence on k and only focussing on the one on d, for "typical" states ρ, σ on $(\mathbf{C}^d)^{\otimes k}$, $\|\rho-\sigma\|_{\mathbf{PPT}_{d,k}}$ is of order 1, like $\|\rho-\sigma\|_{\mathbf{ALL}}$, while $\|\rho-\sigma\|_{\mathbf{SEP}_{d,k}}$ is of order $1/\sqrt{d^{k-1}}$.

In this multipartite setting, another quite natural question is the one of finding states that local observers can poorly distinguish if they remain alone but that they can distinguish substantially better though by gathering into any possible two groups. This type of problem was especially studied in [69]. Here is another result in that direction.

Define $\mathbf{bi{-}SEP}_{d,k}$ as the set of POVMs on $(\mathbf{C}^d)^{\otimes k}$ which are biseparable across any bipartition of $(\mathbf{C}^d)^{\otimes k}$. It may then be shown that for random states ρ, σ, independent and uniformly distributed on the set of states on $(\mathbf{C}^d)^{\otimes k}$, with high probability, $\|\rho-\sigma\|_{\mathbf{bi-SEP}_{d,k}} \simeq d^{-k/4}$ (whereas $\|\rho-\sigma\|_{\mathbf{SEP}_{d,k}} \simeq d^{-(k-1)/2}$ by Theorem 7.6.3). This means that on $(\mathbf{C}^d)^{\otimes k}$, with $k > 2$ fixed, restricting to POVMs which are biseparable across every bipartition is roughly the same as restricting to POVMs which are biseparable across one bipartition, whereas imposing k-separability is a much tougher constraint that implies a dimensional loss in the distinguishing ability. The reader is referred to Chapter 8, Section 8.5, for related analyses.

Remark 7.6.4. *This result might not be as strong as one could hope for. It only shows that $\|\cdot\|_{\mathbf{bi-SEP}_{d,k}}$ typically vanishes slower than $\|\cdot\|_{\mathbf{SEP}_{d,k}}$ when the local dimension d grows, but it does not provide examples of states ρ, σ on $(\mathbf{C}^d)^{\otimes k}$ for which $\|\rho-\sigma\|_{\mathbf{bi-SEP}_{d,k}}$ would be of order 1 while $\|\rho-\sigma\|_{\mathbf{SEP}_{d,k}}$ would tend to zero.*

7.7 Miscellaneous remarks and questions

7.7.1 On the complexity of the classes of separable and PPT POVMs

Having at hand the estimates on the mean width of $K_{\mathbf{SEP}}$ (or $K_{\mathbf{LOCC}}$) and $K_{\mathbf{PPT}}$ provided by Theorem 7.5.1, one may follow the exact same lines as in the proof of Theorem 7.2.1 to show that on $\mathbf{C}^d \otimes \mathbf{C}^d$, $\exp(\Theta(d^4))$ different POVMs are necessary and sufficient to approximate the class **PPT**. For the class **SEP** (or **LOCC**), we lack a complete answer since the same arguments show that the minimal number of POVMs is between $\exp(\Omega(d^3))$ and $\exp(O(d^4))$.

7.7.2 What is the typical performance of the class LO?

While Theorem 7.2.4 shows that the gap between the classes **LO** and **LOCC** may be unbounded, we do not know if this situation is typical or not. Note that asking whether the norms $\|\cdot\|_{\mathbf{LO}}$ and $\|\cdot\|_{\mathbf{LOCC}}$ are comparable in a typical direction is more or less equivalent to asking whether the ratio $\mathrm{vrad}(K_{\mathbf{LOCC}})/\mathrm{vrad}(K_{\mathbf{LO}})$ is bounded as the dimension increases. It is thus a pure convex geometry question, which now only requires to understand how the convex body $K_{\mathbf{LO}}$ looks like.

7.7.3 Can the gap LOCC/SEP be unbounded?

Or conversely, does there exist an absolute constant c such that the inequality $\|\cdot\|_{\mathbf{LOCC}} \geqslant c\|\cdot\|_{\mathbf{SEP}}$ holds for any dimension? Also, we showed the existence of unbounded gaps in the two steps of the hierarchy $\mathbf{LOCC}^{(\mathbf{0})}/\mathbf{LOCC}^{(\mathbf{1})}/\mathbf{LOCC}^{(\mathbf{2})}$. More generally, is it possible to find, for any $r \in \mathbf{N}$, states which can be distinguished very poorly by r rounds of local measurements and classical communication but very well if one extra round is allowed? The main difficulty in generalizing Theorems 7.4.4 and 7.4.5 to the case $r \geqslant 2$ lies in understanding the disturbance induced on the states to be distinguished by all successive rounds of measurements.

7.7.4 Typical value of other locally restricted distance measures

We solved the issue of determining, for several classes of measurements $\mathbf{M}$, what is the typical value of the measured trace distance $\|\rho - \sigma\|_{\mathbf{M}}$ between two states ρ, σ. Several other "filtered through measurements" distances between ρ and σ can be defined in a completely analogous way, such as e.g. the measured fidelity distance $F_{\mathbf{M}}(\rho, \sigma)$ or the measured relative entropy distance $D_{\mathbf{M}}(\rho\|\sigma)$ (see e.g. [147]). These quantities are all closely related to one another by well-known inequalities. Our statements can thus be straightforwardly translated into statements on the typical value of $F_{\mathbf{M}}(\rho, \sigma)$ or $D_{\mathbf{M}}(\rho\|\sigma)$. Besides, such restricted distance measures have already found a tremendous amount of applications in quantum information theory (see e.g. [24] or [136] for two very recent ones, in two quite different topics, and also Chapter 10 of the present manuscript). Understanding better what is their generic scaling (and ultimately the one of their regularised versions) is therefore of prime interest, amongst other, to assess how optimal are the bounds where they appear, what is the efficiency of the quantum information processing protocols where they are involved etc.

7.7.5 Locally restricted measurements on a multipartite quantum system

There are at least two ways for a multipartite system such as $(\mathbf{C}^d)^{\otimes k}$ to be high-dimensional: either with k fixed and d large (few large subsystems) or k large and d fixed (many small subsystems). Theorem 7.6.3 tells us what is the typical discriminating power of k-PPT and k-separable POVMs, but in the first setting only. The extension to the case of many small subsystems seems a challenging problem.

7.8 Appendix: Volume estimates for some Schatten norm unit balls and related convex bodies

In this appendix we gather estimates on the mean width and the volume radius of "standard" sets, which are used in our proofs. These concepts from classical convex geometry (and several others which we alude to here) were introduced in Chapter 4, Section 4.1, to which the reader is referred for further details. We also recall the two following notation specified in Chapter 1, Section 1.3, for the unit balls associated to Schatten norms in the space of self-adjoint operators on $\mathbf{C}^d$:

$$B_1(\mathbf{C}^d) = \{A \in \mathscr{H}(\mathbf{C}^d) \ : \ \|A\|_1 \leqslant 1\},$$

$$B_\infty(\mathbf{C}^d) = \{A \in \mathscr{H}(\mathbf{C}^d) \ : \ \|A\|_\infty \leqslant 1\} = [-\mathrm{Id}, \mathrm{Id}].$$

Moreover, given symmetric convex bodies $K \subset \mathbf{R}^n$ and $K' \subset \mathbf{R}^{n'}$, their projective tensor product is the convex body in $\mathbf{R}^n \otimes \mathbf{R}^{n'}$ defined as

$$K \hat{\otimes} K' = \mathrm{conv}\{x \otimes x' \ : \ x \in K, x' \in K'\} \subset \mathbf{R}^n \otimes \mathbf{R}^{n'}$$

Theorem 7.8.1. *We have*

$$\mathrm{vrad}(B_\infty(\mathbf{C}^d)) \simeq w(B_\infty(\mathbf{C}^d)) \simeq \sqrt{d}.$$

$$\mathrm{vrad}(B_1(\mathbf{C}^d)) \simeq w(B_1(\mathbf{C}^d)) \simeq \frac{1}{\sqrt{d}}.$$

Proof. The estimates on the mean width follow from the semicircle law. Indeed, the standard Gaussian vector in the space of self-adjoint operators on $\mathbf{C}^d$ is exactly a $d \times d$ GUE matrix G, and therefore (see [4], Chapter 2, for a proof of these asymptotic estimates)

$$w_G(B_\infty(\mathbf{C}^d)) = \mathbf{E}\,\|G\|_1 = d^{3/2}\left(\int_{-2}^{2} |x|\frac{\sqrt{4-x^2}}{2\pi}\,\mathrm{d}x + o(1)\right) = d^{3/2}\left(\frac{8}{3\pi} + o(1)\right),$$
$$w_G(B_1(\mathbf{C}^d)) = \mathbf{E}\,\|G\|_\infty = d^{1/2}(2 + o(1)).$$

Hence, setting $\gamma(d) = \mathbf{E}\,\|G\|_2$, which is known to satisfy $\gamma(d) \sim d$ (see again [4], Chapter 2, for a proof of this asymptotic estimate), we get

$$w(B_\infty(\mathbf{C}^d)) = \frac{1}{\gamma(d)} w_G(B_\infty(\mathbf{C}^d)) \sim \frac{8\sqrt{d}}{3\pi},$$
$$w(B_1(\mathbf{C}^d)) = \frac{1}{\gamma(d)} w_G(B_1(\mathbf{C}^d)) \sim \frac{2}{\sqrt{d}}.$$

Since $B_1(\mathbf{C}^d)$ and $B_\infty(\mathbf{C}^d)$ are polar to each other and symmetric, the Santaló and reverse Santaló inequalities (see [160] and [33]) yield

$$\operatorname{vrad}(B_\infty(\mathbf{C}^d))\operatorname{vrad}(B_1(\mathbf{C}^d)) \simeq 1.$$

If we then use the Urysohn inequality (see Theorem 4.1.3 in Chapter 4, Section 4.1), we obtain

$$1 \simeq \operatorname{vrad}(B_\infty(\mathbf{C}^d))\operatorname{vrad}(B_1(\mathbf{C}^d)) \leqslant w(B_\infty(\mathbf{C}^d))w(B_1(\mathbf{C}^d)) \simeq \sqrt{d}\,\frac{1}{\sqrt{d}} \simeq 1,$$

and therefore all these inequalities are sharp up to a multiplicative constant. □

We also need volume estimates on projective tensor products of Schatten norm unit balls.

Theorem 7.8.2. *We have the following estimates*

$$\operatorname{vrad}\big(B_1(\mathbf{C}^d)\hat{\otimes}B_\infty(\mathbf{C}^d)\big) \simeq w\big(B_1(\mathbf{C}^d)\hat{\otimes}B_\infty(\mathbf{C}^d)\big) \simeq \frac{1}{\sqrt{d}}.$$

A very similar proof shows that the estimates of Theorem 7.8.2 are also valid when we consider the full complex Schatten classes, without the self-adjoint constraint. The question of estimating the volume radius of projective tensor product of Schatten classes has been considered in [59], where the question is answered (in a general setting) only up to a factor $\log d$.

Proof. An upper bound on the mean width can be obtained by a discretization argument, which we just sketch since only the lower bound will we used. There is a polytope P with $\exp(Cd)$ vertices such that $B_1(\mathbf{C}^d) \subset P \subset 2B_1(\mathbf{C}^d)$, and a polytope Q with $\exp(Cd^2)$ vertices such that $B_\infty(\mathbf{C}^d) \subset Q \subset 2B_\infty(\mathbf{C}^d)$. The polytope $P\hat{\otimes}Q$ satisfies

$$B_1(\mathbf{C}^d)\hat{\otimes}B_\infty(\mathbf{C}^d) \subset P\hat{\otimes}Q \subset 4B_1(\mathbf{C}^d)\hat{\otimes}B_\infty(\mathbf{C}^d).$$

The polytope $P\hat{\otimes}Q$ is the convex hull of $\exp(C'd^2)$ points with Hilbert–Schmidt norm at most $4\sqrt{d}$. Using standard bounds for the mean width of polytopes (see Lemma 4.1.4 in Chapter 4, Section 4.1) gives the desired estimate $w\big(B_1(\mathbf{C}^d)\hat{\otimes}B_\infty(\mathbf{C}^d)\big) \lesssim 1/\sqrt{d}$.

We now give a lower bound on the volume radius. We denote by $B_1^n \subset \mathbf{R}^n$ the unit ball for $\|\cdot\|_1$ in $\mathbf{R}^n$. We have the following formula.

Lemma 7.8.3. *Let m, n be integers and $K \subset \mathbf{R}^m$ be a symmetric convex body. Then*

$$\operatorname{vol}(B_1^n\hat{\otimes}K) = \frac{(m!)^n}{(mn)!}\operatorname{vol}(K)^n.$$

Consequently,

$$\operatorname{vrad}(B_1^n\hat{\otimes}K) \simeq \frac{1}{\sqrt{n}}\operatorname{vrad}(K).$$

Proof. If $\{e_1, \dots, e_n\}$ denotes the canonical basis of $\mathbf{R}^n$, we have, for any $x_1, \dots, x_n \in \mathbf{R}^m$,

$$\left\| \sum_{i=1}^n e_i \otimes x_i \right\|_{B_1^n \hat{\otimes} K} = \sum_{i=1}^n \|x_i\|_K.$$

So Lemma 7.8.3 follows easily from the formula below, valid for any integer p and any symmetric convex body $L \subset \mathbf{R}^p$,

$$\mathrm{vol}(L) = \frac{1}{p!} \int_{\mathbf{R}^p} \exp(-\|x\|_L)\, \mathrm{d}x. \tag{7.21}$$

Equation (7.21) itself may be obtained by the following chain of equalities

$$\int_{\mathbf{R}^p} e^{-\|x\|_L}\, \mathrm{d}x = \int_{\mathbf{R}^p} \int_{\|x\|_L}^{+\infty} e^{-t}\, \mathrm{d}t\, \mathrm{d}x = \int_0^{+\infty} \int_{\{\|x\|_L < t\}} e^{-t}\, \mathrm{d}x\, \mathrm{d}t = \int_0^{+\infty} e^{-t}\, \mathrm{vol}(tL)\, \mathrm{d}t = \mathrm{vol}(L) p!,$$

the last equality being because $\int_0^{+\infty} t^p e^{-t}\, \mathrm{d}t = p!$. □

Denote by $\{|j\rangle \ : \ 1 \leqslant j \leqslant d\}$ an orthonormal basis of $\mathbf{C}^d$. The family

$$\{|j\rangle\langle j| \ : \ 1 \leqslant j \leqslant d\} \cup \left\{ \frac{1}{\sqrt{2}} (|j\rangle\langle k| + |k\rangle\langle j|) \ : \ 1 \leqslant j < k \leqslant d \right\} \cup \left\{ \frac{i}{\sqrt{2}} (|j\rangle\langle k| - |k\rangle\langle j|) \ : \ 1 \leqslant j < k \leqslant d \right\}$$

is an orthonormal basis of $\mathscr{H}(\mathbf{C}^d)$ whose elements live in $\sqrt{2} B_1(\mathbf{C}^d)$. It follows that

$$\mathrm{vrad}\left(B_1(\mathbf{C}^d) \hat{\otimes} B_\infty(\mathbf{C}^d)\right) \geqslant \frac{1}{\sqrt{2}}\, \mathrm{vrad}\left(B_1^{d^2} \hat{\otimes} B_\infty(\mathbf{C}^d)\right) \gtrsim \frac{1}{d}\, \mathrm{vrad}(B_\infty(\mathbf{C}^d)),$$

the last estimate being a consequence of Lemma 7.8.3.

Using Theorem 7.8.1 one may thus conclude that $\mathrm{vrad}\left(B_1(\mathbf{C}^d) \hat{\otimes} B_\infty(\mathbf{C}^d)\right) \gtrsim 1/\sqrt{d}$. □

We also need a result on the volume radius and the mean width of the set of separable states, which is taken from [16], Theorem 1.

Theorem 7.8.4. *On $\mathbf{C}^d \otimes \mathbf{C}^d$, denoting by $\mathcal{S}$ the set of separable states, we have*

$$d^{-3/2} \simeq \mathrm{vrad}(\mathcal{S}) \leqslant w(\mathcal{S}) \simeq d^{-3/2}.$$

Chapter 8

Relaxations of separability in multipartite quantum systems

Based on "Relaxations of separability in multipartite systems: semidefinite programs, witnesses and volumes", in collaboration with O. Gühne, M. Huber and R. Sengupta [87].

While entanglement is believed to be an important ingredient in understanding quantum many-body physics, the complexity of its characterization scales very unfavorably with the size of the system. Finding super-sets of the set of separable states that admit a simpler description has proven to be a fruitful approach in the bipartite setting. In this chapter we discuss a systematic way of characterizing multiparticle entanglement via various relaxations. We furthermore describe an operational witness construction arising from such relaxations that is capable of detecting every entangled state. Finally, we also derive an analytic upper bound on the volume of biseparable states and show that the volume of the states with a positive partial transpose for any split rapidly outgrows this volume. This proves that simple semidefinite relaxations in the multiparticle case cannot be an equally good approximation for any scenario.

8.1 Introduction

Without a doubt entanglement can be considered one of the most important concepts in quantum physics, clearly distinguishing quantum systems from classical ones. It can be harnessed to enable novel ways of processing quantum information in numerous ways, from communication to computation. Many of these operational tasks require an operational detection or even quantification of this indispensable resource. While in bipartite systems of low dimensions this can be achieved in an efficient way, the complexity of the characterization of entangled states makes a complete and computable framework of entanglement detection impossible in high dimensions and thus also for multipartite systems [106, 89].

A possible way for deriving statements on the presence of entanglement is to discard the actual complex structure of the border between separable and entangled states and try to find good approximations that admit a more amenable description. In the bipartite case positive maps play a central role in such approximations: all separable states remain positive semidefinite under application of a positive, yet not completely positive map to one of its subsystems. The most well-known example of such a map is the partial transposition: this map generally changes the eigenvalues of a matrix, but separable states have a positive partial transpose. The approach of positive maps allows for a characterization of super-sets of the set of separable states using techniques from semidefinite programming, and it nevertheless captures the whole structure: a state remains positive under *all* positive maps if and only if the state is indeed separable [106, 89]. The reader is referred to Chapter 2, Section 2.4, for extra information.

In order to gauge the efficiency of certain maps for characterizing entanglement, one of the most relevant issues is how the volume of states that remain positive under the map in question compares to the volume of separable states [110]. The sobering and non-surprising answer from bipartite systems can be gained from convex geometry considerations and shows that for all known maps the states that remain positive are most likely entangled in high dimensions [25]. For small dimensions, however, a given positive map can detect a large fraction of entangled states. This is, for instance, true for the partial transposition, which delivers a necessary and sufficient criterion for 2×2 and 2×3 systems.

In multipartite systems the characterization of entanglement constitutes an even greater challenge. Since partial separability of multipartite states can no longer be defined as a purely bipartite concept, the application of positive maps to subsystems alone can reveal little more than entanglement across a fixed partition of the multipartite state. Nevertheless, recently several works succeeded in defining suitable mixtures of positive maps, which can be used to develop strong criteria for genuine multiparticle entanglement [88, 112].

In this chapter we first develop a framework that allows for the semidefinite relaxation of partially separable states, opening the possibility for harnessing well developed techniques based on positive maps to detect genuine multipartite entanglement (an approach which has already been shown to yield useful results with different relaxations in [57, 37, 38]). We achieve this goal by first formally defining semidefinite relaxations of partially separable states using positive maps.

Due to the formulation as semidefinite programs these constructions yield versatile criteria for detecting multipartite entanglement in low dimensional systems. To unlock these powerful techniques for more complex quantum states we proceed to discuss a recently introduced program of lifting bipartite witnesses [112]. We prove that it is always possible to exploit witnesses that only reveal bipartite entanglement in order to construct witnesses for genuine multipartite entanglement. This facilitates this notoriously hard problem, and we showcase this technique with some exemplary multipartite entangled states.

In a second step, we ask which fraction of genuinely multipartite entangled states can be detected with such relaxation methods. We prove an upper bound on the volume of the set of biseparable states, and a lower bound on the volume of a set of states that can never be detected with relaxation methods based on the partial transposition. For large dimensions, both values deviate significantly. This shows that while the relaxation approaches are strong for small systems, they fail to deliver a good approximation in the general case.

8.2 Characterizing relaxations of separability with semidefinite programs

The most straightforward relaxation of separability in multipartite systems is again given by positive maps. Trying to justify this assertion is the object of the current section.

Let us start with some basic definitions and notation (see e.g. [89, 70] for a general review on these notions, and Chapter 2, Section 2.4, of this manuscript for a brief recap). On a multipartite system, given a bipartition $\{I|I^c\}$ of the subsystems, we denote by $\mathcal{S}_I$ the set of states which are biseparable across this cut, i.e.

$$\rho \in \mathcal{S}_I \text{ if } \rho = \sum_x q_x |\phi_I^{(x)}\rangle\langle\phi_I^{(x)}| \otimes |\phi_{I^c}^{(x)}\rangle\langle\phi_{I^c}^{(x)}|,$$

where $\{q_x\}_x$ is a convex combination, and for each x, $|\phi_I^{(x)}\rangle, |\phi_{I^c}^{(x)}\rangle$ are pure states on the subsystems in I, I^c respectively. We then define the set $\mathcal{S}_{(2)}$ of biseparable states as being the convex hull of $\{\mathcal{S}_I\}_I$, i.e.

$$\rho \in \mathcal{S}_{(2)} \text{ if } \rho = \sum_I p_I \sigma_I,$$

where $\{p_I\}_I$ is a convex combination, and for each I, $\sigma_I \in \mathcal{S}_I$. If a state is not biseparable it is called *genuinely multipartite entangled* (GME).

This definition admits a simple relaxation with a positive semidefinite characterization. Given, for each bipartition $\{I|I^c\}$, a positive map $\mathcal{N}_I$, acting on the substystems in I, we define the set $\mathcal{R}_{\{\mathcal{N}_I\}_I}$ of $\{\mathcal{N}_I\}_I$-relaxation of $\mathcal{S}_{(2)}$ by

$$\rho \in \mathcal{R}_{\{\mathcal{N}_I\}_I} \text{ if } \rho = \sum_I p_I \sigma_{\mathcal{N}_I}, \tag{8.1}$$

where $\{p_I\}_I$ is a convex combination and for each I, $\mathcal{N}_I \otimes \mathcal{I}d_{I^c}(\sigma_{\mathcal{N}_I}) \geqslant 0$.

Such a relaxation carries the operational advantage that it can be approached via semidefinite programming (SDP). Besides, this definition can easily be extended to ℓ-separable states by applying different maps to the induced partitions. As we are however mainly interested in characterizing the strongest form of multipartite entanglement we focus here on the distinction between biseparable states and genuinely multipartite entangled ones.

For instance, these relaxations can be particularly useful when optimizing convex functions over the set of biseparable states. Indeed, for any function f, we trivially have that, for any set of positive maps $\{\mathcal{N}_I\}_I$,

$$\min_{\sigma\in\mathcal{S}_{(2)}} f(\sigma) \geqslant \min_{\sigma\in\mathcal{R}_{\{\mathcal{N}_I\}_I}} f(\sigma). \tag{8.2}$$

In particular, for any convex function f, equation (8.2) provides a relaxation of the optimization of f over $\mathcal{S}_{(2)}$ which can be cast as an SDP. One of the most straightforward applications of such a strategy is to testing whether a given density matrix ρ is indeed biseparable, i.e. applying it e.g. to $f : \sigma \mapsto \left(\mathrm{Tr}\left[(\rho-\sigma)^2\right]\right)^{1/2}$. It yields in that case the equivalence

$$\forall\ \{\mathcal{N}_I\}_I,\ \min_{\sigma\in\mathcal{R}_{\{\mathcal{N}_I\}_I}} \mathrm{Tr}\big[\sigma(\sigma-2\rho)\big] \leqslant \mathrm{Tr}\left[\rho^2\right] \Leftrightarrow\ \rho\in\mathcal{S}_{(2)}.$$

While this defines, in theory, a necessary and sufficient program for deciding whether a given state is biseparable, checking all possible sets of positive maps is of course not feasible. However even making one particular choice, such as the transposition map for instance, has already proven to yield very strong witnesses, and through the dual of the program one can additionally often extract analytical constructions for important classes of states [88].

Let us specify a bit what we mean in the simplest tripartite case. Given a tripartite state ρ and a positive map $\mathcal{N}$ acting on one subsystem, the condition $\rho\in\mathcal{R}_{\{\mathcal{N}_1,\mathcal{N}_2,\mathcal{N}_3\}}$ is equivalent to the following SDP yielding a non-negative value:

$$\max s \text{ s.t. } \sigma_1,\sigma_2,\sigma_3 \geqslant s\mathrm{Id},\ \rho=\sigma_1+\sigma_2+\sigma_3,\ \mathcal{N}_1\otimes\mathcal{Id}_{23}(\sigma_1),\mathcal{N}_2\otimes\mathcal{Id}_{13}(\sigma_2),\mathcal{N}_3\otimes\mathcal{Id}_{12}(\sigma_3)\geqslant s\mathrm{Id}.$$

While not being quantitative as the above program it can be easily implemented in MATLAB, using the packages YALMIP [135] and the SEDUMI solver [169]. What is more, if an $s\geqslant 0$ is found, the program also returns a feasible $\sigma_1,\sigma_2,\sigma_3$ achieving the maximum, and can thus be used to find actual $\mathcal{N}$-positive decompositions of ρ.

To test the further prospects of such SDP approach, we have first programmed it choosing either the transposition map $\mathcal{T}$ (initially introduced in [146] and [108]) or the Choi map $\mathcal{C}$ (initially introduced in [47]) as positive map (see Chapter 2, Section 2.4, for precise definitions). We have applied them to the family of states $\rho(\{\lambda_1,\lambda_2,\lambda_3\})$ introduced in [112], Example 2. It was shown there that the state $\rho(\lambda):=\rho(\{\lambda,\lambda,\lambda\})$ is GME for $0<\lambda<1/3$. Testing the two programs on this family of states returns a positive under partial transposition map (PPT) or positive under partial Choi map (PPC) decomposition for $1/3\leqslant\lambda<1$, thus showing that on the set of PPT or PPC mixtures the witness presented in [112] is weakly optimal. It furthermore proves that for $1/3\leqslant\lambda<1$ the state can be decomposed into states which are either PPT or PPC.

We have then extended the program to demand simultaneous positivity under the transposition and the Choi maps. That is formally, we looked for

$$\max s \text{ s.t. } \sigma_1,\sigma_2,\sigma_3 \geqslant s\mathrm{Id},\ \rho=\sigma_1+\sigma_2+\sigma_3,\ \begin{cases}\mathcal{C}_1\otimes\mathcal{Id}_{23}(\sigma_1),\mathcal{C}_2\otimes\mathcal{Id}_{13}(\sigma_2),\mathcal{C}_3\otimes\mathcal{Id}_{12}(\sigma_3)\geqslant s\mathrm{Id}\\ \mathcal{T}_1\otimes\mathcal{Id}_{23}(\sigma_1),\mathcal{T}_2\otimes\mathcal{Id}_{13}(\sigma_2),\mathcal{T}_3\otimes\mathcal{Id}_{12}(\sigma_3)\geqslant s\mathrm{Id}\end{cases}.$$

With this program, one can check that, in fact, for every $0<\lambda<1$, $\rho(\lambda)$ is GME. This is because the program shows that, even in the range $1/3\leqslant\lambda\leqslant 1$, $\rho(\lambda)$ cannot be decomposed into a mixture of states which are both PPT and PPC.

This example showcases the versatility of this approach in small dimensions. Indeed, relaxing being biseparable to just being a mixture of states which are positive under partial application of two positive maps already proves quite fruitful. Let us mention though that, in high dimensions, making the SDP so slightly more constraining, by imposing positivity under two maps instead of one, is not expected to drastically improve its efficiency as in this particular case (see [15] for a mathematically precise formulation of this assertion).

8.3 Constructing multiparticle witnesses from bipartite witnesses

While the semidefinite program presented before technically gives sufficient criteria for deciding biseparability, it becomes quickly intractable beyond a few qubits. There are, however, frequent situations in which one can use some additional knowledge to facilitate witness constructions. Genuinely multipartite entangled states of course also have to be entangled across every bipartition of the system. And since the construction of bipartite entanglement witnesses can be a rather straightforward affair (e.g. through positive maps) one can ask whether there is a possibility to construct multipartite entanglement witnesses from a collection of bipartite operators.

8.3.1 A systematic construction for lifting bipartite witnesses

In [112] a general witness construction method was introduced, which enables the construction of multipartite entanglement witnesses from a set of bipartite witnesses across every possible bipartition. Such a problem can be formalized as follows.

Given, for each bipartition $\{I|I^c\}$, a witness W_I, i.e. a self-adjoint operator such that $\mathrm{Tr}(W_I\sigma_I) \geqslant 0$ for all $\sigma_I \in \mathcal{S}_I$, we are looking for a self-adjoint operator W_{GME} with the following property

$$\forall\ I,\ W_{GME} \geqslant W_I. \tag{8.3}$$

The construction of [112] is one particular instance of the following general method. Let Q be some operator, and set, for each I,

$$T_I = Q - W_I. \tag{8.4}$$

Define next W_{GME} as

$$W_{GME} = Q + \sum_I \left[T_I\right]_+, \tag{8.5}$$

where, for any self-adjoint operator A, we denote by $[A]_+$ the projection of A onto the positive semidefinite cone. It is then easy to see that condition (8.3) holds for every I. So of course, the crucial issue here is first of all the one of the generality of such a construction: given a set of bipartite entanglement witnesses $\{W_I\}_I$ for some genuinely multipartite entangles state ρ_{GME}, does there always exist a Q such that W_{GME} as defined by equation (8.5) is a genuinely multipartite entanglement witness for ρ_{GME}? Second, one can also ask the question of the optimal choice of Q (in a sense to be defined).

Using a special choice for Q, it was shown in [112] that there exist genuinely multipartite entangled states and a set of bipartite witnesses for which the expectation value of W_{GME} is negative, proving that this construction can indeed succeed in generating witnesses for multipartite entanglement. Here we begin with proving the following generality result.

Theorem 8.3.1. *For every genuinely multipartite entangled state ρ_{GME}, there exists a set of weakly optimal bipartite entanglement witnesses $\{W_I\}_I$ such that*

$$\mathrm{Tr}\left(\rho_{GME}\left(Q + \sum_I \left[T_I\right]_+\right)\right) < 0,$$

where the operators Q and $\{T_I\}_I$ are defined by equation (8.4).

Proof. Let Q be a genuine multipartite entanglement witness for ρ_{GME}, meaning that $\mathrm{Tr}(Q\sigma) \geqslant 0$ for any $\sigma \in \mathcal{S}_{(2)}$, while $\mathrm{Tr}(Q\rho_{GME}) < 0$. From the former assumption it follows that, for each bipartition $\{I|I^c\}$, $\alpha_I := \min_{\sigma_I \in \mathcal{S}_I} \mathrm{Tr}(Q\sigma_I) \geqslant 0$. As this minimization is convex in the space of states it is clear that the minimal overlap with Q can also be reached by an optimal pure state $|\varphi_I\rangle\langle\varphi_I| \in \mathcal{S}_I$. Now we can choose the following bipartite witness: $W_I = Q - \alpha_I|\varphi_I\rangle\langle\varphi_I|$. We can then verify that, for each bipartition $\{I|I^c\}$, on the one hand

$$\forall\ \sigma_I \in \mathcal{S}_I,\ \mathrm{Tr}(W_I\sigma_I) \geqslant \mathrm{Tr}(Q\sigma_I) - \alpha_I \geqslant 0,\ \text{and}\ \mathrm{Tr}(W_I|\varphi_I\rangle\langle\varphi_I|) = 0,$$

while on the other hand

$$\mathrm{Tr}\left(W_I\rho_{GME}\right) \leqslant \mathrm{Tr}\left(Q\rho_{GME}\right) < 0.$$

So indeed, for each bipartition $\{I|I^c\}$, W_I is a weakly optimal witness detecting that $\rho_{GME} \notin \mathcal{S}_I$.

Inserting this set of optimal bipartite witnesses into the construction from above using $[-\alpha_I|\varphi_I\rangle\langle\varphi_I|]_+ = 0$ yields $W_{GME} = Q$, proving that each genuine multipartite entanglement witness can be gained from the construction using weakly optimal bipartite entanglement witnesses. □

As a direct consequence of Theorem 8.3.1 we have, as wanted, that for any set of bipartite entanglement witnesses $\{W_I\}_I$ for ρ_{GME}, there exists a Q such that W_{GME} as defined by equation (8.5) is a genuinely multipartite entanglement witness for ρ_{GME}.

Corollary 8.3.2. *Every multipartite entanglement witness W_{GME} can be constructed using the framework summarized by equations* (8.4) *and* (8.5). *Furthermore, it is possible to impose that, for every bipartition $\{I|I^c\}$, $W_I = \mathcal{N}_I^* \otimes \mathcal{I}d_{I^c}\left(|\psi_I\rangle\langle\psi_I|\right)$ for some choice of positive map $\mathcal{N}_I$, acting on the subsystems in I, and some choice of pure state $|\psi_I\rangle$.*

Proof. First it is important to notice that, for every state ρ which is entangled across a specific bipartition $\{I|I^c\}$, there exists a positive map $\mathcal{N}_I$, acting on the subsystems in I, such that $\mathcal{N}_I \otimes \mathcal{Id}_{I^c}(\rho)$ is not positive semidefinite, i.e. it is detected to be entangled by this positive map (see [106, 58, 117]). This implies in turn that there exists a unit vector $|\psi_I\rangle$ such that $\langle\psi_I|\mathcal{N}_I \otimes \mathcal{Id}_{I^c}(\rho)|\psi_I\rangle < 0$, i.e. $\mathrm{Tr}\left(\rho \mathcal{N}_I^* \otimes \mathcal{Id}_{I^c}\left(|\psi_I\rangle\langle\psi_I|\right)\right) < 0$. Through continuity this implies that, for every extremal biseparable state $|\phi\rangle\langle\phi|$, there exists a weakly optimal witness of the form $\widetilde{W}_I = \mathcal{N}_I^* \otimes \mathcal{Id}_{I^c}\left(|\psi_I\rangle\langle\psi_I|\right)$ such that $\mathrm{Tr}\big(\widetilde{W}_I|\phi\rangle\langle\phi|\big) = 0$. Now, invoking the Choi-Jamiolkowski isomorphism, we can conclude that every possible hyperplane intersecting the biseparable set corresponds to a witness derived from a positive map. This implies that amongst all the $\{\widetilde{W}_I\}_I$, there is at least one of them $\widetilde{W}_{I_0}$ which satisfies the following: for each bipartition $\{I|I^c\}$, $W_I = \widetilde{W}_{I_0} - \beta_I \mathrm{Id}$, for some β_I, is an optimal bipartite witness for ρ (and $\beta_{I_0} = 0$). This concludes the proof: the $\{W_I\}_I$ are all obtained from shifting the same $\widetilde{W}_{I_0} = \mathcal{N}_{I_0}^* \otimes \mathcal{Id}_{I_0^c}\left(|\psi_{I_0}\rangle\langle\psi_{I_0}|\right)$. □

8.3.2 Illustration on one example

To showcase the strength of the above constructions let us illustrate it with a peculiar example. First of all let us make a specific choice for Q that has already proven to work well in [112]. Given a set of witnesses for every bipartition $\{W_I\}_I$, we will construct Q by finding element wise the matrix of largest common negative matrix entries

$$N = \sum_{j,j'=1}^{d} |j\rangle\langle j'| \min\left[0, \max_I \left[\Re e\langle j|W_I|j'\rangle\right]\right],$$

and the matrix of smallest common positive matrix entries

$$P = \sum_{j,j'=1}^{d} |j\rangle\langle j'| \max\left[0, \min_I \left[\Re e\langle j|W_I|j'\rangle\right]\right],$$

where d denotes the global dimension. With these two matrices we can then define

$$Q := N + P\,.$$

Now for the purpose of elucidating how this construction works in practice let us follow it through step by step in an exemplary three-qutrit case. In order to write down our target state, we first define the vector $|\psi_0\rangle = |111\rangle + |222\rangle + |333\rangle$. We next fix $a_1, a_2, a_3 > 0$ and define the following vectors

$$|\psi_1\rangle = \sqrt{a_1}|112\rangle + \sqrt{\frac{1}{a_1}}|221\rangle, \quad |\psi_2\rangle = \sqrt{a_1}|121\rangle + \sqrt{\frac{1}{a_1}}|212\rangle, \quad |\psi_3\rangle = \sqrt{a_1}|211\rangle + \sqrt{\frac{1}{a_1}}|122\rangle,$$
$$|\psi_4\rangle = \sqrt{a_2}|223\rangle + \sqrt{\frac{1}{a_2}}|332\rangle, \quad |\psi_5\rangle = \sqrt{a_2}|232\rangle + \sqrt{\frac{1}{a_2}}|323\rangle, \quad |\psi_6\rangle = \sqrt{a_2}|322\rangle + \sqrt{\frac{1}{a_2}}|233\rangle,$$
$$|\psi_7\rangle = \sqrt{a_3}|331\rangle + \sqrt{\frac{1}{a_3}}|113\rangle, \quad |\psi_8\rangle = \sqrt{a_3}|313\rangle + \sqrt{\frac{1}{a_3}}|131\rangle, \quad |\psi_9\rangle = \sqrt{a_3}|133\rangle + \sqrt{\frac{1}{a_3}}|311\rangle.$$

We can then construct a three-qutrit mixed state ρ as follows

$$\rho := \frac{\tilde{\rho}}{\mathrm{Tr}\,\tilde{\rho}}, \text{ where } \tilde{\rho} := \sum_{i=0}^{9} |\psi_i\rangle\langle\psi_i| + p(|112\rangle\langle 112| + \mathrm{Id}).$$

By construction this state is PPT across all three cuts. If we choose $a_1 = 10^{-6}$, $a_2 = 300$ and $a_3 = 12 \times 10^{-3}$, then for $p > 0.0003$ it is also PPC across all three cuts. As it is both fully PPT and PPC it is fair to say that if it is entangled it is only very weakly so (from a standard detection tool point of view). The system's size is still small enough to apply the positive map mixer introduced in the previous section, and to reveal that it is nonetheless indeed genuinely multipartite entangled for various values of $p > 0.0003$.

Hence, this is a good opportunity to demonstrate the power of witness liftings. We can use the indecomposable map introduced in [145], choosing $c_1 = 1$, $c_2 = 10^{-3}$ and $c_3 = 10^3$, revealing entanglement across all three cuts. In fact, even without ever calculating any eigenvalues, we can just apply the map on the three subsystems of $|\psi_0\rangle$ only. This means taking

$$W_1 = \mathcal{N}_1^{c_1,c_2,c_3} \otimes \mathcal{Id}_{23}\left(|\psi_0\rangle\langle\psi_0|\right), \; W_2 = \mathcal{N}_2^{c_1,c_2,c_3} \otimes \mathcal{Id}_{13}\left(|\psi_0\rangle\langle\psi_0|\right), \; W_3 = \mathcal{N}_3^{c_1,c_2,c_3} \otimes \mathcal{Id}_{12}\left(|\psi_0\rangle\langle\psi_0|\right).$$

Plugging the three resulting witnesses in our construction for Q we get a GME-witness that is able to reveal genuine multipartite entanglement in the state for a range of $0 \leqslant p \leqslant 0.00069$. This example illustrates that even bound entangled states which are positive with respect to paradigmatic maps can exhibit multipartite entanglement. As the state itself is full rank and not symmetric, there is no other known method that could have revealed it to be GME, clearly demonstrating the power of positive map mixers and witness liftings.

One question at this point is that of how exceptional the fact of being fully PPT and PPC, but nevertheless GME, is. Finding explicit examples of such states is not so easy. Indeed, in small dimensions things are quite contrived and there is not much room to move (as just exemplified), while as the dimensions grow computations quickly become untractable. However, as we shall see in Section 8.5, this feature is actually generic in large dimensions.

8.3.3 Finding a multiparticle witness with semidefinite programming

In [112] and in the previous example, the multipartite witness W_{GME} was constructed by starting with bipartite witnesses W_I for each bipartition $\{I|I^c\}$, and then one possible choice for the operator Q was explicitly constructed. While the presented choice works well for many examples, as just illustrated, it is not clear why it would be the best one. We leave open this optimality question at this level of generality. Let us make two easy observations though. On the one hand, note that the choice $Q = 0$ yields $W_{GME} = \sum_I [W_I]_+ \geqslant 0$, which detects absolutely no GME state. On the other hand, in the case where there exists an I_0 such that, for each I, $W_I = W_{I_0} - \alpha_I \mathrm{Id}$ with $\alpha_I \geqslant 0$, then the choice $Q = W_{I_0}$ yields $W_{GME} = W_{I_0}$, which detects all the GME states whose bipartite entanglement is detected by the $\{W_I\}_I$.

Let us also mention that, for a given state ρ, the optimal multipartite witness W_{GME} can directly be computed as a semidefinite program that is easier to run than the map-mixers.

For that, consider the following constrained optimization problem

$$\text{minimize:} \quad \mathrm{Tr}\big(\rho W_{GME}\big) \quad \text{subject to:} \quad \forall\, I,\ W_{GME} \geqslant W_I .$$

This is a semidefinite program, which can be easily and efficiently solved using standard numerical techniques.

There is, in addition, a variation of the presented method for obtaining a multiparticle witness from a set of bipartite witnesses. The condition imposed by equation (8.3) guarantees that, for each I, $\mathrm{Tr}(\sigma W_{GME}) \geqslant \mathrm{Tr}(\sigma W_I)$ for any state σ, so that W_{GME} is indeed an entanglement witness. However, for being an entanglement witness it suffices that, for each I, $\mathrm{Tr}(\sigma_I W_{GME}) \geqslant \mathrm{Tr}(\sigma_I W_I)$ for any state σ_I which is separable for the bipartition $\{I|I^c\}$. And this is already guaranteed if, for all I, there exists a positive map $\mathcal{N}_I$, acting on subsystems in I, such that

$$\mathcal{N}_I \otimes \mathcal{I}d_{I^c}\left(W_{GME}\right) \geqslant \mathcal{N}_I \otimes \mathcal{I}d_{I^c}\left(W_I\right).$$

So, for computing a multipartite witness for a given state one can also consider the semidefinite program

$$\text{minimize:} \quad \mathrm{Tr}\big(\rho W_{GME}\big) \quad \text{subject to:} \quad \forall\, I,\ \mathcal{N}_I \otimes \mathcal{I}d_{I^c}\left(W_{GME}\right) \geqslant \mathcal{N}_I \otimes \mathcal{I}d_{I^c}\left(W_I\right).$$

These two semidefinite programs can be used to construct the best witness for a given state systematically. They might be useful, if the analytical method from [112] does not work.

It should be noted, however, that the presented formulations are not necessarily the best way to construct a multiparticle witness from bipartite witnesses, if one wishes to use semidefinite programming. The reason is the following: For each bipartite witness one can construct via the Choi-Jamiolkowski isomorphism a positive, but not completely positive map. This map detects more states than the original witness. Given these maps one can then evaluate the corresponding map relaxations, as described in Section 8.2. This criterion will be stronger than the semidefinite programs presented above. For that reason, we do not discuss detailed examples here.

8.4 Relaxations of separability beyond positive maps

So far, we considered only the approximation of the biseparable set by super-sets which are associated to positive maps. One can, however, also use other bipartite separability criteria, such as the computable cross norm (CCNR), aka realignment, criterion [159, 44] or the symmetric extension, aka k-extendibility, criterion [66] (for the latter, see also Chapter 9 of this manuscript). In the following we explain how the CCNR criterion can be used in the multipartite setting.

8.4.1 Description of the method

Let us start by explaining the CCNR criterion. Any quantum state ρ on the bipartite Hilbert space $\mathrm{H}_1 \otimes \mathrm{H}_2$ can be expressed (via the Schmidt decomposition in the space of linear operators) as

$$\rho = \sum_i \lambda_i G_1^{(i)} \otimes G_2^{(i)},$$

where the λ_i are positive coefficients and the $G_1^{(i)}$, $G_2^{(i)}$ are orthogonal observables on H_1, H_2, i.e. they fulfill $\mathrm{Tr}\big(G_1^{(i)} G_1^{(j)}\big) = \mathrm{Tr}\big(G_2^{(i)} G_2^{(j)}\big) = \delta_{ij}$. With this representation one can easily prove that the following holds:

$$\rho \in \mathcal{S}(\mathrm{H}_1{:}\mathrm{H}_2) \;\Rightarrow\; \sum_i \lambda_i \leqslant 1.$$

And this necessary condition for separability is known as the CCNR criterion. The criterion has the advantage that it detects entanglement in many states where the PPT criterion fails. On the other hand, not all two-qubit entangled states can be detected by this test.

From this structure, one can easily write down entanglement witnesses. Namely, any operator of the form

$$W = \mathrm{Id} - \sum_i G_1^{(i)} \otimes G_2^{(i)}$$

is an entanglement witness, as it is positive on all states with $\sum_i \lambda_i \leqslant 1$.

This structure can be used for constructing witnesses for genuine multipartite entanglement as follows. Consider an operator W_{GME} on the k-partite Hilbert space $\mathrm{H}_1 \otimes \cdots \otimes \mathrm{H}_k$ which is such that, for any bipartition $\{I|I^c\}$ of the k subsystems,

$$W_{GME} = P_I + \mathrm{Id} - \sum_i G_I^{(i)} \otimes G_{I^c}^{(i)}, \tag{8.6}$$

where the $G_I^{(i)}$, $G_{I^c}^{(i)}$ are orthogonal observables on H_I, H_{I^c}, and $P_I \geqslant 0$ is positive semidefinite. Clearly, if a state obeys the CCNR criterion for some bipartition, the mean value of the witness W_{GME} will be non-negative. Consequently, the witness is also non-negative on all biseparable states.

8.4.2 Example: the three-qubit GHZ state

The witnesses from the CCNR criterion are more difficult to handle than the witnesses from positive maps. The reason is that no approach via semidefinite programming is possible. Moreover, the condition from equation (8.6) is more difficult to check than the condition in equation (8.3). Nevertheless, we will present an example where known optimal entanglement witnesses have this structure.

Consider first the three-qubit Greenberger-Horne-Zeilinger (GHZ) state $|GHZ\rangle = (|111\rangle + |222\rangle)/\sqrt{2}$. The typical witness for this state is $W = \mathrm{Id}/2 - |GHZ\rangle\langle GHZ|$. Now, the GHZ state can be expressed in terms of its stabilizers as

$$|GHZ\rangle\langle GHZ| = \frac{1}{8}(111 + ZZ1 + Z1Z + 1ZZ + XXX - XYY - YXY - YYX),$$

where $1, X, Y, Z$ represent the Pauli matrices $\mathrm{Id}, \sigma_x, \sigma_y, \sigma_z$, and tensor product signs have been omitted. After a change of the normalization this can be used to write the witness as

$$W = \mathrm{Id} - 2|GHZ\rangle\langle GHZ| = \mathrm{Id} - \left[\frac{Z}{\sqrt{2}}\frac{Z1+1Z}{\sqrt{8}} + \frac{X}{\sqrt{2}}\frac{XX-YY}{\sqrt{8}} + \frac{-Y}{\sqrt{2}}\frac{YX+XY}{\sqrt{8}} + \frac{1}{\sqrt{2}}\frac{11+ZZ}{\sqrt{8}}\right].$$

From this representation, it is clear that W is a witness as in equation (8.6) for the bipartition $\{\{1\}|\{2,3\}\}$, with $P_1 = 0$. Due to the symmetry, this works for all bipartitions.

8.5 Estimating the performance of PPT relaxations in high dimensions

In this final section we want to discuss the overall performance of such relaxations in multipartite systems, using the paradigmatic partial transpose map. In order to estimate the performance of using PPT relaxations

to detect randomly chosen multipartite entangled states we derive lower bounds on the fraction of multipartite entangled states, among states which are positive under partial transposition across every cut. The latter condition is strictly stronger than the relaxation employed in equation (8.1), thus providing an upper bound on the fraction of states in $\mathcal{R}_{\{\mathcal{T}_I\}_I}$ (where, as before, $\mathcal{T}_I$ stands for the transposition map on the subsystems in I) that are also in $\mathcal{S}_{(2)}$.

Our main result can be summarized as follows: for a fixed number of parties, the ratio between the size of fully PPT states and the size of biseparable states (as measured by either the volume radius or the mean width) scales at least as $\sqrt{d}$, where d is the local dimension.

In order to precisely formulate this result, we need to introduce first some of the basic notions and definitions that will be employed in the derivations.

8.5.1 Notation and preliminary technical remarks

All notation, concepts and results from classical convex geometry, which are required throughout our proofs, are gathered in Chapter 4, Section 4.1 (see also Chapter 1, Section 1.3, for general notation that we use in the remainder of this section).

It may, however, be worth mentioning that whenever we use tools from convex geometry in the space $\mathscr{H}(\mathbf{C}^n)$ of Hermitian operators on $\mathbf{C}^n$ (which has real dimension n^2) it is tacitly understood that we use the Euclidean structure induced by the Hilbert–Schmidt inner product $\langle A, B\rangle = \mathrm{Tr}(AB)$. For instance, Definition 4.1.1 of the volume radius of a convex body $K \subset \mathscr{H}(\mathbf{C}^n)$ becomes, denoting by $B_{HS}(\mathbf{C}^n)$ the Hilbert–Schmidt unit ball of $\mathscr{H}(\mathbf{C}^n)$,

$$\mathrm{vrad}(K) = \left(\frac{\mathrm{vol}\, K}{\mathrm{vol}\, B_{HS}(\mathbf{C}^n)}\right)^{1/n^2}.$$

While Definition 4.1.2 of its mean width is, denoting by $S_{HS}(\mathbf{C}^n)$ the Hilbert–Schmidt unit sphere of $\mathscr{H}(\mathbf{C}^n)$ equipped with the uniform probability measure σ,

$$w(K) = \int_{S_{HS}(\mathbf{C}^n)} \max_{M\in K} \mathrm{Tr}(XM)\, \mathrm{d}\sigma(X).$$

As also mentioned in Chapter 4, Section 4.1, the latter quantity can be re-expressed via Gaussian variables, which yields here

$$w(K) = \frac{1}{\gamma(n)}\, \mathbf{E}\left(\max_{M\in K} \mathrm{Tr}(GM)\right),$$

where G is a matrix from the Gaussian Unitary Ensemble (GUE) on $\mathbf{C}^n$ and $\gamma(n) = \mathbf{E}\,\|G\|_2 \sim_{n\to+\infty} n$ (see e.g. [4], Chapter 2, for a proof).

To be fully rigorous, let us make one last comment. All the convex bodies of $\mathscr{H}(\mathbf{C}^n)$ that we shall consider will actually be included in the set $\mathscr{D}(\mathbf{C}^n)$ of density operators on $\mathbf{C}^n$ (i.e. the set of positive and trace 1 operators on $\mathbf{C}^n$). So we will in fact be working in an ambient space of real dimension $n^2 - 1$ (namely the hyperplane of $\mathscr{H}(\mathbf{C}^n)$ composed of trace 1 elements). This subtlety will not be an issue though, since we will be mostly interested in the asymptotic regime $n \to +\infty$. In this setting, the operator that will play for us the role of the origin will naturally be the center of mass of $\mathscr{D}(\mathbf{C}^n)$, i.e. the maximally mixed state Id/n.

Theorem 8.5.1. *On $\mathbf{C}^n$, the volume radius and the mean width of the set $\mathscr{D}(\mathbf{C}^n)$ of all quantum states satisfy the asymptotic estimates,*

$$\mathrm{vrad}\left(\mathscr{D}(\mathbf{C}^n)\right) \underset{n\to+\infty}{\sim} \frac{e^{-1/4}}{\sqrt{n}}, \tag{8.7}$$

$$w\big(\mathscr{D}(\mathbf{C}^n)\big) \underset{n\to+\infty}{\sim} \frac{2}{\sqrt{n}}. \tag{8.8}$$

Proof. Equation (8.7) was established in [167].

Equation (8.8) is a direct consequence of Wigner's semicircle law (see e.g. [4], Chapter 2, for a proof). Indeed, we have by definition

$$w\big(\mathscr{D}(\mathbf{C}^n)\big) = \frac{1}{\gamma(n)}\, \mathbf{E}\left(\max_{\rho\in\mathscr{D}(\mathbf{C}^n)} \mathrm{Tr}\left[G\left(\rho - \frac{\mathrm{Id}}{n}\right)\right]\right) = \frac{1}{\gamma(n)}\, \mathbf{E}\left(\max_{\rho\in\mathscr{D}(\mathbf{C}^n)} \mathrm{Tr}[G\rho]\right) = \frac{1}{\gamma(n)}\, \mathbf{E}\left(\lambda_{max}(G)\right),$$

where G is a GUE matrix on $\mathbf{C}^n$ and we denoted by $\lambda_{max}(G)$ its largest eigenvalue (the second equality being because $\mathbf{E}\, G = 0$). The claimed result then follows from $\gamma(n) \sim_{n\to+\infty} n$ and $\mathbf{E}\left(\lambda_{max}(G)\right) \sim_{n\to+\infty} 2\sqrt{n}$. □

8.5.2 Volume estimates

In the sequel, we shall consider the multipartite system $(\mathbf{C}^d)^{\otimes k}$, and slightly adapt and generalize the notation introduced in Section 8.2. We shall denote by $\mathcal{S}$ and $\mathcal{P}$ the sets of states on $(\mathbf{C}^d)^{\otimes k}$ which are, respectively, separable and PPT across any bi-partition, and by $\mathcal{S}_{(2)}$ and $\mathcal{P}_{(2)}$ the sets of states on $(\mathbf{C}^d)^{\otimes k}$ which are, respectively, bi-separable and bi-PPT. These sets may be more precisely defined in the following way. There are $N_k = 2^{k-1}-1$ different bi-partitions of the k subsystems $\mathbf{C}^d$. Denoting by $\left\{\mathcal{S}^1, \ldots, \mathcal{S}^{N_k}\right\}$ and by $\left\{\mathcal{P}^1, \ldots, \mathcal{P}^{N_k}\right\}$ the sets of states which are, respectively, bi-separable and bi-PPT across one of these, we have

$$\mathcal{S} = \bigcap_{i=1}^{N_k} \mathcal{S}^i \quad \text{and} \quad \mathcal{P} = \bigcap_{i=1}^{N_k} \mathcal{P}^i,$$

$$\mathcal{S}_{(2)} = \operatorname{conv}\left(\bigcup_{i=1}^{N_k} \mathcal{S}^i\right) \quad \text{and} \quad \mathcal{P}_{(2)} = \operatorname{conv}\left(\bigcup_{i=1}^{N_k} \mathcal{P}^i\right).$$

Theorem 8.5.2. *There exist positive constants $c_d \to_{d\to+\infty} 1$ such that, on $(\mathbf{C}^d)^{\otimes k}$, the volume radius and the mean width of the set of states which are PPT across any bi-partition satisfy*

$$w(\mathcal{P}) \geqslant \operatorname{vrad}(\mathcal{P}) \geqslant c_d \, \frac{c^{2^k}}{d^{k/2}}, \tag{8.9}$$

where one may choose $c = e^{-1/4}/4$.

Proof. The first inequality in equation (8.9) is just by the Urysohn inequality (see Theorem 4.1.3 in Chapter 4, Section 4.1).

To show the second inequality in equation (8.9), we will use repeatedly the Milman–Pajor inequality (see Theorem 4.1.6) and more specifically its Corollary 4.1.7 (both of them in Chapter 4, Section 4.1). We will in fact show more precisely that there exist $c_d \to_{d\to+\infty} 1$ such that

$$\operatorname{vrad}(\mathcal{P}) \geqslant c_d \, \frac{c^{N_k} \, e^{-1/4}}{d^{k/2}}. \tag{8.10}$$

The first thing to note is that, denoting by $\Gamma_1, \ldots, \Gamma_{N_k}$ the partial transpositions across the N_k different bi-partitions of the k subsystems $\mathbf{C}^d$, we have

$$\mathcal{P} = \mathcal{D} \cap \mathcal{D}^{\Gamma_1} \cap \cdots \cap \mathcal{D}^{\Gamma_{N_k}}.$$

Now, by Corollary 4.1.7 applied to the convex body $\mathcal{D} \subset \mathscr{H}((\mathbf{C}^d)^{\otimes k})$ (which indeed has the origin Id/d^k as center of mass) and to the isometry Γ_1, we get

$$\operatorname{vrad}\left(\mathcal{D} \cap \mathcal{D}^{\Gamma_1}\right) \geqslant \frac{1}{2} \frac{\operatorname{vrad}(\mathcal{D})^2}{w(\mathcal{D})} \underset{d\to+\infty}{\sim} c \times \frac{e^{-1/4}}{d^{k/2}},$$

the last equivalence being by Theorem 8.5.1. We may then conclude recursively that equation (8.10) actually holds. □

Theorem 8.5.3. *On $(\mathbf{C}^d)^{\otimes k}$, the volume radius and the mean width of the set of bi-separable states satisfy*

$$\operatorname{vrad}(\mathcal{S}_{(2)}) \leqslant w(\mathcal{S}_{(2)}) \leqslant \frac{C + C_{d,k}}{d^{(k+1)/2}}, \tag{8.11}$$

where one may choose $C = \min\left\{6\sqrt{\ln(1+2/\delta)}/(1-2\delta^2)^2 \ : \ 1/10 < \delta < 1/4\right\}$ and $C_{d,k} = \sqrt{8\ln(2)/d^{k-1}}$, so that $C \leqslant 11$ and $C_{d,k} \to_{d\to+\infty} 0$.

Proof. The first inequality in equation (8.11) is just by the Urysohn inequality (see Theorem 4.1.3 in Chapter 4, Section 4.1).

To show the second inequality in equation (8.11), we start from this observation: for each $1 \leqslant i \leqslant N_k$,

$$w(\mathcal{S}^i) \leqslant \frac{C/2}{d^{(k+1)/2}}, \tag{8.12}$$

where $C = \min\left\{6\sqrt{\ln(1+2/\delta)}/(1-2\delta^2)^2 \ : \ 1/10 < \delta < 1/4\right\}$. It relies on the already known fact that there exists a universal constant $\widetilde{C}$ such that, for any $m, n \in \mathbf{N}$ with $m \leqslant n$, the mean width of the set $\mathcal{S}$ of separable

states on $\mathbf{C}^m \otimes \mathbf{C}^n$ is upper bounded by $\widetilde{C}/m\sqrt{n}$. In that way, the upper bound appearing in equation (8.12) is simply the upper bound obtained for one of the k sets of states on $(\mathbf{C}^d)^{\otimes k}$ which are separable across a given bipartite cut $\mathbf{C}^d$:$(\mathbf{C}^d)^{\otimes k-1}$ (the largest of all upper bounds for sets of states on $(\mathbf{C}^d)^{\otimes k}$ which are separable across some bipartite cut). The former result was basically proved in [16], Theorem 1, but since specifically stated there in the balanced case $m = n$ only, for vrad($\mathcal{S}$) rather than $w(\mathcal{S})$ and without specifying that one may choose $\widetilde{C} = C/2$, we briefly recall the argument here.

Let $1/10 < \delta < 1/4$ and consider $\mathcal{A}_\delta$, $\mathcal{B}_\delta$ δ-nets for $\|\cdot\|$ within the Euclidean unit spheres of $\mathbf{C}^m$ and $\mathbf{C}^n$ respectively. Imposing that $\mathcal{A}_\delta$, $\mathcal{B}_\delta$ have minimal cardinality, we know by volumetric arguments (see Lemma 4.2.3 in Chapter 4, Section 4.2) that $|\mathcal{A}_\delta| \leqslant (1+2/\delta)^{2m}$ and $|\mathcal{B}_\delta| \leqslant (1+2/\delta)^{2n}$. Then, it may be checked that

$$\operatorname{conv}(\mathcal{S} \cup -\mathcal{S}) \subset \frac{1}{(1-2\delta^2)^2} \operatorname{conv}\left\{ \pm |x\rangle\langle x| \otimes |y\rangle\langle y| \ : \ |x\rangle \in \mathcal{A}_\delta, \ |y\rangle \in \mathcal{B}_\delta \right\}.$$

So by Lemma 4.1.4 in Chapter 4, Section 4.1, we get

$$\begin{aligned} w(\mathcal{S}) &\leqslant w(\operatorname{conv}(\mathcal{S} \cup -\mathcal{S})) \\ &\leqslant \frac{1}{(1-2\delta^2)^2} \sqrt{\frac{2\ln\left(2(1+2/\delta)^{2m}(1+2/\delta)^{2n}\right)}{(mn)^2}} \\ &= \frac{1}{(1-2\delta^2)^2} \frac{\sqrt{4(m+n)\ln(1+2/\delta) + \ln(4)}}{mn} \\ &\leqslant \frac{3\sqrt{\ln(1+2/\delta)}}{(1-2\delta^2)^2 m\sqrt{n}}, \end{aligned}$$

which is precisely the content of equation (8.12).

Now, we also have that, for each $1 \leqslant i \leqslant N_k$, $\mathcal{S}^i \subset B_1 \subset B_{HS}$. Hence, by Lemma 4.1.5 in Chapter 4, Section 4.1, we get

$$w(\mathcal{S}_{(2)}) \leqslant 2\left(\max_{1 \leqslant i \leqslant N_k} w(\mathcal{S}^i) + \sqrt{\frac{2\ln(N_k)}{(d^k)^2}} \right) \leqslant \frac{C}{d^{(k+1)/2}} + \frac{\sqrt{8\ln(2)k}}{d^k} = \frac{C + C_{d,k}}{d^{(k+1)/2}},$$

where $C_{d,k} = \sqrt{8\ln(2)/d^{k-1}}$. □

The conclusion of Theorems 8.5.2 and 8.5.3 may be phrased as follows. On a multipartite system which is composed of a small number of big subsystems (k fixed and $d \to +\infty$), imposing that a state is PPT across any bi-partition (i.e. the strongest notion of PPT) is still, on average, a much less restrictive constraint than imposing that it is bi-separable (i.e. the weakest notion of separability). Indeed, the "sizes" of these two sets of states (measured by either their volume radii or their mean widths) scale completely differently: the "size" of the former is at least of order $1/d^{k/2}$ while the "size" of the latter is at most of order $1/d^{(k+1)/2}$, hence differing by a factor of order at least $\sqrt{d}$.

8.5.3 A class of fully PPT and GME states

In Section 8.3 an explicit class of GME states which are PPT across all cuts was presented. In small dimensions it is a hard task to find such examples, but the results from the previous section suggest that at least in high dimensions being GME should be a generic feature of fully PPT states. To emphasize this fact we present a construction of random states which are with high probability PPT across all cuts and GME.

Consider the following random state model on $(\mathbf{C}^d)^{\otimes k}$: fix some parameter $0 < \alpha < 1/4$ (independent of d), pick G a traceless GUE matrix on $(\mathbf{C}^d)^{\otimes k}$, and define the "maximally mixed + gaussian noise" state on $(\mathbf{C}^d)^{\otimes k}$

$$\rho_G = \frac{1}{d^k}\left(\mathrm{Id} + \frac{\alpha}{d^{k/2}} G\right). \tag{8.13}$$

Then, typically (i.e. with probability going to 1 as d grows) ρ_G is fully PPT and nevertheless GME.

More quantitatively, we will show that the following result holds.

Theorem 8.5.4. *Let G be a traceless GUE matrix on $(\mathbf{C}^d)^{\otimes k}$. Then, the state ρ_G on $(\mathbf{C}^d)^{\otimes k}$, as defined by equation (8.13), is fully PPT and not bi-separable with probability greater than $1 - \exp(-cd^{k-1})$, for some universal constant $c > 0$.*

Theorem 8.5.4 is a straightforward consequence of Propositions 8.5.5 and 8.5.6 below. Before stating and proving them, let us elude once and for all a slight issue: a GUE matrix on $\mathbf{C}^n$ is the standard Gaussian vector in $\mathscr{H}(\mathbf{C}^n)$, while a traceless GUE matrix on $\mathbf{C}^n$ is the standard Gaussian vector in the hyperplane of $\mathscr{H}(\mathbf{C}^n)$ composed of trace 0 elements. So in the asymptotic regime $n \to +\infty$, all the known results on $n \times n$ GUE matrices that we shall use also hold for traceless $n \times n$ GUE matrices (because the ambient spaces of these two Gaussian vectors have equivalent dimensions in this limit).

Proposition 8.5.5. *Let G be a traceless GUE matrix on $(\mathbf{C}^d)^{\otimes k}$. Then, the state ρ_G on $(\mathbf{C}^d)^{\otimes k}$, as defined by equation (8.13), satisfies*

$$\mathbf{P}\left(\rho_G \notin \mathscr{P}\right) \leqslant N_k e^{-cd^k},$$

where $c > 0$ is a universal constant.

Proof. In [8], a deviation inequality is proved for the smallest eigenvalue of a GUE matrix, namely: Let G be a GUE matrix on $\mathbf{C}^n$ and denote by $\lambda_{min}(G)$ its smallest eigenvalue. Then, for any $\varepsilon > 0$,

$$\mathbf{P}\left(\lambda_{min}(G) < -(2+\varepsilon)\sqrt{n}\right) \leqslant e^{-c\varepsilon^{3/2}n}, \tag{8.14}$$

where $c > 0$ is a universal constant.

Now, observe that G as well as all its partial transpositions G^{Γ_i}, $1 \leqslant i \leqslant N_k$, are GUE matrices on $(\mathbf{C}^d)^{\otimes k}$. Hence, Proposition 8.5.5 follows directly, by choosing for instance $\varepsilon = 1$. Indeed, by assumption on α, we have $3\alpha < 3/4 < 1$, so the probability that ρ_G or any $\rho_G^{\Gamma_i}$, $1 \leqslant i \leqslant N_k$, is not positive is less than e^{-cd^k}. The advertised result follows by the union bound. □

Proposition 8.5.6. *Let G be a traceless GUE matrix on $(\mathbf{C}^d)^{\otimes k}$. Then, the state ρ_G on $(\mathbf{C}^d)^{\otimes k}$, as defined by equation (8.13), satisfies*

$$\mathbf{P}\left(\rho_G \in \mathcal{S}_{(2)}\right) \leqslant e^{-cd^{k-1}},$$

where $c > 0$ is a universal constant.

Proof. Our strategy to show Proposition 8.5.6 is to exhibit a Hermitian M on $(\mathbf{C}^d)^{\otimes k}$ which is with probability greater than $1 - \exp(-cd^{k-1})$ a GME witness for the state ρ_G (i.e. such that $\mathrm{Tr}(\rho_G M) < 0$ while $\mathrm{Tr}(\rho M) > 0$ for any bi-separable state ρ).

Note first of all that, on the one hand,

$$\mathbf{E}\,\mathrm{Tr}(\rho_G G) = \frac{\alpha}{d^{3k/2}}\,\mathbf{E}\,\mathrm{Tr}(G^2) \underset{d\to+\infty}{\sim} \frac{\alpha}{d^{3k/2}} d^{2k} = \alpha d^{k/2}, \tag{8.15}$$

while on the other hand, by Theorem 8.5.3,

$$\mathbf{E} \sup_{\rho\in\mathcal{S}_{(2)}} \mathrm{Tr}(\rho G) \leqslant \frac{C + C_{d,k}}{d^{(k+1)/2}}\,\mathbf{E}[\mathrm{Tr}(G^2)]^{1/2} \underset{d\to+\infty}{\sim} \frac{C}{d^{(k+1)/2}} d^k = C d^{(k-1)/2}, \tag{8.16}$$

where we used that for G a GUE matrix on $\mathbf{C}^n$, $\mathbf{E}\,\mathrm{Tr}(G^2) \sim_{n\to+\infty} n^2$ and $\mathbf{E}[\mathrm{Tr}(G^2)]^{1/2} \sim_{n\to+\infty} n$ (see e.g. [4], Chapter 2, for a proof).

Let us now show that the functions $G \mapsto \mathrm{Tr}(\rho_G G)$ and $G \mapsto \sup_{\rho\in\mathcal{S}_{(2)}} \mathrm{Tr}(\rho G)$ concentrate around their respective average values. In that aim, we shall make use of the following Gaussian deviation inequality (see e.g. [148], Chapter 2, for a proof), which is the analogue of Levy's spherical version appearing as Lemma 4.2.1 in Chapter 4, Section 4.2: Assume that f is a function satisfying, for any Gaussian random variables G, H, $|f(G) - f(H)| \leqslant \sigma_{G,H}\|G - H\|_2$ for some $\sigma_{G,H}$ such that $\mathbf{E}\,\sigma_{G,H} \leqslant L$. Then, for any $\varepsilon > 0$,

$$\mathbf{P}\left(|f - \mathbf{E}\,f| > \varepsilon\right) \leqslant e^{-c_0\varepsilon^2/L^2}, \tag{8.17}$$

where $c_0 > 0$ is a universal constant.

Define $f : G \in \mathscr{H}(\mathbf{C}^n) \mapsto \mathrm{Tr}(G^2)$, and $f_\Sigma : G \in \mathscr{H}(\mathbf{C}^n) \mapsto \sup_{\rho\in\Sigma} \mathrm{Tr}(\rho G)$, for any given set of states Σ on $\mathbf{C}^n$. We have first,

$$|f(G) - f(H)| = |\,\mathrm{Tr}(GG^\dagger) - \mathrm{Tr}(HH^\dagger)| \leqslant \|GG^\dagger - HH^\dagger\|_1 \leqslant (\|G\|_2 + \|H\|_2)\|G - H\|_2,$$

where the last inequality is by the triangle inequality, the Cauchy–Schwarz inequality, and the invariance of $\|\cdot\|_2$ under conjugate transposition, after noticing that $GG^\dagger - HH^\dagger = G(G^\dagger - H^\dagger) + (G-H)H^\dagger$. And second,

$$|f_\Sigma(G) - f_\Sigma(H)| = \left|\sup_{\rho\in\Sigma} \mathrm{Tr}(\rho G) - \sup_{\rho\in\Sigma} \mathrm{Tr}(\rho H)\right| \leqslant \sup_{\rho\in\Sigma} |\mathrm{Tr}(\rho[G-H])| \leqslant \|G-H\|_\infty \leqslant \|G-H\|_2,$$

where the next to last inequality is because Σ is a subset of the 1-norm unit ball of Hermitians on $\mathbf{C}^n$.

Hence, the functions f and f_Σ both satisfy the hypotheses of the Gaussian deviation inequality (8.17), with $L = 2\gamma(d^k) \sim_{d\to+\infty} 2d^k$ and $L = 1$ respectively. So by the mean estimates (8.15) and (8.16), we have that for any $0 < \varepsilon < 1$,

$$\mathbf{P}\left(\mathrm{Tr}\left(\rho_G G\right) < (1-\varepsilon)\alpha d^{k/2}\right) \leqslant \exp\left(-c_0\left(\varepsilon\alpha d^{2k}\right)^2 / \left(2d^k\right)^2\right) = \exp\left(-c_0'\varepsilon^2 d^{2k}\right), \tag{8.18}$$

$$\mathbf{P}\left(\sup_{\rho\in\mathcal{S}_{(2)}} \mathrm{Tr}(\rho G) > (1+\varepsilon)Cd^{(k-1)/2}\right) \leqslant \exp\left(-c_0\left(\varepsilon C d^{(k-1)/2}\right)^2\right) = \exp\left(-c_0'\varepsilon^2 d^{k-1}\right). \tag{8.19}$$

As a consequence, we have that for any β_d satisfying $1/2Cd^{(k-1)/2} \leqslant \beta_d \leqslant 3/2\alpha d^{k/2}$, the Hermitian $M = \mathrm{Id} - \beta_d G$ on $(\mathbf{C}^d)^{\otimes k}$ is a GME witness for ρ_G with probability greater than $1 - \exp(-cd^{k-1})$, where $c > 0$ is a universal constant. Indeed, choosing $\varepsilon = 1/6$ in equation (8.18) and $\varepsilon = 1/2$ in equation (8.19), we get

$$\mathbf{P}\left(\mathrm{Tr}\left(\rho_G M\right) > -\frac{1}{4}\right) \leqslant e^{-cd^{2k}} \text{ and } \mathbf{P}\left(\sup_{\rho\in\mathcal{S}_{(2)}} \mathrm{Tr}(\rho M) < \frac{1}{4}\right) \leqslant e^{-cd^{k-1}},$$

which concludes the proof. □

We may actually say even more on the random state ρ_G defined by equation (8.13). Indeed, define for all $0 < \varepsilon < 1$ the state $\widetilde{\rho}_G(\varepsilon)$ on $(\mathbf{C}^d)^{\otimes k}$ by

$$\widetilde{\rho}_G(\varepsilon) = \varepsilon\rho_G + (1-\varepsilon)\frac{\mathrm{Id}}{d^k} = \frac{1}{d^k}\left(\mathrm{Id} + \varepsilon\frac{\alpha}{d^{k/2}}G\right).$$

Then, what the proof of Proposition 8.5.6 additionally tells us is that, as long as $\varepsilon \geqslant \gamma/\sqrt{d}$, for some constant $\gamma > 0$, $\widetilde{\rho}_G(\varepsilon)$ is with high probability not bi-separable. This means that ρ_G is typically a fully PPT state on $(\mathbf{C}^d)^{\otimes k}$ which is not bi-separable, and whose random robustness of genuinely multipartite entanglement (as defined in [171]) additionally grows at least as $\sqrt{d}$ when $d \to +\infty$.

8.6 Conclusion

The problem of characterizing genuine multipartite entanglement and biseparability is difficult. Therefore, a natural approach lies in the relaxation of the definition of biseparability: instead of considering states which are separable with respect to some bipartition, one replaces this set by an appropriate superset, e.g. defined by the PPT condition or some other positivity under partial application of a positive, yet not completely positive, map.

In this chapter we investigated this angle of attack from several perspectives. First, we established how this relaxation approach with positive maps can be evaluated with semidefinite programming and how it can be used to construct entanglement witnesses for this problem. Then, we showed that, in principle, also other relaxations, besides those obtained from positive maps (e.g. based on the CCNR criterion), are possible. Finally, we studied the accurateness of the relaxation approach. We proved rigorous bounds on the volume of the set of biseparable states as well as on the volume of the set of states which are PPT for any cut. In this way, we showed that in the limit of large dimensional multipartite systems, the relaxation approach detects only a small fraction of the multiparticle entangled states. It must be stressed, however, that this does not mean that the relaxation method is not fruitful. Indeed, it is a well known fact from the theory of two-particle entanglement that, already in such case, simple entanglement criteria miss most of the states if the dimension of the local spaces increases [25, 16, 15]. However, from a practical point of view, relaxation techniques are clearly the best tools for characterizing multiparticle entanglement available at the moment [112, 88].

For future research, there are many open questions to address. First, a more systematic analysis for the various positive maps besides the transposition would be desirable. Then, an approach for characterizing

separability classes besides biseparability (e.g. ℓ-separability, for any ℓ) would be useful. Finally, methods to certify the Schmidt-rank or the dimensionality of entanglement [111] in high-dimensional systems are needed for current experiments. Investigating the generic scaling of these quantities could also be of interest.

Chapter 9

k-extendibility of high-dimensional bipartite quantum states

Based on "k-extendibility of high-dimensional bipartite quantum states" [126].

The idea of detecting the entanglement of a given bipartite state by searching for symmetric extensions of this state was first proposed by Doherty, Parrilo and Spedialeri. The complete family of separability tests it generates, often referred to as the hierarchy of *k-extendibility tests*, has already proved to be most promising. The goal of this chapter is to try and quantify the efficiency of this separability criterion in typical scenarios. For that, we essentially take two approaches. First, we compute the average width of the set of k-extendible states, in order to see how it scales with the one of separable states. And second, we characterize when random-induced states are, depending on the ancilla dimension, with high probability violating or not the k-extendibility test, and compare the obtained result with the corresponding one for entanglement vs separability. The main results can be precisely phrased as follows: on $\mathbf{C}^d \otimes \mathbf{C}^d$, when d grows, the average width of the set of k-extendible states is equivalent to $(2/\sqrt{k})/d$, while random states obtained as partial traces over an environment $\mathbf{C}^s$ of uniformly distributed pure states are violating the k-extendibility test with probability going to 1 if $s < ((k-1)^2/4k)d^2$. Both statements converge to the conclusion that, if k is fixed, k-extendibility is asymptotically a weak approximation of separability, even though any of the other well-studied separability relaxations is outperformed by k-extendibility as soon as k is above a certain (dimension independent) value.

9.1 Introduction

Deciding whether a given bipartite quantum state is entangled or separable (or even just close to separable) is known to be a computationally hard task (see [90] and [80]). Several much more easily checkable necessary conditions for separability do exist though, the most famous and widely used ones being perhaps the positivity of partial transpose criterion [146], the realignment criterion [44] or the k-extendibility criterion [66]. All of them have in common that verifying if a given state fulfils them or not may be cast as a Semi-Definite Program (SDP) and hence be efficiently solved (see Chapter 8 of this manuscript, and the quite extensive review [65], for much more on that topic).

We focus here on a relaxation of the notion of separability of quite different kind: the so-called k-extendibility criterion for separability, which was introduced in [66]. It is especially appealing because it provides a hierarchy of increasingly powerful separability tests (expressible as SDPs of increasing dimension), which is additionally complete, meaning that any entangled state is guaranteed to fail a test after some finite number of steps in the hierarchy. Let us be more precise.

Definition 9.1.1. *Let $k \in \mathbf{N}$. A state ρ_{AB} on a bipartite Hilbert space $\mathrm{A} \otimes \mathrm{B}$ is k-extendible with respect to B if there exists a state ρ_{AB^k} on $\mathrm{A} \otimes \mathrm{B}^{\otimes k}$ which is invariant under any permutation of the B subsystems and such that $\rho_{\mathrm{AB}} = \mathrm{Tr}_{\mathrm{B}^{k-1}} \rho_{\mathrm{AB}^k}$.*

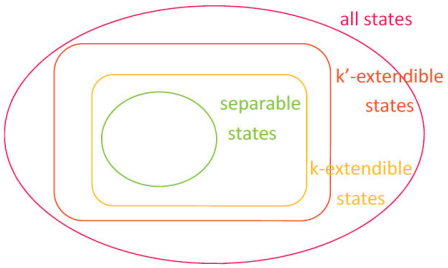
Theorem 9.1.2 (The complete family of k-extendibility criteria for separability, [66])**.** *A state on a bipartite Hilbert space $\mathrm{A} \otimes \mathrm{B}$ is separable if and only if it is k-extendible with respect to B for all $k \in \mathbf{N}$.*

Note that one direction in Theorem 9.1.2 is obvious, namely that a separable state on some bipartite system is necessarily k-extendible for all $k \in \mathbf{N}$ (with respect to both subsystems). Indeed, if $\rho = \sum_x p_x \sigma_x \otimes \tau_x$

is separable, then $\sum_x p_x \sigma_x^{\otimes k} \otimes \tau_x$ and $\sum_x p_x \sigma_x \otimes \tau_x^{\otimes k}$ are symmetric extensions of ρ to k copies of the first and second subsystems respectively. The other direction in Theorem 9.1.2 follows from the quantum finite De Finetti theorem (see e.g. [124, 48] and Chapter 3, Section 3.3, of this manuscript). The latter establishes, roughly speaking, that starting from a permutation-invariant state on some tensor power system and tracing out all except a few of the subsystems, one gets a state that may be well-approximated by a convex combination of tensor power states (with a vanishing error as the initial number of subsystems increases).

It is easy to see that if a state is k-extendible for some $k \in \mathbf{N}$, then it is automatically k'-extendible for all $k' \leqslant k$. Hence, the necessary and sufficient condition for separability provided by Theorem 9.1.2 actually decomposes into a series of increasingly constraining necessary conditions for separability, which are only asymptotically also sufficient (see Figure 9.1). In real life however, checks can only be done up to a finite level in this hierarchy. It thus makes sense to ask, given a finite $k \in \mathbf{N}$, how "powerful" the k-extendibility test is to detect entanglement.

Figure 9.1: The nested and converging sequence of k-extendibility relaxations of separability, $k \in \mathbf{N}$



Actually, various more quantitative versions of Theorem 9.1.2 do exist, that put bounds on how far a k-extendible state can be from separable. Let us mention two quite different statements in that direction. The original result, appearing in [48], establishes that a state on $\mathrm{A} \otimes \mathrm{B}$ which is k-extendible with respect to B is at distance at most $2d_{\mathrm{B}}^2/k$, in 1-norm, from the set of separable states. It is a direct consequence of one of the quantitative versions of the quantum finite De Finetti theorem. A more recent result, proved essentially in [34] and improved in [35], stipulates that such a state is at distance at most $\sqrt{2 \ln d_{\mathrm{A}}/k}$, in $\mathbf{LOCC}^{\rightarrow}$-norm, from the set of separable states (see Chapter 7, Section 7.2, for a precise definition of the operational one-way-LOCC norm). It relies on the observation that a k-extendible state has a small squashed entanglement, and therefore cannot be distinguished well from a separable state by local observers (see Chapter 10, Section 10.4, for a related discussion). The main problem of such estimates is that they become non-trivial only when $k \gg d_{\mathrm{B}}$ or $k \gg \ln d_{\mathrm{A}}$. So in the case where $d_{\mathrm{A}}, d_{\mathrm{B}}$ are "big", can anything interesting still be said for a "not too big" k? On the other hand, these bounds valid for any k-extendible state are known to be close from optimal (there are examples of k-extendible states whose closest separable state is at distance of order d_{B}/k in 1-norm or of order $\sqrt{\ln d_{\mathrm{A}}/k}$ in $\mathbf{LOCC}^{\rightarrow}$-norm). Consequently, one may only hope to make stronger statements about average behaviours.

This is precisely the general question we address here, being especially interested in the case of high-dimensional bipartite quantum systems. We try and quantify in two distinct ways the typical efficiency of the k-extendibility criterion for separability in this asymptotic regime.

The first approach consists in estimating a specific size parameter (known as the *mean width*) of the set of k-extendible states when the dimension of the underlying Hilbert space goes to infinity. Comparing the obtained value with the known asymptotic estimate for the mean width of the set of separable states then tells us how the sizes of these two sets of states scale with one another. The computation is carried out in Section 9.2 (where all needed notions related to high-dimensional convex geometry are properly defined as well) and ends with the concluding Theorem 9.2.5, some technical parts being relegated to Appendix 9.13. In Section 9.3, the result is commented and comparisons are made between the mean-widths of, on the one hand, k-extendible states, and on the other, separable or PPT states. Besides, a smaller upper bound is derived, in Section 9.4 and its companion technical Appendix 9.14, on the mean width of the set of k-extendible states whose extension is required to be PPT (precise definitions and motivations to look at this set of states appear there).

The second approach consists in looking at random mixed states which are obtained by partial tracing over an ancilla space a uniformly distributed pure state, and characterizing when these are, with overwhelming probability as the dimension of the system grows, k-extendible or not. Again, comparing the obtained result

with the known one for separability provides some information on how powerful the k-extendibility test is to detect entanglement. Section 9.5 introduces all required material regarding the considered model of random-induced states and one possible way of detecting their non-k-extendibility. The adopted strategy is next seen through in Section 9.6, relying on technical statements put in Appendix 9.15, and concludes as Theorem 9.6.4. The determined environment dimension below which random-induced states are with high probability violating the k-extendibility criterion is then compared, in Section 9.7, with the previously established ones for violating other separability criteria, and for actually not being separable.

Finally, generalizations to the unbalanced case are stated in Section 9.8 (Theorems 9.8.1 and 9.8.2), while Section 9.9 exposes miscellaneous concluding remarks and loose ends.

The reader may have a look at Table 9.1 for a sample corollary of this study.

Table 9.1: Comparison of the average and typical case performance of the k-extendibility criterion with that of the PPT and realignment criteria

from the point of view of / k-extendibility "beats"	mean width of the set	entanglement detection of random states
PPT	for $k \geqslant 11$	for $k \geqslant 17$
realignment	?	for $k \geqslant 5$

Appendix 9.10 gathers a bunch of standard definitions and facts about the combinatorics of permutations and partitions which are necessary for our purposes. All employed notation on that matter are also introduced there. In Appendix 9.11, the connection is made between computing moments of GUE or Wishart matrices and counting permutations having a certain genus. These general observations play a key role in the moments' derivations of Appendices 9.13, 9.14 and 9.15, which are, as for them, specifically the ones that we need to obtain our various statements. To get tractable expressions, though, a formula relating the number of cycles in some specific permutations is additionally required, whose proof is detailed in Appendix 9.12. Appendix 9.16, finally, is devoted to establishing the last crucial ingredient in most of our reasonings, namely bounding the number of non-geodesic permutations (in terms of the number of geodesic ones) in some particular instances which are of interest to us. Aside, Appendix 9.17 is dedicated to proving more precise results than the ones which are strictly needed on the convergence of the studied random matrix ensembles, and in generalizing the developed method to establish the asymptotic freeness of certain gaussian random matrices.

Notation

Beside the general notation specified in Chapter 1, Section 1.3 (Hermitian and positive semidefinite operators on a Hilbert space, Schatten p-norm etc.). there are a few more specific ones that we will use repeatedly in the remainder of this chapter. Here, we will in fact always consider the case where the state space H of interest is finite-dimensional and bipartite, i.e. $\mathrm{H} = \mathrm{A} \otimes \mathrm{B}$, with A, B finite-dimensional Hilbert spaces. And we introduce the additional notation: $\mathscr{D}(\mathrm{A} \otimes \mathrm{B})$ for the set of all states on H, $\mathcal{S}(\mathrm{A{:}B})$ for the set of separable states on H, $\mathscr{P}(\mathrm{A{:}B})$ for the set of PPT states on H (in both cases in the cut A:B), and for each $k \in \mathbf{N}$, $\mathscr{C}_k(\mathrm{A{:}B})$ for the set of k-extendible states on H (in the cut A:B and with respect to B). What is more, we will actually almost exclusively deal with the balanced case, i.e. when $\mathrm{A} \equiv \mathrm{B} \equiv \mathbf{C}^d$.

Preliminary technical lemma

It will be essential for us in the sequel to express in a more tractable way the quantity $\sup_{\sigma \in \mathscr{C}_k(\mathrm{A:B})} \mathrm{Tr}(M\sigma)$, for any given Hermitian M on $\mathrm{A} \otimes \mathrm{B}$. Such amenable expression is provided by Lemma 9.1.3 below.

Lemma 9.1.3. *Let $k \in \mathbf{N}$. For any $M_{\mathrm{AB}} \in \mathscr{H}(\mathrm{A} \otimes \mathrm{B})$, we have*

$$\sup_{\sigma_{\mathrm{AB}} \in \mathscr{C}_k(\mathrm{A:B})} \mathrm{Tr}\left(M_{\mathrm{AB}} \sigma_{\mathrm{AB}}\right) = \left\| \frac{1}{k} \sum_{j=1}^{k} \widetilde{M}_{\mathrm{AB}^k}(j) \right\|_{\infty},$$

where for each $1 \leqslant j \leqslant k$*, denoting by* $\mathrm{Id}_{\widehat{\mathrm{B}}_j^k}$ *the identity on* $\mathrm{B}_1 \otimes \cdots \otimes \mathrm{B}_{j-1} \otimes \mathrm{B}_{j+1} \otimes \cdots \otimes \mathrm{B}_k$*, we defined* $\widetilde{M}_{\mathrm{AB}^k}(j) = M_{\mathrm{AB}_j} \otimes \mathrm{Id}_{\widehat{\mathrm{B}}_j^k}$.

Before proving Lemma 9.1.3, let us introduce once and for all the following notation, which we shall later use on several occasions: for any $M_{\mathrm{AB}^k} \in \mathscr{H}(\mathrm{A} \otimes \mathrm{B}^{\otimes k})$, we define its symmetrisation with respect to $\mathrm{B}^{\otimes k}$ as

$$\mathrm{Sym}_{\mathrm{A:B}^k}(M_{\mathrm{AB}^k}) = \frac{1}{k!} \sum_{\pi \in \mathfrak{S}(k)} \big(\mathrm{Id}_{\mathrm{A}} \otimes U(\pi)_{\mathrm{B}^k}\big) M_{\mathrm{AB}^k} \big(\mathrm{Id}_{\mathrm{A}} \otimes U(\pi)_{\mathrm{B}^k}\big)^\dagger,$$

where for each permutation $\pi \in \mathfrak{S}(k)$, $U(\pi)_{\mathrm{B}^k}$ denotes the associated permutation unitary on $\mathrm{B}^{\otimes k}$ (see e.g. [93] and Chapter 3, Section 3.1, of this manuscript for further details).

Proof. By definition, the condition $\sigma_{\mathrm{AB}} \in \mathscr{C}_k(\mathrm{A:B})$ is equivalent to the condition $\sigma_{\mathrm{AB}} = \mathrm{Tr}_{\mathrm{B}^{k-1}} \mathrm{Sym}_{\mathrm{A:B}^k}(\sigma_{\mathrm{AB}^k})$ for some $\sigma_{\mathrm{AB}^k} \in \mathscr{D}(\mathrm{A} \otimes \mathrm{B}^{\otimes k})$. Hence, for any $M_{\mathrm{AB}} \in \mathscr{H}(\mathrm{A} \otimes \mathrm{B})$, we have

$$\begin{aligned}
\sup_{\sigma_{\mathrm{AB}} \in \mathscr{C}_k(\mathrm{A:B})} \mathrm{Tr}_{\mathrm{AB}} \left[M_{\mathrm{AB}} \sigma_{\mathrm{AB}}\right] &= \sup_{\sigma_{\mathrm{AB}^k} \in \mathscr{D}(\mathrm{A} \otimes \mathrm{B}^{\otimes k})} \mathrm{Tr}_{\mathrm{AB}} \left[M_{\mathrm{AB}} \, \mathrm{Tr}_{\mathrm{B}^{k-1}} \mathrm{Sym}_{\mathrm{A:B}^k}(\sigma_{\mathrm{AB}^k})\right] \\
&= \sup_{\sigma_{\mathrm{AB}^k} \in \mathscr{D}(\mathrm{A} \otimes \mathrm{B}^{\otimes k})} \mathrm{Tr}_{\mathrm{AB}^k} \left[\left(M_{\mathrm{AB}} \otimes \mathrm{Id}_{\mathrm{B}^{k-1}}\right) \mathrm{Sym}_{\mathrm{A:B}^k}(\sigma_{\mathrm{AB}^k})\right] \\
&= \sup_{\sigma_{\mathrm{AB}^k} \in \mathscr{D}(\mathrm{A} \otimes \mathrm{B}^{\otimes k})} \mathrm{Tr}_{\mathrm{AB}^k} \left[\mathrm{Sym}_{\mathrm{A:B}^k} \left(M_{\mathrm{AB}} \otimes \mathrm{Id}_{\mathrm{B}^{k-1}}\right) \sigma_{\mathrm{AB}^k}\right] \\
&= \left\| \mathrm{Sym}_{\mathrm{A:B}^k} \left(M_{\mathrm{AB}} \otimes \mathrm{Id}_{\mathrm{B}^{k-1}}\right) \right\|_\infty .
\end{aligned}$$

Now, for each $\pi \in \mathfrak{S}(k)$, $\big(\mathrm{Id}_{\mathrm{A}} \otimes U(\pi)_{\mathrm{B}^k}\big)\big(M_{\mathrm{AB}} \otimes \mathrm{Id}_{\mathrm{B}^{k-1}}\big)\big(\mathrm{Id}_{\mathrm{A}} \otimes U(\pi)_{\mathrm{B}^k}\big)^\dagger = M_{\mathrm{AB}_{\pi(1)}} \otimes \mathrm{Id}_{\widehat{\mathrm{B}}_{\pi(1)}^k}$. Therefore, grouping together, for each $1 \leqslant j \leqslant k$, the permutations $\pi \in \mathfrak{S}(k)$ such that $\pi(1) = j$, we get

$$\mathrm{Sym}_{\mathrm{A:B}^k} \left(M_{\mathrm{AB}} \otimes \mathrm{Id}_{\mathrm{B}^{k-1}}\right) = \frac{1}{k} \sum_{j=1}^{k} M_{\mathrm{AB}_j} \otimes \mathrm{Id}_{\widehat{\mathrm{B}}_j^k},$$

and hence the advertised result. □

9.2 Mean width of the set of k-extendible states for "small" k

9.2.1 Preliminaries on convex geometry

Let us introduce a few notions coming from classical convex geometry which we shall need in the sequel. Much more details on that matter appear in Chapter 4, Section 4.1. For any $K \subset \mathscr{H}(\mathbf{C}^n)$ and any $M \in \mathscr{H}(\mathbf{C}^n)$ having unit Hilbert–Schmidt norm, we define the width of K in the direction M as

$$w(K, M) = \sup_{\Delta \in K} \mathrm{Tr}(M\Delta).$$

The mean width of K is then defined as the average of $w(K, \cdot)$ over the whole Hilbert–Schmidt unit sphere $S_{HS}(\mathbf{C}^n)$ of $\mathscr{H}(\mathbf{C}^n)$ (equipped with the Haar measure σ) i.e.

$$w(K) = \int_{M \in S_{HS}(\mathbf{C}^n)} w(K, M) \mathrm{d}\sigma(M) = \int_{M \in S_{HS}(\mathbf{C}^n)} \left[\sup_{\Delta \in K} \mathrm{Tr}(M\Delta)\right] \mathrm{d}\sigma(M).$$

This average width w is an interesting size parameter, on its own, but also because it is related to other important geometric quantities, such as e.g. the volume radius vrad, which is defined as the radius of the Euclidean ball having same volume (i.e. Lebesgue measure). For instance, we have for any convex body K the Urysohn inequality $w(K) \geqslant \mathrm{vrad}(K)$, and for most of the convex bodies K we shall be considering a "reverse" Urysohn inequality $w(K) \leqslant \mu \, \mathrm{vrad}(K)$ for some $\mu \geqslant 1$. These connections and the precise formulation of these convex geometry results are exemplified in Section 9.3.

In order to compute the quantity $w(K)$, it is often convenient to re-express it as a Gaussian rather than spherical averaging. We thus denote by $GUE(n)$ the Gaussian Unitary Ensemble on $\mathbf{C}^n$, which is the standard Gaussian vector in $\mathscr{H}(\mathbf{C}^n)$ (equivalently, $G \sim GUE(n)$ if $G = (H + H^\dagger)/\sqrt{2}$ with H a $n \times n$ matrix having independent complex normal entries). And we define the Gaussian mean width of K as

$$w_G(K) = \mathbf{E}_{G \sim GUE(n)} \left[\sup_{\Delta \in K} \mathrm{Tr}(G\Delta)\right].$$

Just observing that for $G \sim GUE(n)$, $G/\|G\|_2$ is uniformly distributed over $S_{HS}(\mathbf{C}^n)$, and $G/\|G\|_2$, $\|G\|_2$ are independent random variables, we get that the link between both quantities is, setting $\gamma(n) = \mathbf{E}_{G\sim GUE(n)} \|G\|_2$, which is known to satisfy $\gamma(n) \sim_{n\to+\infty} n$ (see e.g. [4], Chapter 2, for a proof),

$$w(K) = \frac{1}{\gamma(n)} w_G(K). \tag{9.1}$$

Remark 9.2.1. *All the sets K that we will consider in the sequel will actually be subsets of $\mathcal{D}(\mathbf{C}^n)$, hence living in the hyperplane of $\mathcal{H}(\mathbf{C}^n)$ composed of trace 1 elements, i.e. in a space of real dimension n^2-1, rather than n^2. It would thus seem more natural to define their mean width $w(K)$ as an average width over a n^2-2, rather than n^2-1, dimensional Euclidean unit sphere. The Gaussian mean width $w_G(K)$, on the other hand, is an intrinsic notion that does not depend on the ambient dimension (because marginals of standard Gaussian vectors are themselves standard Gaussian vectors). As a consequence, we see from equation* (9.1) *that computing the mean width of K as if it was a n^2 dimensional set is asymptotically equivalent to computing it taking into account that it is in fact a n^2-1 dimensional set. We may therefore serenely forget about this issue.*

Our aim is now to estimate, for any fixed $k \in \mathbf{N}$, the mean width of the set of k-extendible states on $\mathrm{A}\otimes\mathrm{B}$ when $\mathrm{A} \equiv \mathrm{B} \equiv \mathbf{C}^d$ and $d \to +\infty$. By the definitions above, we have

$$w\big(\mathscr{E}_k(\mathrm{A}{:}\mathrm{B})\big) = \frac{1}{\gamma(d^2)} \mathbf{E}_{G_{\mathrm{AB}}\sim GUE(d^2)} \left[\sup_{\sigma_{\mathrm{AB}}\in\mathscr{E}_k(\mathrm{A}:\mathrm{B})} \mathrm{Tr}\big(G_{\mathrm{AB}}\sigma_{\mathrm{AB}}\big) \right].$$

Using the result of Lemma 9.1.3, and the notation introduced there, we thus get,

$$w\big(\mathscr{E}_k(\mathrm{A}{:}\mathrm{B})\big) = \frac{1}{\gamma(d^2)} \mathbf{E}_{G_{\mathrm{AB}}\sim GUE(d^2)} \left\| \frac{1}{k} \sum_{j=1}^{k} \widetilde{G}_{\mathrm{AB}^k}(j) \right\|_\infty . \tag{9.2}$$

9.2.2 An operator-norm estimate

As justified above, to obtain the mean width of the set of k-extendible states on $\mathbf{C}^d \otimes \mathbf{C}^d$, what we need is to compute the average operator-norm $\mathbf{E} \left\| \sum_{j=1}^k \widetilde{G}_{\mathrm{AB}^k}(j) \right\|_\infty$, for G a GUE matrix on $\mathbf{C}^d \otimes \mathbf{C}^d$. We will show that the following asymptotic estimate holds.

Proposition 9.2.2. *Fix $k \in \mathbf{N}$. Then,*

$$\mathbf{E}_{G_{\mathrm{AB}}\sim GUE(d^2)} \left\| \sum_{j=1}^{k} \widetilde{G}_{\mathrm{AB}^k}(j) \right\|_\infty \underset{d\to+\infty}{\sim} 2\sqrt{k}d.$$

As a preliminary step towards estimating the sup-norm $\mathbf{E} \left\| \sum_{j=1}^k \widetilde{G}_{\mathrm{AB}^k}(j) \right\|_\infty$, we will look at the $2p$-order moments $\mathbf{E}\,\mathrm{Tr}\left[\left(\sum_{j=1}^k \widetilde{G}_{\mathrm{AB}^k}(j) \right)^{2p} \right]$, $p \in \mathbf{N}$, and show that they can be expressed in terms of the $2p$-order moments of a centered semicircular distribution of appropriate parameter.

So let us recall first a few required definitions. For any $\sigma > 0$, we shall denote by $\mu_{SC(\sigma^2)}$ the centered semicircular distribution of variance parameter σ^2, whose density is given by

$$\mathrm{d}\mu_{SC(\sigma^2)}(x) = \frac{1}{2\pi\sigma^2} \sqrt{4\sigma^2 - x^2}\, \mathbf{1}_{[-2\sigma,2\sigma]}(x)\mathrm{d}x.$$

We shall also denote, for each $p \in \mathbf{N}$, by $\mathrm{M}^{(p)}_{SC(\sigma^2)}$ its p-order moment, i.e. $\mathrm{M}^{(p)}_{SC(\sigma^2)} = \int_{-\infty}^{+\infty} x^p \mathrm{d}\mu_{SC(\sigma^2)}(x)$. It is well-known that

$$\forall\, p \in \mathbf{N},\ \mathrm{M}^{(2p-1)}_{SC(\sigma^2)} = 0 \text{ and } \mathrm{M}^{(2p)}_{SC(\sigma^2)} = \sigma^{2p}\,\mathrm{Cat}_p,$$

where Cat_p is the pth Catalan number defined in Lemma 9.10.4.

Proposition 9.2.3. *Fix $k \in \mathbf{N}$. Then, when $d \to +\infty$, the random matrix $\left(\sum_{j=1}^k \widetilde{G}_{\mathrm{AB}^k}(j) \right)/d$ converges in moments towards a centered semicircular distribution of parameter k. Equivalently, this means that, for any $p \in \mathbf{N}$,*

$$\mathbf{E}_{G_{\mathrm{AB}}\sim GUE(d^2)}\, \mathrm{Tr} \left[\left(\sum_{j=1}^{k} \widetilde{G}_{\mathrm{AB}^k}(j) \right)^{2p} \right] \underset{d\to+\infty}{\sim} \mathrm{M}^{(2p)}_{SC(k)} d^{2p+k+1}.$$

Remark 9.2.4. *Stronger convergence results than the one established in Proposition 9.2.3 may in fact be proved, as discussed in Appendix 9.17.*

Proof of Proposition 9.2.3. Let $p \in \mathbf{N}$. Computing the value of the $2p$-order moment $\mathbf{E}\,\mathrm{Tr}\left[\left(\sum_{j=1}^{k} \widetilde{G}_{\mathrm{AB}^k}(j)\right)^{2p}\right]$ may be done using the Gaussian Wick formula (see Lemma 9.11.1 for the statement and Appendix 9.11.1 for a succinct summary of how to derive moments of GUE matrices from it). In our case, what we get by the computations carried out in Appendix 9.13 and summarized in Proposition 9.13.1 is that, for any $d \in \mathbf{N}$,

$$\mathbf{E}\,\mathrm{Tr}\left[\left(\sum_{j=1}^{k} \widetilde{G}_{\mathrm{AB}^k}(j)\right)^{2p}\right] = \sum_{f:[2p]\to[k]} \mathbf{E}\,\mathrm{Tr}\left[\overrightarrow{\prod_{i=1}^{2p}} \widetilde{G}_{\mathrm{AB}^k}(f(i))\right] \tag{9.3}$$

$$= \sum_{f:[2p]\to[k]} \sum_{\lambda\in\mathfrak{P}^{(2)}(2p)} d^{\sharp(\gamma^{-1}\lambda)+\sharp(\gamma_f^{-1}\lambda)+k-|\,\mathrm{Im}(f)|}, \tag{9.4}$$

where we defined on $\{1,\ldots,2p\}$, $\gamma = (2p\ \ldots\ 1)$ as the canonical full cycle, and for each $f : [2p] \to [k]$, $\gamma_f = \gamma_{f=1}\cdots\gamma_{f=k}$ as the product of the canonical full cycles on each of the level sets of f.

We now have to understand which $\lambda \in \mathfrak{P}^{(2)}(2p)$ and $f : [2p] \to [k]$ contribute to the dominating term in the moment expansion (9.4), i.e. are such that the quantity $\sharp(\gamma^{-1}\lambda) + \sharp(\gamma_f^{-1}\lambda) + k - |\,\mathrm{Im}(f)|$ is maximal.

First of all, for any $\lambda \in \mathfrak{P}^{(2)}(2p)$, we have

$$\sharp(\lambda) + \sharp(\gamma^{-1}\lambda) = 4p - \big(|\lambda| + |\gamma^{-1}\lambda|\big) \leqslant 4p - |\gamma^{-1}| = 4p - (2p-1) = 2p+1, \tag{9.5}$$

where the first equality is by Lemma 9.10.1, while the second inequality is by equation (9.27) in Lemma 9.10.5 and is an equality if and only if the pair-partition λ is non-crossing. Next, for any $\lambda \in \mathfrak{P}^{(2)}(2p)$ and $f : [2p] \to [k]$, we have

$$\sharp(\lambda) + \sharp(\gamma_f^{-1}\lambda) = 4p - \Big(|\lambda| + |\gamma_f^{-1}\lambda|\Big) \leqslant 4p - |\gamma_f^{-1}| = 4p - (2p - |\,\mathrm{Im}(f)|) = 2p + |\,\mathrm{Im}(f)|, \tag{9.6}$$

where the first equality is again by Lemma 9.10.1, while the second inequality is by equation (9.28) in Lemma 9.10.5 and is an equality if and only if the pair-partition λ is non-crossing and is finer than the partition of $\{1,\ldots,2p\}$ induced by γ_f (i.e. f takes the same value on elements belonging to the same pair-block of λ).

Putting equations (9.5) and (9.6) together, we get that for any $\lambda \in \mathfrak{P}^{(2)}(2p)$ and $f : [2p] \to [k]$ (just keeping in mind that necessarily $\sharp(\lambda) = p$),

$$\sharp(\gamma^{-1}\lambda) + \sharp(\gamma_f^{-1}\lambda) + k - |\,\mathrm{Im}(f)| \leqslant 2p + k + 1, \tag{9.7}$$

with equality if and only if $\lambda \in NC^{(2)}(2p)$ and $f \circ \lambda = f$. Since it is well-known that there are Cat_p elements in $NC^{(2)}(2p)$, and for each of these there are k^p functions which are constant on each of its p pair-blocks, we indeed get the asymptotic estimate announced in Proposition 9.2.3, namely

$$\mathbf{E}\,\mathrm{Tr}\left[\left(\sum_{j=1}^{k} \widetilde{G}_{\mathrm{AB}^k}(j)\right)^{2p}\right] \underset{d\to+\infty}{\sim} k^p\,\mathrm{Cat}_p\,d^{2p+k+1} = \mathrm{M}_{SC(k)}^{(2p)} d^{2p+k+1}. \qquad \square$$

Proof of Proposition 9.2.2. The convergence in moments stated in Proposition 9.2.3 implies that the matrix $\left(\sum_{j=1}^{k} \widetilde{G}_{\mathrm{AB}^k}(j)\right)/d$ asymptotically has a largest eigenvalue which is, on average, at least the upper-edge of the support of $\mu_{SC(k)}$, i.e. $2\sqrt{k}$. In other words, it guarantees that there exist positive constants $c_d \to_{d\to+\infty} 1$ such that

$$\mathbf{E}\left\|\sum_{j=1}^{k} \widetilde{G}_{\mathrm{AB}^k}(j)\right\|_\infty \geqslant c_d\,2\sqrt{k}d. \tag{9.8}$$

In the opposite direction, Proposition 9.2.3 only guarantees that the matrix $\left(\sum_{j=1}^{k} \widetilde{G}_{\mathrm{AB}^k}(j)\right)/d$ asymptotically has, on average, no strictly positive fraction of eigenvalues strictly above $2\sqrt{k}$. So to show that the reverse inequality to (9.8) holds too, a little more care is required. Indeed, to say it roughly, we have to make sure that in the moment's expression (9.4), the permutations contributing to the non-dominating terms (in d) are not too numerous.

For $d \in \mathbf{N}$ fixed, it holds that

$$\forall\ p \in \mathbf{N},\ \mathbf{E}\left\|\sum_{j=1}^{k} \widetilde{G}_{\mathrm{AB}^k}(j)\right\|_\infty \leqslant \left(\mathbf{E}\,\mathrm{Tr}\left[\left(\sum_{j=1}^{k} \widetilde{G}_{\mathrm{AB}^k}(j)\right)^{2p}\right]\right)^{1/2p}. \tag{9.9}$$

So let us fix $d \in \mathbf{N}$ and $p \in \mathbf{N}$, and rewrite (9.4) explicitly as an expansion in powers of d, keeping in the sum the permutations not saturating equation (9.7). Being cautious only with the permutations not saturating equation (9.6), and not with those not saturating equation (9.5), we get

$$\mathbf{E}\,\mathrm{Tr}\left[\left(\sum_{j=1}^{k} \widetilde{G}_{\mathrm{AB}^k}(j)\right)^{2p}\right] \leqslant \left(\sum_{f:[2p]\to[k]} \sum_{\delta=0}^{\lfloor (p+k)/2\rfloor} \left|\mathfrak{P}^{(2)}_{f,\delta}(2p)\right| d^{-2\delta}\right) d^{2p+k+1}, \tag{9.10}$$

where we defined, for each $f : [2p] \to [k]$ and each $0 \leqslant \delta \leqslant \lfloor (p+k)/2 \rfloor$,

$$\mathfrak{P}^{(2)}_{f,\delta}(2p) = \left\{\lambda \in \mathfrak{P}^{(2)}(2p)\ :\ \sharp(\gamma_f^{-1}\lambda) = p + |\,\mathrm{Im}(f)| - 2\delta\right\}.$$

In words, $\mathfrak{P}^{(2)}_{f,\delta}(2p)$ is nothing else than the set of permutations which have a defect 2δ from lying on the geodesics between the identity and the product of the canonical full cycles on each of the level sets of f. This justifies in particular *a posteriori* why the summation in (9.10) is only over even defects (see the parity argument in Lemma 9.10.2).

Now, by Lemma 9.16.3, we know that, if $0 \leqslant \delta \leqslant \lfloor p/2 \rfloor$, then

$$\left|\left\{(f,\lambda)\ :\ \lambda \in \mathfrak{P}^{(2)}_{f,\delta}(2p)\right\}\right| \leqslant k^p\,\mathrm{Cat}_p \times \left(\frac{kp^2}{2}\right)^{2\delta}.$$

And if $\lceil p/2 \rceil \leqslant \delta \leqslant \lfloor (p+k)/2 \rfloor$, then trivially

$$\left|\left\{(f,\lambda)\ :\ \lambda \in \mathfrak{P}^{(2)}_{f,\delta}(2p)\right\}\right| \leqslant k^{2p}\,\frac{(2p)!}{2^p p!} \leqslant k^p\,\mathrm{Cat}_p \times \left(\frac{kp^2}{2}\right)^{p}.$$

Putting everything together, we therefore get,

$$\mathbf{E}\,\mathrm{Tr}\left[\left(\sum_{j=1}^{k} \widetilde{G}_{\mathrm{AB}^k}(j)\right)^{2p}\right] \leqslant k^p\,\mathrm{Cat}_p \left(1 + \sum_{\delta=1}^{\lfloor p/2\rfloor} \left(\frac{kp^2}{2d}\right)^{2\delta} + \frac{k}{2}\left(\frac{kp^2}{2d}\right)^{p}\right) d^{2p+k+1}.$$

Yet, $\max\left\{\left(kp^2/2d\right)^{2\delta}\ :\ 1 \leqslant \delta \leqslant \lceil p/2 \rceil\right\}$ is attained for $\delta = 1$, provided $p \leqslant (2d/k)^{1/2}$. So if such is the case,

$$\sum_{\delta=1}^{\lfloor p/2\rfloor} \left(\frac{kp^2}{2d}\right)^{2\delta} + \frac{k}{2}\left(\frac{kp^2}{2d}\right)^{p} \leqslant \frac{p+k}{2}\frac{k^2p^4}{4d^2} \leqslant \frac{k^2p^5}{4d^2},$$

where the last inequality holds as long as $p \geqslant k$. And hence, under all the previous assumptions,

$$\mathbf{E}\,\mathrm{Tr}\left[\left(\sum_{j=1}^{k} \widetilde{G}_{\mathrm{AB}^k}(j)\right)^{2p}\right] \leqslant \mathrm{M}^{(2p)}_{SC(k)}\left(1 + \frac{k^2p^5}{4d^2}\right) d^{2p+k+1}.$$

So set $p_d = (2d/k)^{(2-\epsilon)/5}$ for some $0 < \epsilon < 1$ (which is indeed smaller than $(2d/k)^{1/2}$ and bigger than k for d big enough, in particular bigger than $k^{7/2}/2$). And using inequality (9.9) in the special case $p = p_d$, we eventually get

$$\mathbf{E}\left\|\sum_{j=1}^{k} \widetilde{G}_{\mathrm{AB}^k}(j)\right\|_\infty \leqslant \left(\mathrm{M}^{(2p_d)}_{SC(k)}\left(1 + \frac{k^2p_d^5}{4d^2}\right)\right)^{1/2p_d} d^{1+(k+1)/2p_d} \underset{d\to+\infty}{\sim} 2\sqrt{k}d. \tag{9.11}$$

Combining the lower bound in equation (9.8) and the upper bound in equation (9.11) yields Proposition 9.2.2. □

9.2.3 Conclusion

Combining Proposition 9.2.2 with equation (9.2), we straightforwardly obtain the estimate we were looking for, which is stated in Theorem 9.2.5 below.

Theorem 9.2.5. *Let $k \in \mathbf{N}$. The mean width of the set of k-extendible states on $\mathbf{C}^d \otimes \mathbf{C}^d$ satisfies*

$$w\big(\mathscr{E}_k(\mathbf{C}^d{:}\mathbf{C}^d)\big) \underset{d\to+\infty}{\sim} \frac{2}{\sqrt{k}}\frac{1}{d}.$$

9.3 Discussion and comparison with the mean width of the set of PPT states

It was shown in [16] that the mean width of the set of separable states on $\mathbf{C}^d \otimes \mathbf{C}^d$ is of order $1/d^{3/2}$. And we just showed in Theorem 9.2.5 that, for $k \in \mathbf{N}$ fixed, the mean width of the set of k-extendible states on $\mathbf{C}^d \otimes \mathbf{C}^d$ is of order $1/d$, so that, for d large,

$$w\big(\mathscr{S}(\mathbf{C}^d{:}\mathbf{C}^d)\big) \ll w\big(\mathscr{E}_k(\mathbf{C}^d{:}\mathbf{C}^d)\big).$$

This result is not surprising: it just means that, when d grows, if k does not grow in some way too, then the set of k-extendible states becomes an increasingly poor approximation of the set of separable states on $\mathbf{C}^d \otimes \mathbf{C}^d$. There had been several evidences, already, in that direction, with examples of highly-extendible, though entangled, states (see e.g. [34] and [143]).

It is well-known that the exact same feature is actually exhibited by the set of PPT states on $\mathbf{C}^d \otimes \mathbf{C}^d$, whose mean width is of order $1/d$ too. Let us be more precise.

Proposition 9.3.1. *There exist positive constants $c_d, C_d \to_{d\to+\infty} 1$ such that the mean width of the set of PPT states on $\mathbf{C}^d \otimes \mathbf{C}^d$ satisfies*

$$c_d\,\frac{e^{-1/2}}{d} \leqslant w\big(\mathscr{P}(\mathbf{C}^d{:}\mathbf{C}^d)\big) \leqslant C_d\,\frac{2}{d}.$$

Proof. Proposition 9.3.1 was basically established in [16], but not stated in this exact way and with these exact constants, so we briefly recall the argument here for the sake of completeness (see also Chapter 8, Section 8.5, of the present manuscript for a similar discussion).

To get the asymptotic upper bound, we just use

$$w\big(\mathscr{P}(\mathbf{C}^d{:}\mathbf{C}^d)\big) \leqslant w\big(\mathscr{D}(\mathbf{C}^d \otimes \mathbf{C}^d)\big) \underset{d\to+\infty}{\sim} \frac{2}{d}.$$

The last equivalence is a consequence of Wigner's semicircle law (see e.g. [4], Chapter 2, for a proof) from which it follows that

$$w_G\big(\mathscr{D}(\mathbf{C}^d \otimes \mathbf{C}^d)\big) = \mathbf{E}_{G\sim GUE(d^2)} \sup_{\sigma\in\mathscr{D}(\mathbf{C}^d\otimes\mathbf{C}^d)} \mathrm{Tr}(G\sigma) = \mathbf{E}_{G\sim GUE(d^2)}\,\|G\|_\infty \underset{d\to+\infty}{\sim} 2d.$$

To get the asymptotic lower bound, we will make use of two results from classical convex geometry: Urysohn and Milman–Pajor inequalities (see Theorems 4.1.3 and 4.1.6, respectively, in Chapter 4, Section 4.1). But before that, we need one more definition (also appearing in Chapter 4, Section 4.1): For any convex body K, we denote by $\mathrm{vrad}(K)$ its volume radius, which is defined as the radius of the Euclidean ball having the same volume (i.e. Lebesgue measure) as K. In our case, denoting by Γ the partial transposition, we have $\mathscr{P}(\mathbf{C}^d{:}\mathbf{C}^d) = \mathscr{D}(\mathbf{C}^d \otimes \mathbf{C}^d) \cap \mathscr{D}(\mathbf{C}^d \otimes \mathbf{C}^d)^\Gamma$, with $\mathscr{D}(\mathbf{C}^d \otimes \mathbf{C}^d)$ having the maximally mixed state Id/d^2 as center of gravity and Γ being a linear isometry. Hence,

$$w\big(\mathscr{P}(\mathbf{C}^d{:}\mathbf{C}^d)\big) \geqslant \mathrm{vrad}\left(\mathscr{P}(\mathbf{C}^d{:}\mathbf{C}^d)\right) \geqslant \frac{\mathrm{vrad}\left(\mathscr{D}(\mathbf{C}^d \otimes \mathbf{C}^d)\right)^2}{2w\big(\mathscr{D}(\mathbf{C}^d \otimes \mathbf{C}^d)\big)},$$

the first inequality being by the Urysohn inequality, and the second being by the consequence of Milman–Pajor inequality stated as Corollary 4.1.7 in Chapter 4, Section 4.1. Now, we just argued that $w\big(\mathscr{D}(\mathbf{C}^d \otimes \mathbf{C}^d)\big) \sim_{d\to+\infty} 2/d$, while it was shown in [167] that $\mathrm{vrad}\left(\mathscr{D}(\mathbf{C}^d \otimes \mathbf{C}^d)\right) \sim_{d\to+\infty} e^{-1/4}/d$. Therefore,

$$w\big(\mathscr{P}(\mathbf{C}^d{:}\mathbf{C}^d)\big) \geqslant \frac{\mathrm{vrad}\left(\mathscr{D}(\mathbf{C}^d \otimes \mathbf{C}^d)\right)^2}{2w\big(\mathscr{D}(\mathbf{C}^d \otimes \mathbf{C}^d)\big)} \underset{d\to+\infty}{\sim} \frac{e^{-1/2}}{d}.$$

□

As a straightforward consequence of Theorem 9.2.5 and Proposition 9.3.1, we have, roughly speaking, that for $k \geqslant 11$, the set of k-extendible states becomes asymptotically a "better" approximation of the set of separable states than the set of PPT states, on average. Indeed, if $k \geqslant 11$, then $2/\sqrt{k} < e^{-1/2}$, so that for d large enough

$$w\big(\mathscr{E}_k(\mathbf{C}^d{:}\mathbf{C}^d)\big) < w\big(\mathscr{P}(\mathbf{C}^d{:}\mathbf{C}^d)\big).$$

9.4 Adding the PPT constraint on the extension

The hierarchy of SDPs originally proposed in [66] to detect entanglement was in fact slightly different from the one that would be derived from Theorem 9.1.2. Indeed, for a given bipartite state ρ_{AB}, the kth test would here consist in looking for a symmetric extension ρ_{AB^k} of ρ_{AB}, while in [66] it was additionally imposed that this extension had to be PPT in any cut of the $k+1$ subsystems. This of course increased quite considerably the size of the SDP to be solved at each step, but with the hope that it would at the same time decrease dramatically the number of steps an entangled state would pass.

Another hierarchy of SDPs was later proposed in [143] and [142], built on the exact same ideas as those in [66]. It was noticed there that only demanding that the (Bose) symmetric extension of the state be PPT in one fixed (even) cut of the $k+1$ subsystems already implied a noticeable speed-up in the convergence of the algorithm. It therefore seems worth taking a closer look at the set of states arising from these constraints. The latter is properly defined as follows.

Definition 9.4.1. *Let $k \in \mathbf{N}$. A state ρ_{AB} on a bipartite Hilbert space $\mathrm{A} \otimes \mathrm{B}$ is k-PPT-extendible with respect to B if there exists a state ρ_{AB^k} on $\mathrm{A} \otimes \mathrm{B}^{\otimes k}$ which is PPT in the cut $\mathrm{A} \otimes \mathrm{B}^{\otimes \lfloor k/2 \rfloor} : \mathrm{B}^{\otimes \lceil k/2 \rceil}$, invariant under any permutation of the B subsystems and such that $\rho_{\mathrm{AB}} = \mathrm{Tr}_{\mathrm{B}^{k-1}} \rho_{\mathrm{AB}^k}$. We denote by $\mathscr{E}_k^{PPT}(\mathrm{A}{:}\mathrm{B})$ the set of k-PPT-extendible states on $\mathrm{A} \otimes \mathrm{B}$ (in the cut $\mathrm{A}{:}\mathrm{B}$ and with respect to B).*

Theorem 9.4.2. *Let $k \in \mathbf{N}$. There exist positive constants $C_d \to_{d\to+\infty} 1$ such that the mean width of the set of k-PPT-extendible states on $\mathbf{C}^d \otimes \mathbf{C}^d$ satisfies*

$$w\big(\mathscr{E}_k^{PPT}(\mathbf{C}^d{:}\mathbf{C}^d)\big) \leqslant C_d \, \frac{\sqrt{2}}{\sqrt{k}} \frac{1}{d}.$$

Proof. Using the notation introduced in Lemma 9.1.3, we start from the simple observation that, for any $M_{\mathrm{AB}} \in \mathscr{H}(\mathrm{A} \otimes \mathrm{B})$,

$$\begin{aligned} \sup_{\sigma_{\mathrm{AB}} \in \mathscr{E}_k^{PPT}(\mathrm{A}:\mathrm{B})} \mathrm{Tr}\left[M_{\mathrm{AB}} \sigma_{\mathrm{AB}}\right] &= \sup_{\sigma_{\mathrm{AB}^k} \in \mathscr{P}(\mathrm{AB}^{\lfloor k/2 \rfloor}:\mathrm{B}^{\lceil k/2 \rceil})} \mathrm{Tr}\left[\mathrm{Sym}_{\mathrm{A}:\mathrm{B}^k}\left(M_{\mathrm{AB}} \otimes \mathrm{Id}_{\mathrm{B}^{k-1}}\right)\sigma_{\mathrm{AB}^k}\right] \\ &\leqslant \min\left(\left\| \mathrm{Sym}_{\mathrm{A}:\mathrm{B}^k}\left(M_{\mathrm{AB}} \otimes \mathrm{Id}_{\mathrm{B}^{k-1}}\right)\right\|_\infty, \left\| \left[\mathrm{Sym}_{\mathrm{A}:\mathrm{B}^k}\left(M_{\mathrm{AB}} \otimes \mathrm{Id}_{\mathrm{B}^{k-1}}\right)\right]^\Gamma \right\|_\infty\right), \end{aligned}$$

where Γ stands here for the partial transposition over the $\lceil k/2 \rceil$ last B subsystems, so that in fact

$$\left[\mathrm{Sym}_{\mathrm{A}:\mathrm{B}^k}\left(M_{\mathrm{AB}} \otimes \mathrm{Id}_{\mathrm{B}^{k-1}}\right)\right]^\Gamma = \frac{1}{k}\left(\sum_{j=1}^{\lfloor k/2 \rfloor} M_{\mathrm{AB}_j} \otimes \mathrm{Id}_{\widehat{\mathrm{B}}_j^k} + \sum_{j=\lfloor k/2 \rfloor+1}^{k} M_{\mathrm{AB}_j}^\Gamma \otimes \mathrm{Id}_{\widehat{\mathrm{B}}_j^k}\right),$$

where Γ now stands for the partial transposition over B.

The upper bound in Theorem 9.4.2 will thus be a direct consequence of the sup-norm estimate

$$\mathbf{E}_{G_{\mathrm{AB}} \sim GUE(d^2)} \left\| \sum_{j=1}^{k} \widetilde{G}_{\mathrm{AB}^k}(j)^\Gamma \right\|_\infty \underset{d\to+\infty}{\sim} \sqrt{2}\sqrt{k}d.$$

The latter is proved in the exact same way as Proposition 9.2.2, i.e. by first showing that for any $p \in \mathbf{N}$,

$$\mathbf{E}_{G_{\mathrm{AB}} \sim GUE(d^2)} \mathrm{Tr}\left[\left(\sum_{j=1}^{k} \widetilde{G}_{\mathrm{AB}^k}(j)^\Gamma\right)^{2p}\right] \underset{d\to+\infty}{\sim} 2\mathrm{M}_{SC(k/2)}^{(2p)} d^{2p+k/2+1}, \tag{9.12}$$

and second arguing that also $\mathbf{E}\left\|\sum_{j=1}^{k} \widetilde{G}_{\mathrm{AB}^k}(j)^\Gamma\right\|_\infty \sim_{d\to+\infty} \lim_{p\to+\infty}\left(\mathbf{E}\,\mathrm{Tr}\left[\left(\sum_{j=1}^{k} \widetilde{G}_{\mathrm{AB}^k}(j)^\Gamma\right)^{2p}\right]\right)^{1/2p}$. This last step will be omitted here since the argument is very similar to the one appearing in the proof of Proposition

9.2.2. Concerning the moment estimate (9.12), it is first of all proved in Appendix 9.14 that

$$\mathbf{E}\ \mathrm{Tr}\left[\left(\sum_{j=1}^{k}\widetilde{G}_{\mathrm{AB}^k}(j)^\Gamma\right)^{2p}\right]\underset{d\to+\infty}{\sim}\sum_{\lambda\in\mathfrak{P}^{(2)}(2p)}\ \sum_{f:[2p]\to[\lfloor k/2\rfloor]\ \mathrm{or}\ [\lceil k/2\rceil]} d^{\sharp(\gamma^{-1}\lambda)+\sharp(\gamma_f^{-1}\lambda)+k-|\,\mathrm{Im}(f)|}.$$

And by the same arguments as the in the proof of Proposition 9.2.3, we can then identify which λ and f actually contribute to the dominant order in the latter expression, yielding

$$\begin{aligned}\mathbf{E}\ \mathrm{Tr}\left[\left(\sum_{j=1}^{k}\widetilde{G}_{\mathrm{AB}^k}(j)^\Gamma\right)^{2p}\right]&\underset{d\to+\infty}{\sim}\sum_{\lambda\in NC^{(2)}(2p)}\left(\sum_{\substack{f:[2p]\to[\lfloor k/2\rfloor]\\ f\circ\lambda=f}} d^{2p+\lfloor k/2\rfloor+1}+\sum_{\substack{f:[2p]\to[\lceil k/2\rceil]\\ f\circ\lambda=f}} d^{2p+\lceil k/2\rceil+1}\right)\\ &\underset{d\to+\infty}{\sim}\mathrm{Cat}_p\left(\lfloor k/2\rfloor^p d^{2p+\lfloor k/2\rfloor+1}+\lceil k/2\rceil^p d^{2p+\lceil k/2\rceil+1}\right),\end{aligned}$$

which is the announced moment estimate (9.12). □

Comparing Theorem 9.4.2 to Theorem 9.2.5, we see that the asymptotic mean width of the set of k-PPT-extendible states is at least $\sqrt{2}$ smaller than the asymptotic mean width of the set of k-extendible states. For instance, the set of 2-PPT-extendible states is, on average, asymptotically smaller than the set of 4-extendible states. This however does not really shed light on why adding the constraint, at each step in the sequence of tests, that the symmetric extension is PPT across one fixed (even) cut would make the entanglement detection notably faster.

9.5 Preliminaries on random-induced states and witnesses

We will employ the notation $\rho\sim\mu_{n,s}$ to mean that $\rho=\mathrm{Tr}_{\mathbf{C}^s}|\psi\rangle\langle\psi|$ with $|\psi\rangle$ a random Haar-distributed pure state on $\mathbf{C}^n\otimes\mathbf{C}^s$ (i.e. ρ describes an n-dimensional system which is obtained by partial-tracing over an s-dimensional ancilla space a uniformly distributed pure state on the global "system+ancilla" space). An equivalent mathematical characterization of such random state model is $\rho=W/\mathrm{Tr}W$ with $W\sim\mathscr{W}_{n,s}$ an (n,s)-Wishart matrix, i.e. $W=GG^\dagger$ with G a $n\times s$ matrix having independent complex normal entries (see e.g. [166]).

Let $K\subset\mathscr{D}(\mathbf{C}^n)$ be a convex body. For any $\rho\in\mathscr{D}(\mathbf{C}^n)$, a standard way of showing that $\rho\notin K$ is to produce a "not belonging to K witness", i.e. some $M\in\mathscr{H}_+(\mathbf{C}^n)$ which is such that

$$\sup_{\sigma\in K}\mathrm{Tr}(M\sigma)<\mathrm{Tr}(M\rho).$$

By testing ρ itself as possible such "not belonging to K witness", we have

$$\sup_{\sigma\in K}\mathrm{Tr}(\rho\sigma)<\mathrm{Tr}(\rho^2)\ \Rightarrow\ \rho\notin K. \tag{9.13}$$

Crucially for the applications we have in mind, the functions $\rho\mapsto\mathrm{Tr}(\rho^2)$ and $\rho\mapsto\sup_{\sigma\in K}\mathrm{Tr}(\rho\sigma)$ both have nice concentration properties around their average. More precisely, we have the two following results.

Proposition 9.5.1. *Let $n,s\in\mathbf{N}$. Then, there exist universal constants $c,c'>0$ such that, for any $\eta>0$, first of all*

$$\mathbf{P}_{\rho\sim\mu_{n,s}}\left(\left|\mathrm{Tr}(\rho^2)-\mathbf{E}_{\tau\sim\mu_{n,s}}\left[\mathrm{Tr}(\tau^2)\right]\right|\geqslant\eta\right)\leqslant e^{-cs}+e^{-c'n^3s\eta^2},$$

and second of all, for any convex body $K\subset\mathscr{D}(\mathbf{C}^n)$,

$$\mathbf{P}_{\rho\sim\mu_{n,s}}\left(\left|\sup_{\sigma\in K}\mathrm{Tr}(\rho\sigma)-\mathbf{E}_{\tau\sim\mu_{n,s}}\left[\sup_{\sigma\in K}\mathrm{Tr}(\tau\sigma)\right]\right|\geqslant\eta\right)\leqslant e^{-cs}+e^{-c'n^2s\eta^2}.$$

Proof. To show Proposition 9.5.1, we will make essential use of a local version of Levy's lemma. The usual Levy lemma is recalled as Lemma 4.2.1 in Chapter 4, Section 4.2. Here we will rely on the following refinement (see [18], Lemma 3.4, for a proof): Let $\Omega\subset S^{m-1}$ be a subset of the Euclidean unit sphere of $\mathbf{R}^m$ satisfying

$\mathbf{P}(\Omega) \geqslant 7/8$. Let also $f : S^{m-1} \to \mathbf{R}$ be a function whose restriction to Ω is L-Lipschitz and M be a central value for f (i.e. $\mathbf{P}(\{f \geqslant M\}) \geqslant 1/4$ and $\mathbf{P}(\{f \leqslant M\}) \geqslant 1/4$). Then, for any $\eta > 0$,

$$\mathbf{P}\left(\{|f - M| \geqslant \eta\}\right) \leqslant \mathbf{P}\left(S^{m-1} \setminus \Omega\right) + e^{-c_0 m\eta^2/L^2},$$

where $c_0 > 0$ is a universal constant.

It is well-known (see e.g. [166] for a proof) that $\rho \sim \mu_{n,s}$ is equivalent to $\rho = XX^\dagger$ with X uniformly distributed over the Hilbert–Schmidt unit sphere of $n \times s$ complex matrices, and the latter can be identified with the real Euclidean unit sphere S^{2ns-1}. Therefore, one may apply Levy's lemma above with $\Omega = \left\{X \in S^{2ns-1} \,:\, \|X\|_\infty \leqslant 3/\sqrt{n}\right\}$, which is such that $\mathbf{P}(S^{2ns-1} \setminus \Omega) \leqslant e^{-cs}$ for some universal constant $c > 0$ (see e.g. [17], Lemma 6 and Appendix B, for a proof).

Consider first $f : X \in S^{2ns-1} \mapsto \mathrm{Tr}((XX^\dagger)^2)$, which is $36/n$-Lipschitz on Ω. Indeed, for any $X, Y \in \Omega$,

$$\begin{aligned}
|f(X) - f(Y)| &\leqslant \left\| \left(XX^\dagger\right)^2 - \left(YY^\dagger\right)^2 \right\|_1 \\
&\leqslant \left(\|XX^\dagger\|_\infty + \|YY^\dagger\|_\infty\right) \left(\|X\|_2 + \|Y\|_2\right) \|X - Y\|_2 \\
&\leqslant \frac{36}{n} \|X - Y\|_2.
\end{aligned}$$

The second inequality is just by Hölder's inequality (more specifically $\|ABC\|_1 \leqslant \|\mathrm{A}\|_\infty \|B\|_2 \|C\|_2$) and the triangle inequality, after noticing that $\left(XX^\dagger\right)^2 - \left(YY^\dagger\right)^2 = XX^\dagger\Delta + \Delta YY^\dagger$ with $\Delta = X(X^\dagger - Y^\dagger) + (X - Y)Y^\dagger$. And the third inequality is because, by assumption, for any $Z \in \Omega$, $\|Z\|_2 = 1$ and $\|ZZ^\dagger\|_\infty = \|Z\|_\infty^2 \leqslant 9/n$.

Now, the fact that $\mathbf{P}(S^{2ns-1} \setminus \Omega) \leqslant e^{-cs}$, combined with the fact that $|f|$ is bounded by 1 on S^{2ns-1}, implies that the average of $|f|$ on $S^{2ns-1} \setminus \Omega$ is bounded by e^{-cs}, which tends to 0 when s tends to infinity. While the Lipschitz estimate for f on Ω implies that the average of f on Ω differs from its median by at most $C/n^{3/2}s^{1/2}$, which also tends to 0 when n, s tend to infinity. We can therefore conclude that the average of f is a central value of f for n, s big enough. Hence, taking $M = \mathbf{E}\, f$ as central value for f, we get the concentration estimate

$$\mathbf{P}_X\left(\left|\mathrm{Tr}\left(\left(XX^\dagger\right)^2\right) - \mathbf{E}_Y \mathrm{Tr}\left(\left(YY^\dagger\right)^2\right)\right| \geqslant \eta\right) \leqslant e^{-cs} + e^{-c' n^3 s \eta^2}.$$

Take next $f : X \in S^{2ns-1} \mapsto \sup_{\sigma \in K} \mathrm{Tr}(XX^\dagger\sigma)$, which is $6/\sqrt{n}$-Lipschitz on Ω. Indeed, for any $X, Y \in \Omega$,

$$\begin{aligned}
|f(X) - f(Y)| &\leqslant \left|\sup_{\sigma \in K} \mathrm{Tr}\left(\left(XX^\dagger - YY^\dagger\right)\sigma\right)\right| \\
&\leqslant \left\|XX^\dagger - YY^\dagger\right\|_\infty \\
&\leqslant \left(\|X\|_\infty + \|Y\|_\infty\right) \|X - Y\|_\infty \\
&\leqslant \frac{6}{\sqrt{n}} \|X - Y\|_2.
\end{aligned}$$

The second inequality is just by duality, since K is contained in the unit ball for the 1-norm. The third inequality is by the triangle inequality, after noticing that $XX^\dagger - YY^\dagger = (X - Y)X^\dagger + Y(X^\dagger - Y^\dagger)$. And the fourth inequality is by the norm inequality $\|\cdot\|_\infty \leqslant \|\cdot\|_2$ and because, by assumption, for any $Z \in \Omega$, $\|Z\|_\infty \leqslant 3/\sqrt{n}$.

Arguing as before, we see that the average of f is a central value of f for n, s big enough (this time, the average of $|f|$ on $S^{2ns-1} \setminus \Omega$ is bounded by e^{-cs} while the average of f on Ω differs from its median by at most $C/ns^{1/2}$). Hence, taking $M = \mathbf{E}\, f$ as central value for f, we get the concentration estimate

$$\mathbf{P}_X\left(\left\{\left|\sup_{\sigma \in K} \mathrm{Tr}(XX^\dagger\sigma) - \mathbf{E}_Y \sup_{\sigma \in K} \mathrm{Tr}(YY^\dagger\sigma)\right| \geqslant \eta\right\}\right) \leqslant e^{-cs} + e^{-c' n^2 s \eta^2}.$$

Hence, we indeed have the two announced deviation probability bounds. □

Combining the two statements in Proposition 9.5.1, together with equation (9.13), we get as a consequence: Let $K \subset \mathscr{D}(\mathbf{C}^n)$ be a convex body. Then, for any $\eta > 0$,

$$\mathbf{E}_{\rho \sim \mu_{n,s}}\left[\mathrm{Tr}(\rho^2)\right] - \mathbf{E}_{\rho \sim \mu_{n,s}}\left[\sup_{\sigma \in K} \mathrm{Tr}(\rho\sigma)\right] > \eta \;\Rightarrow\; \mathbf{P}_{\rho \sim \mu_{n,s}}\left(\rho \notin K\right) \geqslant 1 - e^{-cs \min(1, n^2\eta^2)}, \tag{9.14}$$

where $c > 0$ is a universal constant.

From now on, we will in fact consider random-induced states on the bipartite space $\mathbf{C}^d \otimes \mathbf{C}^d$. So let $K \subset \mathcal{D}(\mathbf{C}^d \otimes \mathbf{C}^d)$ (such as e.g. $\mathcal{P}(\mathbf{C}^d{:}\mathbf{C}^d)$ or $\mathcal{E}_k(\mathbf{C}^d{:}\mathbf{C}^d)$, $k \in \mathbf{N}$). It follows from equation (9.14) that, for any $\eta > 0$,

$$\mathbf{E}_{\rho\sim\mu_{d^2,s}}\big[\mathrm{Tr}(\rho^2)\big] - \mathbf{E}_{\rho\sim\mu_{d^2,s}}\left[\sup_{\sigma\in K}\mathrm{Tr}(\rho\sigma)\right] > \eta \;\Rightarrow\; \mathbf{P}_{\rho\sim\mu_{d^2,s}}\big(\rho \notin K\big) \geqslant 1 - e^{-cs\min(1,d^4\eta^2)}, \tag{9.15}$$

where $c > 0$ is a universal constant.

9.6 Non k-extendibility of random-induced states for "small" k

9.6.1 Strategy

Our goal in the sequel will be to identify a range of environment size s for which random-induced states on $\mathbf{C}^d \otimes \mathbf{C}^d$ are, with high-probability, not k-extendible. In view of equation (9.15), this may be done by characterizing

$$\left\{ s \in \mathbf{N} \;:\; \mathbf{E}_{\rho\sim\mu_{d^2,s}}\left[\sup_{\sigma\in\mathcal{E}_k(\mathbf{C}^d:\mathbf{C}^d)}\mathrm{Tr}(\rho\sigma)\right] < \mathbf{E}_{\rho\sim\mu_{d^2,s}}\big[\mathrm{Tr}(\rho^2)\big] \right\}.$$

Yet by Lemma 9.1.3, and using the notation introduced there, we have that for any state ρ_{AB} on $\mathrm{A}\otimes\mathrm{B}$,

$$\sup_{\sigma_{\mathrm{AB}}\in\mathcal{E}_k(\mathrm{A:B})}\mathrm{Tr}\big(\rho_{\mathrm{AB}}\sigma_{\mathrm{AB}}\big) = \left\|\sum_{j=1}^{k}\widetilde{\rho}_{\mathrm{AB}^k}(j)\right\|_\infty .$$

9.6.2 An operator-norm estimate

As explained above, to know when random-induced states on $\mathrm{A}\otimes\mathrm{B}$ are not k-extendible, what we need first is to compute the average operator-norm $\mathbf{E}\left\|\sum_{j=1}^{k}\widetilde{W}_{\mathrm{AB}^k}(j)\right\|_\infty$, for W a (d^2,s)-Wishart matrix. We will proceed in a very similar way to what was done in Section 9.2, and establish what can be seen as the analogues of Propositions 9.2.2 and 9.2.3 but for Wishart instead of GUE matrices.

Proposition 9.6.1. *Fix $k \in \mathbf{N}$ and $c > 0$. Then,*

$$\mathbf{E}_{W_{\mathrm{AB}}\sim\mathcal{W}_{d^2,cd^2}}\left\|\sum_{j=1}^{k}\widetilde{W}_{\mathrm{AB}^k}(j)\right\|_\infty \underset{d\to+\infty}{\sim} (\sqrt{ck}+1)^2 d^2.$$

As a preliminary step towards estimating the sup-norm $\mathbf{E}\left\|\sum_{j=1}^{k}\widetilde{W}_{\mathrm{AB}^k}(j)\right\|_\infty$, we will look at the p-order moments $\mathbf{E}\,\mathrm{Tr}\left[\left(\sum_{j=1}^{k}\widetilde{W}_{\mathrm{AB}^k}(j)\right)^p\right]$, $p \in \mathbf{N}$, and show that they can be expressed in terms of the p-order moments of a Marčenko-Pastur distribution of appropriate parameter.

So let us recall first a few required definitions. For any $\lambda > 0$, we shall denote by $\mu_{MP(\lambda)}$ the Marčenko-Pastur distribution of parameter λ, whose density is given by

$$\mathrm{d}\mu_{MP(\lambda)}(x) = \begin{cases} f_\lambda(x)\mathrm{d}x \text{ if } \lambda > 1 \\ (1-\lambda)\,\delta_0 + \lambda f_\lambda(x)\mathrm{d}x \text{ if } \lambda \leqslant 1 \end{cases},$$

where, setting $\lambda_\pm = (\sqrt{\lambda}\pm 1)^2$, we defined the function f_λ by

$$f_\lambda(x) = \frac{\sqrt{(\lambda_+ - x)(x-\lambda_-)}}{2\pi\lambda x}\mathbf{1}_{[\lambda_-,\lambda_+]}(x).$$

We shall also denote, for each $p \in \mathbf{N}$, by $\mathrm{M}^{(p)}_{MP(\lambda)}$ its p-order moment, i.e. $\mathrm{M}^{(p)}_{MP(\lambda)} = \int_{-\infty}^{+\infty} x^p \mathrm{d}\mu_{MP(\lambda)}(x)$. It is well-known that

$$\forall\; p \in \mathbf{N},\; \mathrm{M}^{(p)}_{MP(\lambda)} = \sum_{m=1}^{p}\lambda^m \mathrm{Nar}^m_p,$$

where Nar^m_p is the $(p,m)^{th}$ Narayana number defined in Lemma 9.10.4. In particular, $\mathrm{M}^{(p)}_{MP(1)} = \mathrm{Cat}_p$, the pth Catalan number defined in Lemma 9.10.4 as well.

Proposition 9.6.2. *Fix $k \in \mathbf{N}$ and $c > 0$. Then, when $d \to +\infty$, the random matrix $\left(\sum_{j=1}^{k} \widetilde{W}_{\mathrm{AB}^k}(j)\right)/d^2$ converges in moments towards a Marčenko-Pastur distribution of parameter ck. Equivalently, this means that, for any $p \in \mathbf{N}$,*

$$\mathbf{E}_{W_{\mathrm{AB}} \sim \mathscr{W}_{d^2, cd^2}} \operatorname{Tr}\left[\left(\sum_{j=1}^{k} \widetilde{W}_{\mathrm{AB}^k}(j)\right)^p\right] \underset{d \to +\infty}{\sim} \mathrm{M}_{MP(ck)}^{(p)} d^{2p+k+1}.$$

Remark 9.6.3. *Stronger convergence results than the one established in Proposition 9.6.2 may in fact be proved, as discussed in Appendix 9.17.*

Proof of Proposition 9.6.2. Let $p \in \mathbf{N}$. Computing the value of the p-order moment $\mathbf{E} \operatorname{Tr}\left[\left(\sum_{j=1}^{k} \widetilde{W}_{\mathrm{AB}^k}(j)\right)^p\right]$ may be done using the Gaussian Wick formula (see Lemma 9.11.1 for the statement and Appendix 9.11.2 for a succinct summary of how to derive moments of Wishart matrices from it). In our case, we get by the computations carried out in Appendix 9.15 and summarized in Proposition 9.15.1 that, for any $d, s \in \mathbf{N}$,

$$\begin{aligned}\mathbf{E}_{W_{\mathrm{AB}} \sim \mathscr{W}_{d^2, s}} \operatorname{Tr}\left[\left(\sum_{j=1}^{k} \widetilde{W}_{\mathrm{AB}^k}(j)\right)^p\right] &= \sum_{f:[p]\to[k]} \mathbf{E}_{W_{\mathrm{AB}} \sim \mathscr{W}_{d^2, s}} \operatorname{Tr}\left[\overrightarrow{\prod_{i=1}^{p}} \widetilde{W}_{\mathrm{AB}^k}(f(i))\right] \\ &= \sum_{f:[p]\to[k]} \sum_{\alpha \in \mathfrak{S}(p)} d^{\sharp(\gamma^{-1}\alpha)+\sharp(\gamma_f^{-1}\alpha)+k-|\operatorname{Im}(f)|} s^{\sharp(\alpha)},\end{aligned}$$

where we defined on $\{1, \ldots, p\}$, $\gamma = (p \ldots 1)$ as the canonical full cycle, and for each $f : [p] \to [k]$, $\gamma_f = \gamma_{f=1} \cdots \gamma_{f=k}$ as the product of the canonical full cycles on each of the level sets of f.

Hence, in the case where $s = cd^2$, for some constant $c > 0$, we have

$$\mathbf{E}_{W_{\mathrm{AB}} \sim \mathscr{W}_{d^2, cd^2}} \operatorname{Tr}\left[\left(\sum_{j=1}^{k} \widetilde{W}_{\mathrm{AB}^k}(j)\right)^p\right] = \sum_{f:[p]\to[k]} \sum_{\alpha \in \mathfrak{S}(p)} c^{\sharp(\alpha)} d^{2\sharp(\alpha)+\sharp(\gamma^{-1}\alpha)+\sharp(\gamma_f^{-1}\alpha)+k-|\operatorname{Im}(f)|}. \tag{9.16}$$

We now have to understand which $\alpha \in \mathfrak{S}(p)$ and $f : [p] \to [k]$ contribute to the dominating term in the moment expansion (9.16), i.e. are such that the quantity $2\sharp(\alpha) + \sharp(\gamma^{-1}\alpha) + \sharp(\gamma_f^{-1}\alpha) + k - |\operatorname{Im}(f)|$ is maximal.

First of all, for any $\alpha \in \mathfrak{S}(p)$, we have

$$\sharp(\alpha) + \sharp(\gamma^{-1}\alpha) = 2p - (|\alpha| + |\gamma^{-1}\alpha|) \leqslant 2p - |\gamma| = p + \sharp(\gamma) = p + 1, \tag{9.17}$$

where the first equality is by Lemma 9.10.1, whereas the second inequality is by equation (9.27) in Lemma 9.10.5 and is an equality if and only if $\alpha \in NC(p)$. Next, for any $\alpha \in \mathfrak{S}(p)$ and $f : [p] \to [k]$, we have

$$\sharp(\alpha) + \sharp(\gamma_f^{-1}\alpha) = 2p - (|\alpha| + |\gamma_f^{-1}\alpha|) \leqslant 2p - |\gamma_f| = p + \sharp(\gamma_f) = p + |\operatorname{Im}(f)|, \tag{9.18}$$

where the first equality is once more by Lemma 9.10.1, whereas the second inequality is by equation (9.28) in Lemma 9.10.5 and is an equality if and only if $\alpha \in NC(p)$ and $f \circ \alpha = f$. So equations (9.17) and (9.18) together yield that, for any $\alpha \in \mathfrak{S}(p)$ and $f : [p] \to [k]$,

$$2\sharp(\alpha) + \sharp(\gamma^{-1}\alpha) + \sharp(\gamma_f^{-1}\alpha) + k - |\operatorname{Im}(f)| \leqslant 2p + k + 1, \tag{9.19}$$

with equality if and only if $\alpha \in NC(p)$ and $f \circ \alpha = f$.

We thus get the asymptotic estimate

$$\mathbf{E} \operatorname{Tr}\left[\left(\sum_{j=1}^{k} \widetilde{W}_{\mathrm{AB}^k}(j)\right)^p\right] \underset{d \to +\infty}{\sim} \left(\sum_{\alpha \in NC(p)} \sum_{\substack{f:[p]\to[k] \\ f \circ \alpha = f}} c^{\sharp(\alpha)}\right) d^{2p+k+1}.$$

Yet, a function f satisfying $f \circ \alpha = f$ is fully characterized by its value on each of the $\sharp(\alpha)$ cycles of α. So there are $k^{\sharp(\alpha)}$ such functions. Hence in the end, the asymptotic estimate

$$\mathbf{E} \operatorname{Tr}\left[\left(\sum_{j=1}^{k} \widetilde{W}_{\mathrm{AB}^k}(j)\right)^p\right] \underset{d \to +\infty}{\sim} \left(\sum_{\alpha \in NC(p)} (ck)^{\sharp(\alpha)}\right) d^{2p+k+1} = \mathrm{M}_{MP(ck)}^{(p)} d^{2p+k+1},$$

the last equality being because, for any $\lambda > 0$, $\sum_{\alpha \in NC(p)} \lambda^{\sharp(\alpha)} = \sum_{m=1}^{p} \lambda^m \mathrm{Nar}_p^m = \mathrm{M}_{MP(\lambda)}^{(p)}$. □

Proof of Proposition 9.6.1. The argument will follow the exact same lines as the one used to derive Proposition 9.2.2 from Proposition 9.2.3.

As pointed out there, showing the inequality "$\geqslant$" in Proposition 9.6.1 is easy. Indeed, the convergence in moments established in Proposition 9.6.2 implies that, asymptotically, the matrix $\left(\sum_{j=1}^{k} \widetilde{W}_{\mathrm{AB}^k}(j)\right)/d^2$ has a largest eigenvalue which is, on average, at least the upper-edge of the support of $\mu_{MP(ck)}$, i.e. $(\sqrt{ck}+1)^2$. In other words, it guarantees that there exist positive constants $c_d \to_{d\to+\infty} 1$ such that

$$\mathbf{E}\left\|\sum_{j=1}^{k} \widetilde{W}_{\mathrm{AB}^k}(j)\right\|_\infty \geqslant c_d\,(\sqrt{ck}+1)^2 d^2. \tag{9.20}$$

Let us now turn to the more tricky part, which is showing the inequality "$\leqslant$" in Proposition 9.6.1. For $d \in \mathbf{N}$ fixed, it holds that

$$\forall\ p \in \mathbf{N},\ \mathbf{E}\left\|\sum_{j=1}^{k} \widetilde{W}_{\mathrm{AB}^k}(j)\right\|_\infty \leqslant \left(\mathbf{E}\,\mathrm{Tr}\left[\left(\sum_{j=1}^{k} \widetilde{W}_{\mathrm{AB}^k}(j)\right)^p\right]\right)^{1/p}. \tag{9.21}$$

So let us fix $d \in \mathbf{N}$ and $p \in \mathbf{N}$, and rewrite (9.16) explicitly as an expansion in powers of d, keeping in the sum the permutations not saturating equation (9.19). Being cautious only regarding the permutations not saturating equation (9.18), and not regarding those not saturating equation (9.17), we thus get the upper bound

$$\mathbf{E}\,\mathrm{Tr}\left[\left(\sum_{j=1}^{k} \widetilde{W}_{\mathrm{AB}^k}(j)\right)^p\right] \leqslant \left(\sum_{f:[p]\to[k]} \sum_{\delta=0}^{\lfloor (p+k)/2\rfloor} \sum_{m=1}^{p} |\mathfrak{S}_{f,\delta,m}(p)|\, c^m d^{-2\delta}\right) d^{2p+k+1}, \tag{9.22}$$

where we defined, for each $f : [p] \to [k]$, each $1 \leqslant m \leqslant p$ and each $0 \leqslant \delta \leqslant \lfloor (p+k)/2 \rfloor$,

$$\mathfrak{S}_{f,\delta,m}(p) = \left\{\alpha \in \mathfrak{S}(p)\ :\ \sharp(\alpha) = m \text{ and } \sharp(\alpha) + \sharp(\gamma_f^{-1}\alpha) = p + |\,\mathrm{Im}(f)| - 2\delta\right\}.$$

$\mathfrak{S}_{f,\delta,m}(p)$ is thus nothing else than the set of permutations which are composed of m cycles and have a defect 2δ from lying on the geodesics between the identity and the product of the canonical full cycles on each of the level sets of f. This justifies in particular *a posteriori* why the summation in (9.22) is only over even defects (see the parity argument in Lemma 9.10.2). Note that the definition of $\mathfrak{S}_{f,\delta,m}(p)$ can actually be extended to all $m \in \mathbf{N}$, with $|\mathfrak{S}_{f,\delta,m}(p)| = 0$ if $m \geqslant p + |\,\mathrm{Im}(f)| - 2\delta$, which we shall do in what follows for writing convenience.

Now, by Lemma 9.16.4, we know that, if $0 \leqslant \delta \leqslant \lfloor p/2 \rfloor$, then for any $1 \leqslant m \leqslant p$,

$$\left|\left\{(f,\alpha)\ :\ \alpha \in \mathfrak{S}_{f,\delta,m}(p)\right\}\right| \leqslant \left(\sum_{\epsilon=0}^{2\delta} k^{m-\epsilon}\mathrm{Nar}_p^{m-\epsilon}\right) \times \left(2k^2p^2\right)^{2\delta}.$$

And if $\lceil p/2 \rceil \leqslant \delta \leqslant \lfloor (p+k)/2 \rfloor$, then trivially for any $1 \leqslant m \leqslant p$,

$$\left|\left\{(f,\alpha)\ :\ \alpha \in \mathfrak{S}_{f,\delta,m}(p)\right\}\right| \leqslant k^p\, p! \leqslant \left(\sum_{\epsilon=0}^{p} k^{m-\epsilon}\mathrm{Nar}_p^{m-\epsilon}\right) \times \left(2k^2p^2\right)^{p}.$$

What is more, for a given $0 \leqslant \delta \leqslant \lceil p/2 \rceil$, we have, making the change of summation index $m \mapsto m - \epsilon$,

$$\begin{aligned}
\sum_{m=1}^{p} c^m \left(\sum_{\epsilon=0}^{2\delta} k^{m-\epsilon}\mathrm{Nar}_p^{m-\epsilon}\right) &= \left(\sum_{\epsilon=0}^{2\delta} c^{-\epsilon}\right)\left(\sum_{m=1}^{p} (ck)^m \mathrm{Nar}_p^m\right) \\
&\leqslant (1+c)^{2\delta} \sum_{m=1}^{p} (ck)^m \mathrm{Nar}_p^m \\
&= (1+c)^{2\delta}\mathrm{M}_{MP(ck)}^{(p)}.
\end{aligned}$$

Putting everything together, we therefore get,

$$\mathbf{E}\,\mathrm{Tr}\left[\left(\sum_{j=1}^{k} \widetilde{W}_{\mathrm{AB}^k}(j)\right)^p\right] \leqslant \mathrm{M}_{MP(ck)}^{(p)}\left(1 + \sum_{\delta=1}^{\lfloor p/2\rfloor}\left(\frac{2(1+c)k^2p^2}{d}\right)^{2\delta} + \frac{k}{2}\left(\frac{2(1+c)k^2p^2}{d}\right)^{p}\right) d^{2p+k+1}.$$

Yet, $\max\left\{\left(2(1+c)k^2p^2/d\right)^{2\delta} \ : \ 1\leqslant\delta\leqslant\lceil p/2\rceil\right\}$ is attained for $\delta=1$, provided $p\leqslant(d/2(1+c)k^2)^{1/2}$. So if such is the case,

$$\sum_{\delta=1}^{\lfloor p/2\rfloor}\left(\frac{2(1+c)k^2p^2}{d}\right)^{2\delta}+\frac{k}{2}\left(\frac{2(1+c)k^2p^2}{d}\right)^{p}\leqslant\frac{p+k}{2}\frac{4(1+c)^2k^4p^4}{d^2}\leqslant\frac{4(1+c)^2k^4p^5}{d^2},$$

where the last inequality holds as long as $p\geqslant k$. And hence, under all the previous assumptions,

$$\mathbf{E}\,\mathrm{Tr}\left[\left(\sum_{j=1}^{k}\widetilde{W}_{\mathrm{AB}^k}(j)\right)^{p}\right]\leqslant\mathrm{M}^{(p)}_{MP(ck)}\left(1+\frac{4(1+c)^2k^4p^5}{d^2}\right)d^{2p+k+1}.$$

So set $p_d=(d/2(1+c)k^2)^{(2-\epsilon)/5}$ for some $0<\epsilon<1$ (which is indeed smaller than $(d/2(1+c)k^2)^{1/2}$ and bigger than k for d big enough, in particular bigger than $2(1+c)k^{9/2}$). And using inequality (9.21) in the special case $p=p_d$, we eventually get

$$\mathbf{E}\left\|\sum_{j=1}^{k}\widetilde{W}_{\mathrm{AB}^k}(j)\right\|_{\infty}\leqslant\left(\mathrm{M}^{(p_d)}_{MP(ck)}\left(1+\frac{4(1+c)^2p_d^4}{d^2}\right)\right)^{1/p_d}d^{2+(k+1)/p_d}\underset{d\to+\infty}{\sim}(\sqrt{ck}+1)^2d^2. \tag{9.23}$$

Combining the lower bound in equation (9.20) and the upper bound in equation (9.23) yields Proposition 9.6.1. □

9.6.3 Conclusion

Having at hand the operator-norm estimate from Proposition 9.6.1, we can now easily answer our initial question. It is the content of Theorem 9.6.4 below.

Theorem 9.6.4. *Let $k\in\mathbf{N}$, and for any $0<\epsilon<1/2$ define $c_\epsilon(k)=\frac{(k-1)^2}{4k}(1-\epsilon)$. Then, there exists a constant $C_{k,\epsilon}>0$ such that*

$$\mathbf{P}_{\rho\sim\mu_{d^2,c_\epsilon(k)d^2}}\left(\rho\notin\mathscr{C}_k(\mathbf{C}^d{:}\mathbf{C}^d)\right)\geqslant1-e^{-C_{k,\epsilon}d^2}.$$

One can take $C_{k,\epsilon}=C\epsilon^2/k$ for some universal constant $C>0$.

Proof. As a direct consequence of Proposition 9.6.1, we have

$$\mathbf{E}_{W_{\mathrm{AB}}\sim\mathscr{W}_{d^2,cd^2}}\left[\sup_{\sigma_{\mathrm{AB}}\in\mathscr{C}_k(\mathrm{A}:\mathrm{B})}\mathrm{Tr}\left(W_{\mathrm{AB}}\sigma_{\mathrm{AB}}\right)\right]\underset{d\to+\infty}{\sim}\frac{(\sqrt{ck}+1)^2}{k}d^2.$$

And since $\mathbf{E}_{W\sim\mathscr{W}_{d^2,s}}\mathrm{Tr}W\sim_{d,s\to+\infty}d^2s$ (see e.g. [56] or Appendix 9.11.2), the result we eventually come to after renormalizing by $\mathrm{Tr}W$ is

$$\mathbf{E}_{\rho\sim\mu_{d^2,cd^2}}\left[\sup_{\sigma\in\mathscr{C}_k(\mathbf{C}^d:\mathbf{C}^d)}\mathrm{Tr}(\rho\sigma)\right]\underset{d\to+\infty}{\sim}\frac{1}{d^2cd^2}\frac{(\sqrt{ck}+1)^2}{k}d^2=\frac{(\sqrt{ck}+1)^2}{ck}\frac{1}{d^2}. \tag{9.24}$$

On the other hand, $\mathbf{E}_{W\sim\mathscr{W}_{d^2,s}}\mathrm{Tr}(W^2)\sim_{d,s\to+\infty}d^2s^2+(d^2)^2s$ (see e.g. [56] or Appendix 9.11.2), so we also have

$$\mathbf{E}_{\rho\sim\mu_{d^2,cd^2}}\left[\mathrm{Tr}(\rho^2)\right]\underset{d\to+\infty}{\sim}\frac{d^2(cd^2)^2+(d^2)^2cd^2}{(d^2cd^2)^2}=\left(1+\frac{1}{c}\right)\frac{1}{d^2}.$$

Now, if $c=(1-\epsilon)(k-1)^2/4k$ for some $0<\epsilon<1/2$, then

$$\frac{(\sqrt{ck}+1)^2}{ck}-\left(1+\frac{1}{c}\right)=\frac{4\epsilon}{(k-1)(1-\epsilon)}<\frac{8\epsilon}{k-1}.$$

So by equation (9.15), we have in such case

$$\mathbf{P}_{\rho\sim\mu_{d^2,cd^2}}\left(\rho\notin\mathscr{C}_k(\mathbf{C}^d{:}\mathbf{C}^d)\right)\geqslant1-\exp\left(-Ckd^2d^4(\epsilon/kd^2)^2\right)=1-\exp\left(-Cd^2\epsilon^2/k\right),$$

for some universal constant $C>0$. □

9.7 Discussion and comparison with other separability criteria

For each $k \in \mathbf{N}$, define c_{k-ext} as the smallest constant c such that a random state ρ on $\mathbf{C}^d \otimes \mathbf{C}^d$ induced by an environment of dimension cd^2 is not k-extendible with high probability when d is large. That is,

$$c_{k-ext} = \inf \left\{ c \ : \ \mathbf{P}_{\rho \sim \mu_{d^2, cd^2}} \left(\rho \notin \mathscr{E}_k(\mathbf{C}^d : \mathbf{C}^d) \right) \underset{d \to +\infty}{\to} 1 \right\}.$$

What we established in Theorem 9.6.4 is that $c_{k-ext} \leqslant (k-1)^2/4k$.

Yet, we know from [18] that for $c > 0$ fixed, $\rho \sim \mu_{d^2, cd^2}$ is with high probability entangled when $d \to +\infty$: the threshold for $\rho \sim \mu_{d^2, s(d)}$ being with high probability either entangled or separable occurs for some $s(d) = s_0(d)$ with $d^3 \lesssim s_0(d) \lesssim d^3 \log^2 d$. So what we proved is that if $c < (k-1)^2/4k$, i.e. if $k > 2c + 2\sqrt{c(c+1)} + 1$, then this generic entanglement will be generically detected by the k-extendibility test.

Furthermore, it is well-known (see e.g. [166]) that $\rho \sim \mu_{d^2, d^2}$ is equivalent to ρ being uniformly distributed on the set of mixed states on $\mathbf{C}^d \otimes \mathbf{C}^d$ (for the Haar measure induced by the Hilbert–Schmidt distance). As just mentioned, when $d \to +\infty$, such states are typically not separable. Now, for $k \geqslant 6$, $(k-1)^2/4k > 1$, so such states are also typically not k-extendible. Hence, entanglement of uniformly distributed mixed states on $\mathbf{C}^d \otimes \mathbf{C}^d$ is typically detected by the k-extendibility test for $k \geqslant 6$.

Let us define, in a similar way to what was done for the k-extendibility criterion, c_{ppt}, resp. c_{ra}, as the smallest constant c such that a random state ρ on $\mathbf{C}^d \otimes \mathbf{C}^d$ induced by an environment of dimension cd^2 is, with probability tending to one when d tends to infinity, not satisfying the PPT, resp. realignment (see Chapter 8, Section 8.4), criterion. We know from [10] that $c_{ppt} = 4$, whereas we know from [13] that $c_{ra} = (8/3\pi)^2$. Now, for $k \geqslant 17$, $(k-1)^2/4k > 4$, and for $k \geqslant 5$, $(k-1)^2/4k > (8/3\pi)^2$. So roughly speaking, this means that the k-extendibility criterion for separability becomes "better" than the PPT one at most for $k \geqslant 17$, and "better" than the realignment one at most for $k \geqslant 5$. This is to be taken in the following sense: if $k \geqslant 17$, resp. $k \geqslant 5$, then there is a range of environment dimensions for which random-induced states have a generic entanglement which is generically detected by the k-extendibility test but not detected by the PPT, resp. realignment, test.

Note also that for the reduction criterion [107], it was established in [118] that the threshold for a random-induced state on $\mathbf{C}^d \otimes \mathbf{C}^d$ either passing or failing it with high probability occurs at an environment dimension d, hence much smaller than for all previously mentioned criteria.

9.8 The unbalanced case

For the sake of simplicity, we previously focussed on the case where $\mathrm{H} = \mathrm{A} \otimes \mathrm{B}$ is a balanced bipartite Hilbert space. One may now wonder what happens, more generally, when $\mathrm{A} \equiv \mathbf{C}^{d_\mathrm{A}}$ and $\mathrm{B} \equiv \mathbf{C}^{d_\mathrm{B}}$ with d_A and d_B being possibly different. It is easy to see that the results from Theorems 9.6.4 and 9.2.5 straightforwardly generalize to the case where d_A and d_B both tend to infinity (but possibly at different rates). The corresponding statements appear in Theorem 9.8.1 below.

Theorem 9.8.1. *Let $k \in \mathbf{N}$ and let $d_\mathrm{A}, d_\mathrm{B} \in \mathbf{N}$. The mean width of the set of k-extendible states on $\mathbf{C}^{d_\mathrm{A}} \otimes \mathbf{C}^{d_\mathrm{B}}$ (with respect to $\mathbf{C}^{d_\mathrm{B}}$) satisfies*

$$w\big(\mathscr{E}_k(\mathbf{C}^{d_\mathrm{A}} : \mathbf{C}^{d_\mathrm{B}})\big) \underset{d_\mathrm{A}, d_\mathrm{B} \to +\infty}{\sim} \frac{2}{\sqrt{k}} \frac{1}{\sqrt{d_\mathrm{A} d_\mathrm{B}}}.$$

Also, when $d_\mathrm{A}, d_\mathrm{B} \to +\infty$, a random state on $\mathbf{C}^{d_\mathrm{A}} \otimes \mathbf{C}^{d_\mathrm{B}}$ which is sampled from $\mu_{d_\mathrm{A} d_\mathrm{B}, c d_\mathrm{A} d_\mathrm{B}}$, with $c < (k-1)^2/4k$, is with high probability not k-extendible (with respect to $\mathbf{C}^{d_\mathrm{B}}$).

Oppositely, when one of the two subsystems has a fixed dimension and the other one only has an increasing dimension, the sets of k-extendible states with respect to either the smaller or the bigger subsystem exhibit different size scalings. This is made precise in Theorem 9.8.2 below.

Theorem 9.8.2. *Let $k \in \mathbf{N}$ and let $d_\mathrm{A}, d_\mathrm{B} \in \mathbf{N}$. If d_A is fixed, the mean width of the set of k-extendible states on $\mathbf{C}^{d_\mathrm{A}} \otimes \mathbf{C}^{d_\mathrm{B}}$ (with respect to $\mathbf{C}^{d_B}$) satisfies*

$$w\big(\mathscr{E}_k(\mathbf{C}^{d_\mathrm{A}} : \mathbf{C}^{d_\mathrm{B}})\big) \underset{d_\mathrm{B} \to +\infty}{\sim} \frac{2}{\sqrt{k}} \frac{1}{\sqrt{d_\mathrm{A} d_\mathrm{B}}}.$$

Whereas if d_{B} is fixed, the mean width of the set of k-extendible states on $\mathbf{C}^{d_{\mathrm{A}}} \otimes \mathbf{C}^{d_{\mathrm{B}}}$ (with respect to $\mathbf{C}^{d_{\mathrm{B}}}$) satisfies

$$w\big(\mathscr{E}_k(\mathbf{C}^{d_{\mathrm{A}}}:\mathbf{C}^{d_{\mathrm{B}}})\big) \underset{d_{\mathrm{A}}\to+\infty}{\sim} \frac{2\,C(d_{\mathrm{B}},k)}{\sqrt{k}} \frac{1}{\sqrt{d_{\mathrm{A}} d_{\mathrm{B}}}},$$

with $C(d_{\mathrm{B}},k) \geqslant \big(1+(k-1)/d_{\mathrm{B}}^2\big)^{1/4}$.

Proof. Using the same notation as in the proof of Proposition 9.2.3, we start in both cases from the exact expression for the $2p$-order moment (slightly generalizing Proposition 9.13.1)

$$\mathbf{E}_{G_{\mathrm{AB}}\sim GUE(d_{\mathrm{A}}d_{\mathrm{B}})} \operatorname{Tr}\left[\left(\sum_{j=1}^{k} \widetilde{G}_{\mathrm{AB}^k}(j)\right)^{2p}\right] = \sum_{f:[2p]\to[k]} \sum_{\lambda\in\mathfrak{P}^{(2)}(2p)} d_{\mathrm{A}}^{\sharp(\gamma^{-1}\lambda)} d_{\mathrm{B}}^{\sharp(\gamma_f^{-1}\lambda)+k-|\operatorname{Im}(f)|}. \tag{9.25}$$

First, fix d_{A}. The argument then follows the exact same lines as in the proof of Proposition 9.2.3. Indeed, the pair partitions $\lambda \in \mathfrak{P}^{(2)}(2p)$ contributing to the dominant order in d_B in the expansion (9.25) are the Cat_p non-crossing pair partitions $\lambda \in NC^{(2)}(2p)$, for which $\sharp(\gamma^{-1}\lambda) = p+1$. Moreover, for each of these λ, the functions $f:[2p]\to[k]$ contributing to the dominant order in d_B in the expansion (9.25) are the k^p functions which are such that $f\circ\lambda = f$, for which $\sharp(\gamma_f^{-1}\lambda)+k-|\operatorname{Im}(f)| = p+k$. So we eventually get

$$\mathbf{E}_{G_{\mathrm{AB}}\sim GUE(d_{\mathrm{A}}d_{\mathrm{B}})} \operatorname{Tr}\left[\left(\sum_{j=1}^{k} \widetilde{G}_{\mathrm{AB}^k}(j)\right)^{2p}\right] \underset{d_{\mathrm{B}}\to+\infty}{\sim} \mathrm{Cat}_p\, k^p d_{\mathrm{A}}^{p+1} d_{\mathrm{B}}^{p+k}.$$

Now, fix d_{B}. Again, the pair partitions $\lambda \in \mathfrak{P}^{(2)}(2p)$ contributing to the dominant order in d_{A} in the expansion (9.25) are the Cat_p non-crossing pair partitions $\lambda \in NC^{(2)}(2p)$, for which $\sharp(\gamma^{-1}\lambda) = p+1$. So consider one of these λ. Observe that, for any $0 \leqslant \delta \leqslant \lfloor (p+k)/2 \rfloor$, if $f:[2p]\to[k]$ is such that there are exactly δ pair blocks of λ on which f takes two values, then necessarily $\sharp(\gamma_f^{-1}\lambda)+k-|\operatorname{Im}(f)| \geqslant p+k-2\delta$. Indeed, the case $\delta = 0$ is already known. So let us describe precisely what happens in the case $\delta = 1$, i.e. when there is exactly 1 pair block of λ on which f takes 2 values.

• If amongst these 2 values, at least 1 of them is also taken on another pair block of λ, then there exist transpositions τ,τ' and a function g satisfying $g\circ\lambda = g$, such that $\gamma_f = \gamma_g\tau\tau'$ and $|\operatorname{Im}(f)| = |\operatorname{Im}(g)|$. Hence,

$$\sharp(\gamma_f^{-1}\lambda) = \sharp(\tau'\tau\gamma_f^{-1}\lambda) \geqslant \sharp(\gamma_g^{-1}\lambda) - 2 = p + |\operatorname{Im}(g)| - 2 = p + |\operatorname{Im}(f)| - 2.$$

• If none of these 2 values is also taken on another pair block of λ, then there exist a transposition τ and a function g satisfying $g\circ\lambda = g$, such that $\gamma_f = \gamma_g\tau$ and $|\operatorname{Im}(f)| = |\operatorname{Im}(g)| + 1$. Hence,

$$\sharp(\gamma_f^{-1}\lambda) = \sharp(\tau\gamma_f^{-1}\lambda) \geqslant \sharp(\gamma_g^{-1}\lambda) - 1 = p + |\operatorname{Im}(g)| - 1 = p + |\operatorname{Im}(f)| - 2.$$

And this generalizes in a similar way to $\delta > 1$. Yet, for a given $0 \leqslant \delta \leqslant \lfloor (p+k)/2 \rfloor$, there are $\binom{p}{\delta} k^p (k-1)^\delta$ functions which take 2 values on exactly δ pair blocks of λ (assuming of course that $p \geqslant k$). So we eventually get

$$\begin{aligned}
\mathbf{E}_{G_{\mathrm{AB}}\sim GUE(d_{\mathrm{A}}d_{\mathrm{B}})} \operatorname{Tr}\left[\left(\sum_{j=1}^{k} \widetilde{G}_{\mathrm{AB}^k}(j)\right)^{2p}\right] &\geqslant \mathrm{Cat}_p\, k^p \left(\sum_{\delta=0}^{\lfloor (p+k)/2 \rfloor} \binom{p}{\delta} (k-1)^\delta d_{\mathrm{B}}^{-2\delta}\right) d_{\mathrm{B}}^{p+k} d_{\mathrm{A}}^{p+1} \\
&\geqslant \mathrm{Cat}_p\, k^p \left(\sum_{\delta=0}^{\lfloor (p+k)/2 \rfloor} \binom{\lfloor (p+k)/2 \rfloor}{\delta} \left(\frac{k-1}{d_{\mathrm{B}}^2}\right)^\delta\right) d_{\mathrm{B}}^{p+k} d_{\mathrm{A}}^{p+1} \\
&= \mathrm{Cat}_p\, k^p \left(1+\frac{k-1}{d_{\mathrm{B}}^2}\right)^{\lfloor (p+k)/2 \rfloor} d_{\mathrm{B}}^{p+k} d_{\mathrm{A}}^{p+1}.
\end{aligned}$$

One can then argue as in the proof of the derivation of Proposition 9.2.2 from Proposition 9.2.3 that we additionally have $\mathbf{E}\left\|\sum_{j=1}^{k} \widetilde{G}_{\mathrm{AB}^k}\right\|_\infty \sim \lim_{p\to+\infty} \left(\mathbf{E}\operatorname{Tr}\left[\left(\sum_{j=1}^{k} \widetilde{G}_{\mathrm{AB}^k}\right)^{2p}\right]\right)^{1/2p}$, when either $d_{\mathrm{B}} \to +\infty$ or $d_{\mathrm{A}} \to +\infty$. This automatically yields the two announced statements on the mean width of $\mathscr{E}_k(\mathbf{C}^{d_{\mathrm{A}}}:\mathbf{C}^{d_{\mathrm{B}}})$. □

Remark 9.8.3. *In the situation where d_{B} is fixed, if we had an exact expression*

$$\forall\ p \in \mathbf{N},\ \mathbf{E}_{G_{\mathrm{AB}} \sim GUE(d_{\mathrm{A}} d_{\mathrm{B}})}\, \mathrm{Tr} \left[\left(\sum_{j=1}^{k} \widetilde{G}_{\mathrm{AB}^k}(j) \right)^{2p} \right] \underset{d_{\mathrm{A}} \to +\infty}{\sim} M(d_{\mathrm{A}}, p),$$

then we would be able to conclude without any further argument that

$$\mathbf{E} \left\| \sum_{j=1}^{k} \widetilde{G}_{\mathrm{AB}^k} \right\|_{\infty} \underset{d_{\mathrm{A}} \to +\infty}{\sim} \lim_{p \to +\infty} M(d_{\mathrm{A}}, p)^{1/2p}.$$

This would indeed follow from the convergence result of [91] for non-commutative polynomials in multi-variables with matrix coefficients (in our case, d_{B}^2 variables and $d_{\mathrm{B}}^k \times d_{\mathrm{B}}^k$ coefficients).

The asymmetry in the definition of k-extendibility appears more strikingly in this unbalanced setting. Indeed, for a finite $k \in \mathbf{N}$, a given state on $\mathrm{A} \otimes \mathrm{B}$ may be k-extendible with respect to B but not k-extendible with respect to A. It is only in the limit $k \to +\infty$ that there is equivalence between the two notions: a state on $\mathrm{A} \otimes \mathrm{B}$ is k-extendible with respect to B for all $k \in \mathbf{N}$ if and only if it is k-extendible with respect to A for all $k \in \mathbf{N}$ (and if and only if it is separable).

However, what Theorem 9.8.1 stipulates is that, even for a finite $k \in \mathbf{N}$, when both subsystems grow, being k-extendible with respect to either one or the other are two constraints which are, on average, equivalently restricting. On the contrary, what Theorem 9.8.2 shows is that when only one subsystem grows, and the other remains of fixed size, being k-extendible with respect to the bigger one is, on average, a tougher constraint than being k-extendible with respect to the smaller one (as one would have probably expected).

This is to be put in perspective with some of the original observations made in [66]. It was indeed noticed that checking whether a state on $\mathbf{C}^d \otimes \mathbf{C}^{d'}$ is k-extendible with respect to $\mathbf{C}^{d'}$ requires space resources which scale as $\left[\binom{d'+k-1}{k} d \right]^2$ when implemented. It was therefore advised that in the unbalanced situation of d "big" and d' "small", one should check k-extendibility with respect to $\mathbf{C}^{d'}$ rather than $\mathbf{C}^d$, the former being much more economical. On the other hand, it comes out from our study that, in this case, an entangled state is likely to fail passing the k-extendibility test for a smaller k when the extension is searched with respect to $\mathbf{C}^d$ than when it is searched with respect to $\mathbf{C}^{d'}$. But understanding the precise trade-off seems out of reach at the moment.

9.9 Miscellaneous questions

9.9.1 What about the mean width of the set of k-extendible states for "big" k?

All the statements proven sofar, regarding either the k-extendibility of random-induced states or the mean width of the set of k-extendible states, converge towards the same (expected) conclusion: for any given $k \in \mathbf{N}$, the k-extendibility criterion becomes a very weak necessary condition for separability when the dimension of the considered bipartite system increases. So the natural question at that point is: what can be said about the k-extendibility criterion on $\mathbf{C}^d \otimes \mathbf{C}^d$ when $k \equiv k(d)$ is allowed to grow in some way with d? Unfortunately, most of the results we established rely at some point on the assumption that k is a fixed parameter, and therefore do not seem to be directly generalizable to the case where k depends on d.

There is at least one estimate though that remains valid in this setting, which is the lower bound on the mean width of k-extendible states.

Theorem 9.9.1. *There exist positive constants $c_d \to_{d \to +\infty} 1$ such that, for any $k(d) \in \mathbf{N}$, the mean width of the set of $k(d)$-extendible states on $\mathbf{C}^d \otimes \mathbf{C}^d$ satisfies*

$$w\left(\mathscr{E}_{k(d)}(\mathbf{C}^d : \mathbf{C}^d) \right) \geqslant c_d \, \frac{2}{\sqrt{k(d)}} \frac{1}{d}.$$

Proof. Let $d, k(d) \in \mathbf{N}$. For any $p \in \mathbf{N}$, the exact expression for the $2p$-order moment established in Proposition 9.13.1 of course remains true. So by the same arguments as in the proof of Proposition 9.2.3, we still have in

that case at least the lower bound

$$\mathbf{E}_{G_{\mathrm{AB}}\sim GUE(d^2)}\mathrm{Tr}\left[\left(\sum_{j=1}^{k(d)}\widetilde{G}_{\mathrm{AB}^{k(d)}}(j)\right)^{2p}\right] = \sum_{f:[2p]\to[k(d)]}\sum_{\lambda\in\mathfrak{P}^{(2)}(2p)} d^{\sharp(\gamma^{-1}\lambda)+\sharp(\gamma_f^{-1}\lambda)+k(d)-|\,\mathrm{Im}(f)|} \geqslant \mathrm{Cat}_p\, k(d)^p d^{2p+k(d)+1}.$$

This lower bound on moments in turn guarantees, as explained in the derivation of Proposition 9.2.2 from Proposition 9.2.3, that there exist positive constants $c_d \to_{d\to+\infty} 1$ such that we have the inequality

$$\mathbf{E}_{G_{\mathrm{AB}}\sim GUE(d^2)}\left\|\sum_{j=1}^{k(d)}\widetilde{G}_{\mathrm{AB}^{k(d)}}(j)\right\|_\infty \geqslant c_d\, 2\sqrt{k(d)}d,$$

which yields the announced lower bound for the mean width of $\mathscr{E}_{k(d)}(\mathbf{C}^d{:}\mathbf{C}^d)$. □

Theorem 9.9.1 only provides a lower bound on the asymptotic mean width of $\mathscr{E}_k(\mathbf{C}^d{:}\mathbf{C}^d)$ when k is allowed to depend on d. It is nevertheless already an interesting piece of information. Indeed, as mentioned in Section 9.3, we know from [16] that the mean width of the set of separable states on $\mathbf{C}^d\otimes\mathbf{C}^d$ is of order $1/d^{3/2}$. Theorem 9.9.1 therefore asserts that, on $\mathbf{C}^d\otimes\mathbf{C}^d$, one has to go at least to k of order d to obtain a set of k-extendible states whose mean width scales as the one of the set of separable states.

Furthermore, it may be worth mentioning that the proof of Proposition 9.2.2 actually provides additional information, namely an upper bound on the mean width of k-extendible states which remains valid for a quite wide range of k.

Theorem 9.9.2. *For any $d, k(d) \in \mathbf{N}$, provided that $k(d) < d^{2/7}$ and d is big enough, the mean width of the set of $k(d)$-extendible states on $\mathbf{C}^d\otimes\mathbf{C}^d$ satisfies*

$$w\big(\mathscr{E}_{k(d)}(\mathbf{C}^d{:}\mathbf{C}^d)\big) \leqslant \frac{2}{\sqrt{k(d)}}\frac{1}{d}\exp\left(\frac{k(d)^{7/5}\ln d}{d^{2/5}}\right).$$

Proof. Let $d, k(d) \in \mathbf{N}$ with $k(d) < d^{2/7}$. Taking $p_d = (d/k(d))^{2/5}$ in equation (9.11) (which is indeed, as required, bigger than $k(d)$ and smaller than $(2d/k(d))^{1/2}$ for d big enough) we get

$$\mathbf{E}_{G_{\mathrm{AB}}\sim GUE(d^2)}\left\|\sum_{j=1}^{k(d)}\widetilde{G}_{\mathrm{AB}^{k}}(j)\right\|_\infty \leqslant 2\sqrt{k(d)}\, d\,\exp\left(\frac{k(d)^{7/5}}{2d^{2/5}}\ln d + \frac{k(d)^{2/5}}{2d^{2/5}}\left[\ln d + \ln\left(\frac{5}{4}\right)\right]\right).$$

The latter quantity is smaller than $2\sqrt{k(d)}\, d\,\exp\big(k(d)^{7/5}\ln d/d^{2/5}\big)$ for d big enough, which yields the advertised upper bound for the mean width of $\mathscr{E}_{k(d)}(\mathbf{C}^d{:}\mathbf{C}^d)$. □

Of course, the upper bound provided by Theorem 9.9.2 is interesting only for $k(d) < k_0(d) = d^{2/7}/(\ln d)^{5/7}$. Nevertheless, since the set of k-extendible states contains the set of k'-extendible states for all $k' \geqslant k$, we also have as a (potentially weak) consequence of Theorem 9.9.2 that for $k(d) \geqslant k_0(d)$ and d big enough,

$$w\big(\mathscr{E}_{k(d)}(\mathbf{C}^d{:}\mathbf{C}^d)\big) \leqslant w\big(\mathscr{E}_{k_0(d)}(\mathbf{C}^d{:}\mathbf{C}^d)\big) \leqslant \frac{2e\,(\ln d)^{5/14}}{d^{8/7}}.$$

Theorems 9.9.1 and 9.9.2 together imply in particular the following: in the regime where $k(d)$ grows with d slower than d itself, both the ratio $w\big(\mathcal{S}(\mathbf{C}^d{:}\mathbf{C}^d)\big)/w\big(\mathscr{E}_{k(d)}(\mathbf{C}^d{:}\mathbf{C}^d)\big)$ and the ratio $w\big(\mathscr{E}_{k(d)}(\mathbf{C}^d{:}\mathbf{C}^d)\big)/w\big(\mathscr{D}(\mathbf{C}^d\otimes\mathbf{C}^d)\big)$ are unbounded. To rephrase it, for $k(d)$ having this growth rate, the set of $k(d)$-extendible states lies "strictly in between" the set of separable states and the set of all states from an asymptotic size point of view.

9.9.2 When is a random-induced state with high probability k-extendible?

The result provided by Theorem 9.6.4 is only one-sided: it tells us that if $s < s(k,d) = d^2(k-1)^2/4k$, then a random mixed state on $\mathbf{C}^d\otimes\mathbf{C}^d$ obtained by partial tracing on $\mathbf{C}^s$ a uniformly distributed pure state on $\mathbf{C}^d\otimes\mathbf{C}^d\otimes\mathbf{C}^s$ is with high probability not k-extendible. But what can be said about the case $s > s(k,d)$? Or more generally, can one find a reasonable $s'(k,d)$ such that if $s > s'(k,d)$, then a random mixed state on $\mathbf{C}^d\otimes\mathbf{C}^d$ obtained by partial tracing on $\mathbf{C}^s$ a uniformly distributed pure state on $\mathbf{C}^d\otimes\mathbf{C}^d\otimes\mathbf{C}^s$ is with high probability k-extendible?

By the arguments discussed in extensive depth in [18], one can assert at least that there exists a universal constant $c > 0$ such that $ckd^2 \log^2 d$ is a possible value for such $s'(k,d)$. We will not repeat the whole reasoning here, but let us still give the key ideas underlying it.

Define $\overline{\mathscr{E}}_k(\mathbf{C}^d{:}\mathbf{C}^d)$ the translation of $\mathscr{E}_k(\mathbf{C}^d{:}\mathbf{C}^d)$ by its center of mass, the maximally mixed state Id/d^2, i.e.

$$\overline{\mathscr{E}}_k(\mathbf{C}^d{:}\mathbf{C}^d) = \left\{ \rho - \frac{\mathrm{Id}}{d^2} \ : \ \rho \in \mathscr{E}_k(\mathbf{C}^d{:}\mathbf{C}^d) \right\}.$$

Define also $\overline{\mathscr{E}}_k(\mathbf{C}^d{:}\mathbf{C}^d)^\circ$ the convex body polar to $\overline{\mathscr{E}}_k(\mathbf{C}^d{:}\mathbf{C}^d)$, i.e.

$$\overline{\mathscr{E}}_k(\mathbf{C}^d{:}\mathbf{C}^d)^\circ = \left\{ \Delta \in \mathscr{H}(d^2) \ : \ \forall \ X \in \overline{\mathscr{E}}_k(\mathbf{C}^d{:}\mathbf{C}^d), \ \mathrm{Tr}\,(\Delta X) \leqslant 1 \right\}.$$

What then has to be specifically determined is (see [18], Section 2, for further comments)

$$\left\{ s \in \mathbf{N} \ : \ \mathbf{E}_{\rho \sim \mu_{d^2,s}} \sup_{\Delta \in \overline{\mathscr{E}}_k(\mathbf{C}^d{:}\mathbf{C}^d)^\circ} \mathrm{Tr}\left(\left(\rho - \frac{\mathrm{Id}}{d^2} \right) \Delta \right) < 1 \right\},$$

One can first of all use the fact that, roughly speaking, when $d, s \to +\infty$, the random matrix $\rho - \mathrm{Id}/d^2$ for $\rho \sim \mu_{d^2,s}$ "looks the same as" the random matrix $G/d^2\sqrt{s}$ for $G \sim GUE(d^2)$ (see [18], Proposition 3.1 and Remark 3.2 as well as Appendices A and B, for precise majorization statements and proofs). In particular, there exists a constant $C > 0$ such that, for all $d, s \in \mathbf{N}$ with (say) $d^2 \leqslant s \leqslant d^3$, we have the upper bound

$$\mathbf{E}_{\rho \sim \mu_{d^2,s}} \sup_{\Delta \in \overline{\mathscr{E}}_k(\mathbf{C}^d{:}\mathbf{C}^d)^\circ} \mathrm{Tr}\left(\left(\rho - \frac{\mathrm{Id}}{d^2} \right) \Delta \right) \leqslant \frac{C}{d^2\sqrt{s}} \, \mathbf{E}_{G \sim GUE(d^2)} \sup_{\Delta \in \overline{\mathscr{E}}_k(\mathbf{C}^d{:}\mathbf{C}^d)^\circ} \mathrm{Tr}\,(G\Delta).$$

Next, due to the fact that, again putting it vaguely, the convex body $\overline{\mathscr{E}}_k(\mathbf{C}^d{:}\mathbf{C}^d)$ is "sufficiently well-balanced" (see [18], Section 4 as well as Appendices C and D, for a complete exposition of the ℓ-position argument), we know that there exists a constant $C' > 0$ such that, for all $d \in \mathbf{N}$, we have the upper bound

$$\left(\mathbf{E}_{G \sim GUE(d^2)} \sup_{\Delta \in \overline{\mathscr{E}}_k(\mathbf{C}^d{:}\mathbf{C}^d)^\circ} \mathrm{Tr}\,(G\Delta) \right) \left(\mathbf{E}_{G \sim GUE(d^2)} \sup_{\Delta \in \overline{\mathscr{E}}_k(\mathbf{C}^d{:}\mathbf{C}^d)} \mathrm{Tr}\,(G\Delta) \right) \leqslant C' d^4 \log d.$$

Now, $\mathbf{E}_{G \sim GUE(d^2)} \sup_{\Delta \in \overline{\mathscr{E}}_k(\mathbf{C}^d{:}\mathbf{C}^d)} \mathrm{Tr}\,(G\Delta)$ is nothing else than the Gaussian mean width of $\overline{\mathscr{E}}_k(\mathbf{C}^d{:}\mathbf{C}^d)$, which is the same as the Gaussian mean width of $\mathscr{E}_k(\mathbf{C}^d{:}\mathbf{C}^d)$, so for which we have an estimate thanks to Theorem 9.2.5, namely $w_G\left(\overline{\mathscr{E}}_k(\mathbf{C}^d{:}\mathbf{C}^d)\right) \sim_{d \to +\infty} 2d/\sqrt{k}$.

Putting everything together, we see that

$$\mathbf{E}_{\rho \sim \mu_{d^2,s}} \sup_{\Delta \in \overline{\mathscr{E}}_k(\mathbf{C}^d{:}\mathbf{C}^d)^\circ} \mathrm{Tr}\left(\left(\rho - \frac{\mathrm{Id}}{d^2} \right) \Delta \right) \leqslant \widetilde{C} \frac{\sqrt{k} d \log d}{\sqrt{s}},$$

for some constant $\widetilde{C} > 0$ independent of $d, s, k \in \mathbf{N}$, which implies as claimed that if $s > \widetilde{C}^2 k d^2 \log^2 d$, then $\mathbf{E}_{\rho \sim \mu_{d^2,s}} \sup_{\Delta \in \overline{\mathscr{E}}_k(\mathbf{C}^d{:}\mathbf{C}^d)^\circ} \mathrm{Tr}\left(\left(\rho - \mathrm{Id}/d^2 \right) \Delta \right) < 1$.

Remark 9.9.3. *Let us briefly comment on a notable difference, from a convex geometry point of view, between the k-extendibility criterion and other common separability criteria. In the case of k-extendibility, computing the support function of $\overline{\mathscr{E}}_k$ is easier than computing the support function of its polar $\overline{\mathscr{E}}_k^\circ$, while for other separability relaxations it is usually the opposite. Indeed, for a given traceless unit Hilbert-Schmidt norm Hermitian Δ on $\mathbf{C}^d \otimes \mathbf{C}^d$, we have for instance the closed formulas*

$$h_{\overline{\mathscr{E}}_k(\mathbf{C}^d{:}\mathbf{C}^d)}(\Delta) = \sup\left\{ \mathrm{Tr}(\Delta X) \ : \ X \in \overline{\mathscr{E}}_k(\mathbf{C}^d{:}\mathbf{C}^d) \right\} = \left\| \widetilde{\Delta} \right\|_\infty,$$

$$h_{\overline{\mathscr{P}}(\mathbf{C}^d{:}\mathbf{C}^d)^\circ}(\Delta) = \sup\left\{ \mathrm{Tr}(\Delta X) \ : \ X \in \overline{\mathscr{P}}(\mathbf{C}^d{:}\mathbf{C}^d)^\circ \right\} = d^2 \left\| \Delta^\Gamma \right\|_\infty,$$

whereas the dual quantities $h_{\overline{\mathscr{E}}_k(\mathbf{C}^d{:}\mathbf{C}^d)^\circ}(\Delta)$ and $h_{\overline{\mathscr{P}}(\mathbf{C}^d{:}\mathbf{C}^d)}(\Delta)$ cannot be written in such a simple way.

This explains why in the case of $\mathscr{E}_k$ it is the mean width that can be exactly computed, contrary to the threshold value which can only be approximated, while for other approximations of $\mathcal{S}$ the reverse generally happens.

9.10 Appendix A: Combinatorics of permutations and partitions: short summary of standard facts

Let $p \in \mathbf{N}$. We denote by $\mathfrak{S}(p)$ the set of permutations on $\{1,\dots,p\}$. For any $\pi \in \mathfrak{S}(p)$, we denote by $\sharp(\pi)$ the number of cycles in the decomposition of π into a product of disjoint cycles, and by $|\pi|$ the minimal number of transpositions in the decomposition of π into a product of transpositions. We also define $\gamma \in \mathfrak{S}(p)$ as the canonical full cycle $(p \dots 1)$. More generally, we shall say that c is the canonical full cycle on a set $\{i_1,\dots,i_p\}$ with $i_1 < \dots < i_p$ if $c = (i_p \dots i_1)$.

Some standard results related to $\mathfrak{S}(p)$ are gathered below (see e.g. [144], Lectures 9 and 23, for more details).

Lemma 9.10.1. *For any $\pi \in \mathfrak{S}(p)$, $\sharp(\pi) + |\pi| = p$.*

Lemma 9.10.2. *$d : (\pi,\varsigma) \in \mathfrak{S}(p) \times \mathfrak{S}(p) \mapsto |\pi^{-1}\varsigma|$ defines a distance on $\mathfrak{S}(p)$, so that for any $\pi,\varsigma \in \mathfrak{S}(p)$,*

$$|\varsigma^{-1}\pi| + |\pi| = d(\varsigma,\pi) + d(\mathrm{id},\pi) \geqslant d(\mathrm{id},\varsigma) = |\varsigma|, \tag{9.26}$$

with equality in (9.26) if and only if π lies on the geodesic between id *and ς. And whenever this is not the case, there exists $\delta \in \{1,\dots,p-1\}$ such that $|\varsigma^{-1}\pi| + |\pi| = |\varsigma| + 2\delta$.*

Definition 9.10.3. *A partition of $\{1,\dots,p\}$ is a family $\lambda = \{I_1,\dots,I_L\}$ of disjoint non-empty subsets of $\{1,\dots,p\}$ whose union is $\{1,\dots,p\}$. The sets $I_1,\dots,I_L$ are called the blocks of λ. If each of them contains exactly 2 elements, λ is said to be a pair partition of $\{1,\dots,p\}$. We shall denote by $\mathfrak{P}(p)$ the set of partitions of $\{1,\dots,p\}$, and by $\mathfrak{P}^{(2)}(p)$ the set of pair partitions of $\{1,\dots,p\}$. Note that $\mathfrak{P}^{(2)}(p) = \emptyset$ if p is odd. Remark also that, whenever p is even, the set of pair partitions of $\{1,\dots,p\}$ is in bijection with the set of pairings on $\{1,\dots,p\}$ (i.e. the set of permutations on $\{1,\dots,p\}$ which are a product of $p/2$ disjoint transpositions). We shall therefore make no distinction between both.*

A partition of $\{1,\dots,p\}$ is said to be non-crossing if there does not exist $i < j < k < l$ in $\{1,\dots,p\}$ such that i,k belong to the same block, j,l belong to the same block, and i,j belong to different blocks. We shall denote by $NC(p)$ the set of non-crossing partitions of $\{1,\dots,p\}$, and by $NC^{(2)}(p)$ the set of pair non-crossing partitions of $\{1,\dots,p\}$. Note that $NC^{(2)}(p) = \emptyset$ if p is odd.

A well-known combinatorial result regarding non-crossing partitions is the following.

Lemma 9.10.4. *The number of non-crossing partitions of $\{1,\dots,p\}$ and the number of pair non-crossing partitions of $\{1,\dots,2p\}$ are both equal to the pth Catalan number*

$$\mathrm{Cat}_p = \frac{1}{p+1}\binom{2p}{p}.$$

More precisely, for any $1 \leqslant m \leqslant p$, the number of non-crossing partitions of $\{1,\dots,p\}$ which are composed of exactly m blocks is equal to the $(p,m)^{th}$ Narayana number

$$\mathrm{Nar}_p^m = \frac{1}{p+1}\binom{p+1}{m}\binom{p-1}{m-1}.$$

Obviously, these numbers are such that $\sum_{m=1}^{p} \mathrm{Nar}_p^m = \mathrm{Cat}_p$.

With these definitions in mind, we can now state a special case of particular interest of Lemma 9.10.2.

Lemma 9.10.5. *Denote by γ the canonical full cycle on $\{1,\dots,p\}$. Then, for any $\pi \in \mathfrak{S}(p)$,*

$$|\gamma^{-1}\pi| + |\pi| \geqslant |\gamma| = p - 1, \tag{9.27}$$

with equality in (9.27) if and only if π lies on the geodesic between id *and γ. The latter subset of $\mathfrak{S}(p)$ is in bijection with the set of non-crossing partitions of $\{1,\dots,p\}$ (by the mapping which associates to a given partition the product of the canonical full cycles on each of its blocks). We shall thus write $\pi \in NC(p)$ in such case, not distinguishing a geodesic permutation from its corresponding non-crossing partition.*

More generally, let $\{I_1,\dots,I_L\}$ be a partition of $\{1,\dots,p\}$ and denote by $\gamma_1,\dots,\gamma_L$ the canonical full cycles on $I_1,\dots,I_L$. Then, for any $\pi \in \mathfrak{S}(p)$,

$$|(\gamma_1 \cdots \gamma_L)^{-1}\pi| + |\pi| \geqslant |\gamma_1 \cdots \gamma_L| = p - L, \tag{9.28}$$

with equality in (9.28) if and only if π lies on the geodesic between id *and $\gamma_1 \cdots \gamma_L$. The latter subset of $\mathfrak{S}(p)$ is in bijection with the set of non-crossing partitions of $\{1,\dots,p\}$ which are finer than $I_1 \sqcup \cdots \sqcup I_L$, which itself is in bijection with $NC(|I_1|) \times \cdots \times NC(|I_L|)$.*

Combining Lemma 9.10.5 with Lemma 9.10.4, we can in fact say the following: Let $\varsigma \in \mathfrak{S}(p)$ and assume that its decomposition into disjoint cycles is $\varsigma = c_1 \cdots c_L$ where, for each $1 \leqslant i \leqslant L$, c_i is of length p_i (hence with $p_1 + \cdots + p_L = p$). Then, for any $\pi \in \mathfrak{S}(p)$, $|\varsigma^{-1}\pi| + |\pi| \geqslant p - L$, and

$$\left|\left\{\pi \in \mathfrak{S}(p) \ : \ |\varsigma^{-1}\pi| + |\pi| = p - L\right\}\right| = \mathrm{Cat}_{p_1} \times \cdots \times \mathrm{Cat}_{p_L}.$$

Having this easy observation in mind might be useful later on.

9.11 Appendix B: Computing moments of random matrices: Wick formula and genus expansion

When computing expectations of Gaussian random variables, a useful tool is the Wick formula (see e.g. [178] or [144], Lecture 22, for a proof).

Lemma 9.11.1 (Gaussian Wick formula)**.** *Let $X_1, \ldots, X_q$ be jointly Gaussian centered random variables (real or complex).*

If $q = 2p+1$ is odd, then $\mathbf{E}\left[X_1 \cdots X_q\right] = 0$.

If $q = 2p$ is even, then $\mathbf{E}\left[X_1 \cdots X_q\right] = \sum_{\{\{i_1,j_1\},\ldots,\{i_p,j_p\}\} \in \mathfrak{P}^{(2)}(2p)} \prod_{m=1}^{p} \mathbf{E}\left[X_{i_m} X_{j_m}\right]$.

9.11.1 Moments of GUE matrices

A first important application of Lemma 9.11.1 is to the computation of the moments of matrices from the Gaussian Unitary Ensemble. Indeed, for any $q \in \mathbf{N}$, we have

$$\mathbf{E}_{G \sim GUE(n)} \mathrm{Tr}\left(G^q\right) = \sum_{1 \leqslant l_1, \ldots, l_q \leqslant n} \mathbf{E}\left[G_{l_1,l_2} \cdots G_{l_q,l_1}\right],$$

where the $G_{i,j}$, $1 \leqslant i,j \leqslant n$, are centered Gaussian random variables satisfying $\mathbf{E}[G_{i,j}G_{i',j'}] = \delta_{i=j',j=i'}$. So what we get applying the Wick formula is that, for any $p \in \mathbf{N}$,

$$\mathbf{E}_{G \sim GUE(n)} \mathrm{Tr}\left(G^{2p+1}\right) = 0$$

$$\mathbf{E}_{G \sim GUE(n)} \mathrm{Tr}\left(G^{2p}\right) = \sum_{\lambda \in \mathfrak{P}^{(2)}(2p)} n^{\flat(\lambda)},$$

where for each pair partition $\lambda = \{\{i_1,j_1\}, \ldots, \{i_p,j_p\}\}$ of $\{1,\ldots,2p\}$, $\flat(\lambda)$ is the number of free parameters $l_1, \ldots, l_{2p} \in \{1,\ldots,n\}$ when imposing that $\forall\ 1 \leqslant m \leqslant p$, $l_{i_m+1} = l_{j_m}$, $l_{j_m+1} = l_{i_m}$. Identifying the pair partition $\{\{i_1,j_1\}, \ldots, \{i_p,j_p\}\}$ with the pairing $(i_1\, j_1) \ldots (i_p\, j_p)$ and denoting by γ the canonical full cycle $(2p \ \ldots \ 1)$, the latter condition can be written as $\forall\ 1 \leqslant i \leqslant 2p$, $l_{\gamma^{-1}\lambda(i)} = l_i$. So in fact, $\flat(\lambda) = \sharp(\gamma^{-1}\lambda)$ and the expression above becomes

$$\mathbf{E}_{G \sim GUE(n)} \mathrm{Tr}\left(G^{2p}\right) = \sum_{\lambda \in \mathfrak{P}^{(2)}(2p)} n^{\sharp(\gamma^{-1}\lambda)}.$$

We thus have the so-called *genus expansion* (see e.g. [144], Lecture 22)

$$\mathbf{E}_{G \sim GUE(n)} \mathrm{Tr}\left(G^{2p}\right) = \sum_{\delta=0}^{\lfloor p/2 \rfloor} P(2p,\delta) n^{p+1-2\delta},$$

where for each $0 \leqslant \delta \leqslant \lfloor p/2 \rfloor$, we defined $P(2p,\delta)$ as the number of pairings of $\{1,\ldots,2p\}$ having *genus* δ, i.e.

$$P(2p,\delta) = \left|\left\{\lambda \in \mathfrak{P}^{(2)}(2p) \ : \ \sharp(\gamma^{-1}\lambda) = p+1-2\delta\right\}\right| = \left|\left\{\lambda \in \mathfrak{P}^{(2)}(2p) \ : \ \sharp(\gamma^{-1}\lambda) + \sharp(\lambda) = 2p+1-2\delta\right\}\right|.$$

Equivalently, $P(2p,\delta)$ is the number of pairings of $\{1,\ldots,2p\}$ having a defect 2δ of being on the geodesics between id and γ. Hence, $P(2p,0)$ is the number of pairings of $\{1,\ldots,2p\}$ lying exactly on the geodesics between id and γ, i.e. the number of non-crossing pair partitions of $\{1,\ldots,2p\}$. So $P(2p,0) = \left|NC^{(2)}(2p)\right| = \mathrm{Cat}_p$, and we recover the well-known asymptotic estimate

$$\mathbf{E}_{G \sim GUE(n)} \mathrm{Tr}\left(G^{2p}\right) \underset{n \to +\infty}{\sim} \mathrm{Cat}_p\, n^{p+1}.$$

9.11.2 Moments of Wishart matrices

A second important application of Lemma 9.11.1 is to the computation of the moments of matrices from the Wishart Ensemble. In such case, a graphical way of visualising the Wick formula has been developed in [56], to which the reader is referred for further details and proofs, a brief summary only being provided here.

In the graphical formalism, a matrix $X : \mathbf{C}^m \to \mathbf{C}^n$ is represented by a "box" with two "gates", one specifying the size m at its entrance and the other specifying the size n at its exit. For $X : \mathbf{C}^m \to \mathbf{C}^n$ and $Y : \mathbf{C}^n \to \mathbf{C}^m$, the product $XY : \mathbf{C}^n \to \mathbf{C}^n$ is represented by a wire connecting the exit of Y to the entrance of X. For $Z : \mathbf{C}^m \to \mathbf{C}^m$, the trace $\mathrm{Tr}(Z)$ is represented by a wire connecting the exit and the entrance of Z.

Let W be a (n, s)-Wishart matrix, i.e. $W = GG^\dagger$ with G a $n \times s$ matrix with independent complex normal entries. Representing by ♦ a n-dimensional gate and by ▼ a s-dimensional gate, the quantity $\mathrm{Tr}(W^p)$ is then graphically represented by p boxes G and p boxes $G^\dagger$ connected by wires in the following way.



For any $\alpha \in \mathfrak{S}(p)$, we will denote by $\mathscr{G}_\alpha$ the diagram obtained from the one above by "erasing" the boxes, just keeping their gates, and then connecting, for each $1 \leqslant i \leqslant p$, the entrance of the ith box G to the exit of the $\alpha(i)^{th}$ box $G^\dagger$, and the exit of the ith box G to the entrance of the $\alpha(i)^{th}$ box $G^\dagger$. Doing so, $\sharp(\gamma^{-1}\alpha)$ loops connecting n-dimensional gates and $\sharp(\alpha)$ loops connecting s-dimensional gates are obtained. And the graphical version of the Wick formula tells us that

$$\mathbf{E}_{W\sim\mathscr{W}_{n,s}} \mathrm{Tr}(W^p) = \sum_{\alpha\in\mathfrak{S}(p)} \mathscr{D}_\alpha = \sum_{\alpha\in\mathfrak{S}(p)} n^{\sharp(\gamma^{-1}\alpha)} s^{\sharp(\alpha)}.$$

In the special case where $s = n$, this can be rewritten as a so-called *genus expansion* (see e.g. [56])

$$\mathbf{E}_{W\sim\mathscr{W}_{n,n}} \mathrm{Tr}(W^p) = \sum_{\delta=0}^{\lfloor p/2\rfloor} S(p,\delta) n^{p+1-2\delta},$$

where for each $0 \leqslant \delta \leqslant \lfloor p/2\rfloor$, we defined $S(p,\delta)$ as the number of permutations on $\{1,\ldots,p\}$ having *genus* δ, i.e. $S(p,\delta) = \left|\{\alpha \in \mathfrak{S}(p) \ : \ \sharp(\gamma^{-1}\alpha) + \sharp(\alpha) = p+1-2\delta\}\right|$. Since $\{\alpha \in \mathfrak{S}(p) \ : \ \sharp(\gamma^{-1}\alpha)+\sharp(\alpha) = p+1\} = NC(p)$, we have $S(p,0) = \mathrm{Cat}_p$ and hence recover the well-known asymptotic estimate

$$\mathbf{E}_{W\sim\mathscr{W}_{n,n}} \mathrm{Tr}\left(W^p\right) \underset{n\to+\infty}{\sim} \mathrm{Cat}_p\, n^{p+1}.$$

9.12 Appendix C: A needed combinatorial fact: relating the number of cycles in some specific permutations on either $[p]\times[k]$ or $[p]$

Let $\alpha \in \mathfrak{S}(p)$ and $f : [p] \to [k]$. We define $\hat{\alpha}_f$ on $[p]\times[k]$ as

$$\forall\ (i,r) \in [p]\times[k],\ \hat{\alpha}_f(i,r) = \begin{cases} (\alpha(i), f(\alpha(i))) \text{ if } r = f(i) \\ (i,r) \text{ if } r \neq f(i) \end{cases}.$$

We also define $\hat{\gamma}$ on $[p]\times[k]$ as (γ, id), where $\gamma \in \mathfrak{S}(p)$ is the canonical full cycle $(p\ \ldots\ 1)$.

We would like to understand what is the number of cycles in $\hat{\gamma}^{-1}\hat{\alpha}_f$. For that, it will be convenient to do a bit of rewriting. Let us first extend the definition of $\hat{\alpha}_f$ and $\hat{\gamma}$ to $[p]\times(\{0\}\cup[k])$. We shall denote by $\bar{\alpha}_f$ and $\bar{\gamma}$ the respective extensions. Note that since f takes values in $[k]$, we have $\bar{\alpha}_f(i,0) = (i,0)$ for all $i \in [p]$.

We will now make two easy observations.

Fact 9.12.1. *For any $f : [p] \to [k]$, define for each $i \in [p]$, $\bar{\tau}_f^{(i)}$ as the transposition on $[p]\times(\{0\}\cup[k])$ which swaps $(i,0)$ and $(i,f(i))$, and set $\bar{\beta}_f = \bar{\tau}_f^{(1)}\cdots\bar{\tau}_f^{(p)}$. We then have, for any $\alpha \in \mathfrak{S}(p)$,*

$$\bar{\alpha}_f = \bar{\beta}_f^{-1}\bar{\alpha}'\bar{\beta}_f,\ \textit{where}\ \forall\ (i,r) \in [p]\times(\{0\}\cup[k]),\ \bar{\alpha}'(i,r) = \begin{cases} (\alpha(i),r)\ \textit{if}\ r = 0 \\ (i,r)\ \textit{if}\ r \neq 0 \end{cases}.$$

The advantage of expressing $\bar{\alpha}_f$ in this way is that $\bar{\alpha}'$ is particularly simple: it acts as $\alpha \times \mathrm{id}$ on $[p] \times \{0\}$ and does nothing on $[p] \times [k]$. Furthermore, due to the cyclicity of $\sharp(\cdot)$, a direct consequence of Fact 9.12.1 is that $\sharp(\bar{\gamma}^{-1}\bar{\alpha}_f) = \sharp(\bar{\gamma}_f^{-1}\bar{\alpha}')$, where $\bar{\gamma}_f = \bar{\beta}_f \bar{\gamma} \bar{\beta}_f^{-1}$. It may then be easily checked that $\bar{\gamma}_f$ decomposes into $k+1$ disjoint cycles as stated in Fact 9.12.2 below.

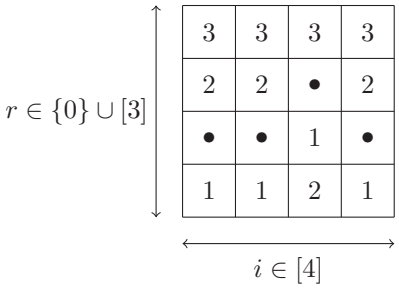
Fact 9.12.2. *For any $f : [p] \to [k]$, we have*

$$\bar{\gamma}_f = \bar{c}_1 \cdots \bar{c}_k \bar{c},$$

with $\bar{c} = ((p, f(p)) \dots (1, f(1)))$, and for each $r \in [k]$, $\bar{c}_r = ((p, s_r(p)) \dots (1, s_r(1)))$, where for each $i \in [p]$, $s_r(i) = 0$ if $f(i) = r$ and $s_r(i) = r$ if $f(i) \neq r$.

Example 9.12.3. *For the sake of concreteness, let us have a look at a simple example. In the case where $p = 4$, $k = 3$, and f is defined by $f(1) = f(2) = f(4) = 1$, $f(3) = 2$, we obtain that the cycles in $\bar{\gamma}_f$ are $\bar{c}_1 = ((4,0)(3,1)(2,0)(1,0))$, $\bar{c}_2 = ((4,2)(3,0)(2,2)(1,2))$, $\bar{c}_3 = ((4,3)(3,3)(2,3)(1,3))$, $\bar{c} = ((4,1)(3,2)(2,1)(1,1))$. This is schematically represented in Figure 9.2, where the elements in $\bar{c}_i$ are marked by "i", for $i \in \{1,2,3\}$, and the elements in $\bar{c}$ are marked by "$\bullet$".*

Figure 9.2: $f : [4] \to [3]$ such that $f^{-1}(1) = \{1,2,4\}$, $f^{-1}(2) = \{3\}$, $f^{-1}(3) = \emptyset$.



Lemma 9.12.4. *Let $f : [p] \to [k]$ and define $\gamma_f \in \mathfrak{S}(p)$ as $\gamma_f = \gamma_{f=1} \cdots \gamma_{f=k}$, where for each $r \in [k]$, $\gamma_{f=r}$ is the canonical full cycle on $f^{-1}(r)$. Then, for any $\alpha \in \mathfrak{S}(p)$,*

$$\sharp(\bar{\gamma}_f^{-1}\bar{\alpha}') = \sharp(\gamma_f^{-1}\alpha) + 1 + k - |\mathrm{Im}(f)|. \tag{9.29}$$

Proof. In $\bar{\gamma}_f^{-1}\bar{\alpha}'$ there are, first of all:
• $k - |\mathrm{Im}(f)|$ cycles of the form $((1,r) \dots (p,r))$ for $r \in [k] \setminus \mathrm{Im}(f)$, because for any $(i,r) \in [p] \times ([k] \setminus \mathrm{Im}(f))$, $\bar{\gamma}_f^{-1}\bar{\alpha}'(i,r) = \bar{\gamma}_f^{-1}(i,r) = (\gamma^{-1}(i), r)$.
• 1 cycle $((1, f(1)) \dots (p, f(p)))$, because for any $i \in [p]$, $\bar{\gamma}_f^{-1}\bar{\alpha}'(i, f(i)) = \bar{\gamma}_f^{-1}(i, f(i)) = (\gamma^{-1}(i), f(\gamma^{-1}(i)))$.
For the cycles in $\bar{\gamma}_f^{-1}\bar{\alpha}'$ which belong to none of these two categories, there are two crucial observations to be made. First, for any $i, j \in [p]$, $(i,0)$ and $(j,0)$ belong to the same cycle of $\bar{\gamma}_f^{-1}\bar{\alpha}'$ if and only if i and j belong to the same cycle of $\gamma_f^{-1}\alpha$. And second, for each $i \in [p]$ and $r \in [k] \setminus \{f(\alpha(i))\}$, there exists $j \in [p]$ such that (i,r) belongs to the same cycle of $\bar{\gamma}_f^{-1}\bar{\alpha}'$ as $(j,0)$. Indeed, for any $i \in [p]$, we have on the one hand

$$\gamma_f^{-1}\alpha(i) = (\gamma^{-1})^{L+1}\alpha(i), \text{ with } L = \inf\{l \geqslant 0 \ : \ f((\gamma^{-1})^{l+1}\alpha(i)) = f(\alpha(i))\}.$$

While we have on the other hand,

$$f(\gamma^{-1}\alpha(i)) = f(\alpha(i)) \ \Rightarrow \ \bar{\gamma}_f^{-1}\bar{\alpha}'(i,0) = (\gamma^{-1}\alpha(i), 0) = (\gamma_f^{-1}\alpha(i), 0),$$
$$f(\gamma^{-1}\alpha(i)) \neq f(\alpha(i)) \ \Rightarrow \ \begin{cases} \forall\ 0 \leqslant l \leqslant L-1, \ (\bar{\gamma}_f^{-1}\bar{\alpha}')^l(i,0) = ((\gamma^{-1})^l\gamma^{-1}\alpha(i), f(\alpha(i))) \\ (\bar{\gamma}_f^{-1}\bar{\alpha}')^L(i,0) = ((\gamma^{-1})^L\gamma^{-1}\alpha(i), 0) = (\gamma_f^{-1}\alpha(i), 0) \end{cases}.$$

So there are in fact exactly $\sharp(\gamma_f^{-1}\alpha)$ remaining cycles in $\bar{\gamma}_f^{-1}\bar{\alpha}'$. □

Example 9.12.5. *Looking at the same example as before, namely $p = 4$, $k = 3$, and f such that $f^{-1}(1) = \{1,2,4\}$, $f^{-1}(2) = \{3\}$, $f^{-1}(3) = \emptyset$, we see that, for $\alpha = (14)(23)$, the cycles in $\bar{\gamma}_f^{-1}\bar{\alpha}'$ are:*
• $((1,3)(2,3)(3,3)(4,3))$, *because $3 \notin \mathrm{Im}(f)$.*
• $((1,1)(2,1)(3,2)(4,1))$, *because $f(1) = 1$, $f(2) = 1$, $f(3) = 2$ and $f(4) = 1$.*
• $(1,0)$ *and* $((2,0)(2,2)(1,2)(4,2)(3,0)(4,0)(3,1))$, *corresponding to the cycles* (1) *and* $(2,3,4)$ *in* $\gamma_f^{-1}\alpha$.

Putting together these preliminary technical results, we straightforwardly obtain Proposition 9.12.6 below.

Proposition 9.12.6. *Let $f : [p] \to [k]$ and define $\gamma_f \in \mathfrak{S}(p)$ as $\gamma_f = \gamma_{f=1} \cdots \gamma_{f=k}$, where for each $r \in [k]$, $\gamma_{f=r}$ is the canonical full cycle on $f^{-1}(r)$. Then, for any $\alpha \in \mathfrak{S}(p)$,*

$$\sharp(\hat{\gamma}^{-1}\hat{\alpha}_f) = \sharp(\gamma_f^{-1}\alpha) + k - |\operatorname{Im}(f)|. \tag{9.30}$$

Proof. This is a direct consequence of Lemma 9.12.4, just noticing that $\sharp(\bar{\gamma}^{-1}\bar{\alpha}_f) = \sharp(\hat{\gamma}^{-1}\hat{\alpha}_f) + 1$. □

9.13 Appendix D: Moments of "modified" GUE matrices (proof)

The goal of this Appendix is to generalize the methodology described in Appendix 9.11.1 in order to compute the $2p$-order moments of the matrix $\sum_{j=1}^{k} \widetilde{G}_{\mathrm{AB}^k}(j)$. Recall that this issue arises when trying to estimate the mean width of the set of k-extendible states. We are thus dealing here, not with standard GUE matrices, but with d^2-dimensional GUE matrices which are tensorized with d^{k-1}-dimensional identity matrices.

For any $i_1, \ldots, i_{2p} \in [k]$, we can write

$$\mathrm{Tr}\left[\overrightarrow{\prod_{j=1}^{2p}} \widetilde{G}_{\mathrm{AB}^k}(i_j)\right] = \sum_{\vec{l}_1,\ldots,\vec{l}_{2p} \in [d]^{k+1}} \widetilde{G}_{\mathrm{AB}^k}(i_1)_{\vec{l}_1,\vec{l}_2} \cdots \widetilde{G}_{\mathrm{AB}^k}(i_{2p})_{\vec{l}_{2p},\vec{l}_1},$$

where for each $j \in [2p]$ and each $\vec{l}_j = \left(a_j, b_j^1, \ldots, b_j^k\right), \vec{l}_{j+1} = \left(a_{j+1}, b_{j+1}^1, \ldots, b_{j+1}^k\right) \in [d]^{k+1}$, we have

$$\widetilde{G}_{\mathrm{AB}^k}(i_j)_{\vec{l}_j,\vec{l}_{j+1}} = G_{(a_j,b_j^{i_j}),(a_{j+1},b_{j+1}^{i_j})} \delta_{\vec{b}_j \setminus b_j^{i_j} = \vec{b}_{j+1} \setminus b_{j+1}^{i_j}}.$$

Consequently, for any $f : [2p] \to [k]$, we have

$$\mathrm{Tr}\left[\overrightarrow{\prod_{i=1}^{2p}} \widetilde{G}_{\mathrm{AB}^k}(f(i))\right] = \sum_{a_1,\ldots,a_{2p} \in [d]} \sum_{\vec{b}_1,\ldots,\vec{b}_{2p} \in I_f} G_{(a_1,b_1^{f(1)}),(a_2,b_2^{f(1)})} \cdots G_{(a_{2p},b_{2p}^{f(2p)}),(a_1,b_1^{f(2p)})},$$

where $I_f = \left\{\vec{b}_1, \ldots, \vec{b}_{2p} \in [d]^k \ : \ \forall\, i \in [2p],\ \forall\, r \in [k] \setminus \{f(i)\},\ b_{i+1}^r = b_i^r\right\}$.

What we therefore get by the Wick formula for Gaussian matrices is that, for any $f : [2p] \to [k]$,

$$\mathbf{E}_{G_{\mathrm{AB}} \sim GUE(d^2)} \mathrm{Tr}\left[\overrightarrow{\prod_{i=1}^{2p}} \widetilde{G}_{\mathrm{AB}^k}(f(i))\right] = \sum_{\lambda \in \mathfrak{P}^{(2)}(2p)} d^{\flat(\lambda) + \flat(\hat{\lambda}_f)},$$

where for each pair partition $\lambda = \{\{i_1, j_1\}, \ldots, \{i_p, j_p\}\}$ of $\{1, \ldots, 2p\}$, $\flat(\lambda)$ is the number of free parameters $a_1, \ldots, a_{2p} \in [d]$ when imposing that $\forall\ 1 \leqslant m \leqslant p,\ a_{i_m+1} = a_{j_m},\ a_{j_m+1} = a_{i_m}$, and $\flat(\hat{\lambda}_f)$ is the number of free parameters $\vec{b}_1, \ldots, \vec{b}_{2p} \in I_f$ when imposing that $\forall\ 1 \leqslant m \leqslant p,\ b_{i_m+1}^{f(i_m)} = b_{j_m}^{f(j_m)},\ b_{j_m+1}^{f(j_m)} = b_{i_m}^{f(i_m)}$. As noticed before, identifying the pair partition $\{\{i_1, j_1\}, \ldots, \{i_p, j_p\}\}$ with the pairing $(i_1\, j_1) \ldots (i_p\, j_p)$ and denoting by γ the canonical full cycle $(2p \ldots 1)$, the latter conditions may be written as

$$\forall\ i \in [2p],\ a_{\gamma^{-1}\lambda(i)} = a_i \text{ and } \forall\ r \in [k],\ \begin{cases} b_{\gamma^{-1}\lambda(i)}^{f(\lambda(i))} = b_i^r \text{ if } r = f(i) \\ b_{\gamma^{-1}(i)}^r = b_i^r \text{ if } r \neq f(i) \end{cases}.$$

So in fact, $\flat(\lambda) = \sharp(\gamma^{-1}\lambda)$ and $\flat(\hat{\lambda}_f) = \sharp(\hat{\gamma}^{-1}\hat{\lambda}_f)$, where

$$\forall\ (i,r) \in [2p] \times [k],\ \hat{\gamma}(i,r) = (\gamma(i), r) \text{ and } \hat{\lambda}_f(i,r) = \begin{cases} (\lambda(i), f(\lambda(i))) \text{ if } r = f(i) \\ (i,r) \text{ if } r \neq f(i) \end{cases}.$$

What is more, we know by Proposition 9.12.6 that $\sharp(\hat{\gamma}^{-1}\hat{\lambda}_f) = \sharp(\gamma_f^{-1}\lambda) + k - |\operatorname{Im}(f)|$, where γ_f is the product of the canonical full cycles on the level sets of f.

Let us summarize.

Proposition 9.13.1. *For any $d \in \mathbf{N}$ and any $p \in \mathbf{N}$, we have*

$$\begin{aligned} \mathbf{E}_{G_{\mathrm{AB}} \sim GUE(d^2)} \mathrm{Tr}\left[\left(\sum_{j=1}^{k} \widetilde{G}_{\mathrm{AB}^k}(j)\right)^{2p}\right] &= \sum_{f:[2p]\to[k]} \mathbf{E}_{G_{\mathrm{AB}} \sim GUE(d^2)} \mathrm{Tr}\left[\overrightarrow{\prod_{i=1}^{2p}} \widetilde{G}_{\mathrm{AB}^k}(f(i))\right] \\ &= \sum_{f:[2p]\to[k]} \sum_{\lambda \in \mathfrak{P}^{(2)}(2p)} d^{\sharp(\gamma^{-1}\lambda) + \sharp(\gamma_f^{-1}\lambda) + k - |\operatorname{Im}(f)|}. \end{aligned}$$

9.14 Appendix E: Moments of partially transposed "modified" GUE matrices (proof)

The goal of this Appendix is to compute the $2p$-order moments of $\sum_{j=1}^k \widetilde{G}_{\mathrm{AB}^k}(j)^\Gamma$, where Γ stands here for the partial transposition over the $\lceil k/2 \rceil$ last B subsystems. Recall that this issue arises when trying to estimate the mean width of the set of k-PPT-extendible states.

Using the same notation as in Appendix 9.13, and reasoning in a completely analogous way, we have that, for any $f : [2p] \to [k]$,

$$\mathrm{Tr}\left[\overrightarrow{\prod_{i=1}^{2p}} \widetilde{G}_{\mathrm{AB}^k}(f(i))^\Gamma\right] = \sum_{a_1,\ldots,a_{2p}\in[d]} \sum_{\vec{b}_1,\ldots,\vec{b}_{2p}\in I_f} G_{(a_1,b_{x_1}^{f(1)}),(a_2,b_{\bar{x}_1}^{f(1)})} \cdots G_{(a_{2p},b_{x_{2p}}^{f(2p)}),(a_1,b_{\bar{x}_{2p}}^{f(2p)})},$$

where for each $1 \leqslant i \leqslant 2p$, $x_i = \begin{cases} i \text{ if } f(i) \leqslant \lfloor k/2 \rfloor \\ i+1 \text{ if } f(i) > \lfloor k/2 \rfloor \end{cases}$ and $\bar{x}_i = \begin{cases} i+1 \text{ if } f(i) \leqslant \lfloor k/2 \rfloor \\ i \text{ if } f(i) > \lfloor k/2 \rfloor \end{cases}$.

What we therefore get by the Wick formula for Gaussian matrices is that, for any $f : [2p] \to [k]$,

$$\mathbf{E}_{G_{\mathrm{AB}}\sim GUE(d^2)} \mathrm{Tr}\left[\overrightarrow{\prod_{i=1}^{2p}} \widetilde{G}_{\mathrm{AB}^k}(f(i))^\Gamma\right] = \sum_{\lambda\in\mathfrak{P}^{(2)}(2p)} d^{\flat(\lambda)+\flat(\check{\lambda}_f)},$$

where for each pair partition $\lambda = \{\{i_1, j_1\}, \ldots, \{i_p, j_p\}\}$ of $\{1, \ldots, 2p\}$, $\flat(\check{\lambda}_f)$ is the number of free parameters $\vec{b}_1, \ldots, \vec{b}_{2p} \in [d]^k$ when imposing that for all $i \in [2p]$, first $b^r_{\gamma^{-1}(i)} = b^r_i$ if $r \neq f(i)$, and second the one condition $b^{f(\lambda(i))}_{\gamma^{-1}\lambda(i)} = b^{f(i)}_i$ if $f(i), f(\lambda(i)) \leqslant \lfloor k/2 \rfloor$ or $f(i), f(\lambda(i)) > \lfloor k/2 \rfloor$, while the two conditions $b^{f(\lambda(i))}_{\gamma^{-1}\lambda(i)} = b^{f(i)}_i$ and $b^{f(\lambda(i))}_{\lambda(i)} = b^{f(i)}_{\gamma^{-1}(i)}$ if $f(i) \leqslant \lfloor k/2 \rfloor$, $f(\lambda(i)) > \lfloor k/2 \rfloor$ or $f(i) > \lfloor k/2 \rfloor$, $f(\lambda(i)) \leqslant \lfloor k/2 \rfloor$.

Let us rephrase what we just established. Fix $\lambda \in \mathfrak{P}^{(2)}(2p)$. For functions $f : [2p] \to \{1, \ldots, \lfloor k/2 \rfloor\} \equiv [\lfloor k/2 \rfloor]$ or $f : [2p] \to \{\lfloor k/2 \rfloor + 1, \ldots, k\} \equiv [\lceil k/2 \rceil]$, the number of free parameters associated to the pair (λ, f) is the same as the one observed in Appendix 9.13. On the contrary, for functions f which are not of this form, extra matching conditions are imposed. So these will for sure not contribute to the dominating term in the expansion of $\mathbf{E}\,\mathrm{Tr}\left[\left(\sum_{j=1}^k \widetilde{G}_{\mathrm{AB}^k}(j)^\Gamma\right)^{2p}\right]$ into powers of d. Consequently, we have the asymptotic estimate

$$\mathbf{E}_{G_{\mathrm{AB}}\sim GUE(d^2)} \mathrm{Tr}\left[\left(\sum_{j=1}^k \widetilde{G}_{\mathrm{AB}^k}(j)^\Gamma\right)^{2p}\right] \underset{d\to+\infty}{\sim} \sum_{\lambda\in\mathfrak{P}^{(2)}(2p)} \sum_{f:[2p]\to[\lfloor k/2\rfloor] \text{ or } [\lceil k/2\rceil]} d^{\sharp(\gamma^{-1}\lambda)+\sharp(\gamma_f^{-1}\lambda)+k-|\,\mathrm{Im}(f)|}.$$

9.15 Appendix F: Moments of "modified" Wishart matrices (proof)

The goal of this Appendix is to generalize the methodology described in Appendix 9.11.2 in order to compute the p-order moments of the matrix $\sum_{j=1}^k \widetilde{W}_{\mathrm{AB}^k}(j)$. Recall that this issue arises when trying to characterize k-extendibility of random-induced states. We are thus dealing here, not with standard Wishart matrices, but with (d^2, s)-Wishart matrices which are tensorized with d^{k-1}-dimensional identity matrices.

Representing by $\bullet$ a d-dimensional gate and by $\blacktriangledown$ a s-dimensional gate, the matrix $\widetilde{W}_{\mathrm{AB}^k}(1)$, for instance, may be graphically represented as in Figure 9.3.

The products $\widetilde{W}_{\mathrm{AB}^k}(1)\widetilde{W}_{\mathrm{AB}^k}(1)$ and $\widetilde{W}_{\mathrm{AB}^k}(1)\widetilde{W}_{\mathrm{AB}^k}(2)$, for instance, are then obtained by the wirings represented in Figure 9.4.

So what we get by the graphical Wick formula for Wishart matrices is that for any $f : [p] \to [k]$,

$$\mathbf{E}_{W_{\mathrm{AB}}\sim\mathscr{W}_{d^2,s}} \mathrm{Tr}\left[\overrightarrow{\prod_{i=1}^{p}} \widetilde{W}_{\mathrm{AB}^k}(f(i))\right] = \sum_{\alpha\in\mathfrak{S}(p)} \mathscr{D}_{f,\alpha} = \sum_{\alpha\in\mathfrak{S}(p)} d^{\sharp(\gamma^{-1}\alpha)} d^{\sharp(\hat{\gamma}^{-1}\hat{\alpha}_f)} s^{\sharp(\alpha)},$$

where $\hat{\alpha}_f$ is defined by

$$\hat{\alpha}_f : (i, r) \in [p] \times [k] \mapsto \begin{cases} (\alpha(i), f(\alpha(i))) \text{ if } r = f(i) \\ (i, r) \text{ if } r \neq f(i) \end{cases},$$

Figure 9.3: $\widetilde{W}_{\mathrm{AB}^k}(1) = X_{\mathrm{AB}^k}(1)X_{\mathrm{AB}^k}(1)^\dagger$, with $X_{\mathrm{AB}^k}(1) = G_{\mathrm{AB}_1} \otimes \mathrm{Id}_{\mathrm{B}_2 \ldots \mathrm{B}_k}$

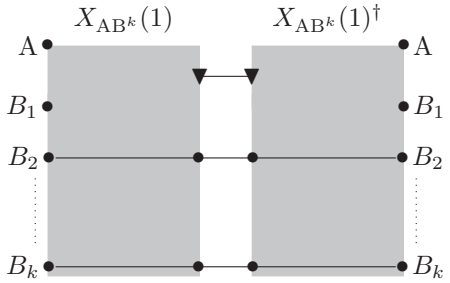
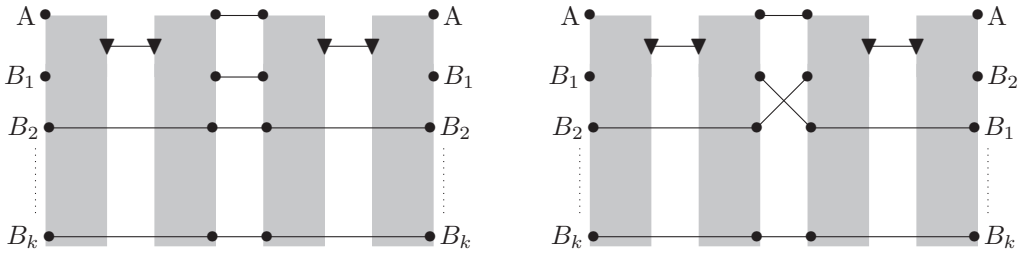


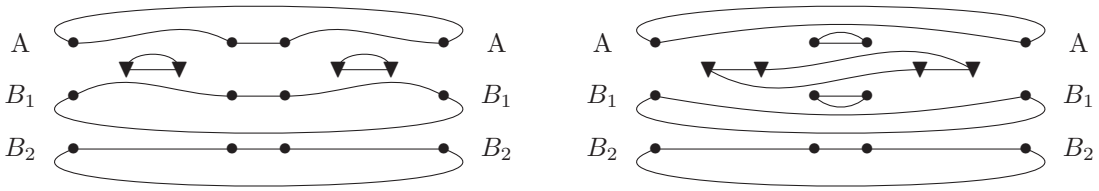
Figure 9.4: $\widetilde{W}_{\mathrm{AB}^k}(1)\widetilde{W}_{\mathrm{AB}^k}(1)$ (on the left) and $\widetilde{W}_{\mathrm{AB}^k}(1)\widetilde{W}_{\mathrm{AB}^k}(2)$ (on the right)



and where $\hat{\gamma}$ stands for γ applied to the first argument. Indeed, for each $\alpha \in \mathfrak{S}(p)$, there are $\sharp(\gamma^{-1}\alpha)$ loops connecting the d-dimensional gates corresponding to A, $\sharp(\hat{\gamma}^{-1}\hat{\alpha}_f)$ loops connecting the d-dimensional gates corresponding to $B_1, \ldots, B_k$, and $\sharp(\alpha)$ loops connecting s-dimensional gates. This is because for each $1 \leqslant i \leqslant p$, on subsystems A and $B_{f(i)}$, the entrances (respectively the exit) of the ith box $X_{\mathrm{AB}^k}(f(i))$ are connected to the exits (respectively the entrance) of the $\alpha(i)^{th}$ box $X_{\mathrm{AB}^k}(f(\alpha(i)))^\dagger$.

What happens in the special case $p = 2$ and $k = 2$ is detailed in Figures 9.5 and 9.6 below as an illustration.

Figure 9.5: $f(1) = f(2) = 1$. On the left, $\alpha = \mathrm{id}$: $\mathscr{D}_{f,\alpha} = d^3 s^2$. On the right, $\alpha = (1\,2)$: $\mathscr{D}_{f,\alpha} = d^5 s$.



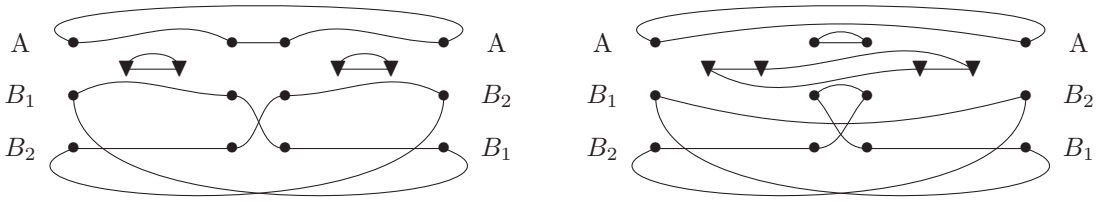
Finally, we also know by Proposition 9.12.6 that for any $\alpha \in \mathfrak{S}(p)$ and $f : [p] \to [k]$, denoting by γ_f the product of the canonical full cycles on the level sets of f, we have $\sharp(\hat{\gamma}^{-1}\hat{\alpha}_f) = \sharp(\gamma_f^{-1}\alpha) + k - |\operatorname{Im}(f)|$.

Putting everything together, we eventually come to the result summarized in Proposition 9.15.1 below.

Proposition 9.15.1. *For any $d, s \in \mathbf{N}$ and any $p \in \mathbf{N}$, we have*

$$\begin{aligned}\mathbf{E}_{W_{\mathrm{AB}} \sim \mathscr{W}_{d^2,s}} \operatorname{Tr}\left[\left(\sum_{j=1}^{k} \widetilde{W}_{\mathrm{AB}^k}(j)\right)^p\right] &= \sum_{f:[p]\to[k]} \mathbf{E}_{W_{\mathrm{AB}} \sim \mathscr{W}_{d^2,s}} \operatorname{Tr}\left[\overrightarrow{\prod_{i=1}^{p}} \widetilde{W}_{\mathrm{AB}^k}(f(i))\right] \\ &= \sum_{f:[p]\to[k]} \sum_{\alpha \in \mathfrak{S}(p)} d^{\sharp(\gamma^{-1}\alpha)+\sharp(\gamma_f^{-1}\alpha)+k-|\operatorname{Im}(f)|} s^{\sharp(\alpha)}.\end{aligned}$$

Figure 9.6: $f(1)=1$, $f(2)=2$. On the left, $\alpha=\mathrm{id}$: $\mathscr{D}_{f,\alpha}=d^3s^2$. On the right, $\alpha=(1\,2)$: $\mathscr{D}_{f,\alpha}=d^3s$.



9.16 Appendix G: Counting geodesics vs non-geodesics pairings and permutations

Let us recall once and for all two notation that we will use repeatedly in this section, and that were introduced in Lemma 9.10.4. For any $p,m\in\mathbf{N}$ with $m\leqslant p$, we denote by $\mathrm{Cat}_p=\frac{1}{p+1}\binom{2p}{p}$ the pth Catalan number, and by $\mathrm{Nar}_p^m=\frac{1}{p+1}\binom{p+1}{m}\binom{p-1}{m-1}$ the $(p,m)^{th}$ Narayana number.

9.16.1 Number of pairings of $2p$ elements which are not on the geodesics between the identity and the canonical full cycle

Lemma 9.16.1. *Let $p\in\mathbf{N}$ and denote by γ the canonical full cycle on $\{1,\dots,2p\}$. For any $0\leqslant\delta\leqslant\lfloor p/2\rfloor$, define the set of pairings having a defect 2δ of being on the geodesics between* id *and γ as*

$$\mathfrak{P}_\delta^{(2)}(2p)=\{\lambda\in\mathfrak{P}^{(2)}(2p)\ :\ \sharp(\gamma^{-1}\lambda)=p+1-2\delta\}.$$

Then, the cardinality of $\mathfrak{P}_\delta^{(2)}(2p)$ is upper bounded by $\mathrm{Cat}_p\left(p^4/4\right)^\delta$.

To prove Lemma 9.16.1 (and later on Lemma 9.16.3) we will need the simple observation below. Roughly speaking, it will allow us to assume without loss of generality that, in the decomposition of an element of $\mathfrak{P}_\delta^{(2)}(2p)$ into p disjoint transpositions, the ones "creating" the 2δ geodesic defects are the 2δ first ones.

Fact 9.16.2. *Let ς be a permutation on $\{1,\dots,q\}$ and τ_1,τ_2,τ_3 be three disjoint transpositions on $\{1,\dots,q\}$, for some integer $q\geqslant 6$. Define $\varsigma^{(1)}=\varsigma\,\tau_1$, $\varsigma^{(2)}=\varsigma\,\tau_1\tau_2$, $\varsigma^{(3)}=\varsigma\,\tau_1\tau_2\tau_3$, and assume that*

$$\sharp(\varsigma^{(1)})=\sharp(\varsigma)+1,\ \sharp(\varsigma^{(2)})=\sharp(\varsigma)+2,\ \sharp(\varsigma^{(3)})=\sharp(\varsigma)+1. \tag{9.31}$$

Then, there exists a permutation π of the three indices $\{1,2,3\}$ such that, defining this time $\varsigma_\pi^{(1)}=\varsigma\,\tau_{\pi(1)}$, $\varsigma_\pi^{(2)}=\varsigma\,\tau_{\pi(1)}\tau_{\pi(2)}$, $\varsigma_\pi^{(3)}=\varsigma\,\tau_{\pi(1)}\tau_{\pi(2)}\tau_{\pi(3)}$, we have

$$\sharp(\varsigma_\pi^{(1)})=\sharp(\varsigma)+1,\ \sharp(\varsigma_\pi^{(2)})=\sharp(\varsigma),\ \sharp(\varsigma_\pi^{(3)})=\sharp(\varsigma)+1.$$

Proof. Assume that ς and $\tau_1=(i_1\,j_1),\tau_2=(i_2\,j_2),\tau_3=(i_3\,j_3)$ satisfy equation (9.31). This means that i_1,j_1 belong to the same cycle of ς, i_2,j_2 belong to the same cycle of $\varsigma^{(1)}$, and i_3,j_3 belong to two different cycles of $\varsigma^{(2)}$. So let us inspect all the scenarios which may occur.
• $c_1\overset{(i_1\,j_1)}{\to}c_1^xc_1^y$ and $c_2\overset{(i_2\,j_2)}{\to}c_2^xc_2^y$, with c_1,c_2 two different cycles of ς: If $i_3\in c_1^x$ and $j_3\in c_1^y$ then the re-ordering $1,3,2$ is suitable. If $i_3\in c_2^x$ and $j_3\in c_2^y$ then the re-ordering $2,3,1$ is suitable. If $i_3\in c_1^a$ and $j_3\in c_2^b$, for $a,b\in\{x,y\}$, then both re-orderings $1,3,2$ and $2,3,1$ are suitable. And similarly when the roles of i_3 and j_3 are exchanged.
• $c\overset{(i_1\,j_1)}{\to}c'c''\overset{(i_2\,j_2)}{\to}c^xc^yc^z$, with c a cycle of ς, while $c^z=c''$ and $c'\overset{(i_2\,j_2)}{\to}c^xc^y$: If $i_3\in c^x$ and $j_3\in c^y$ then the re-ordering $2,3,1$ is suitable. If $i_3\in c^a$, for $a\in\{x,y\}$, and $j_3\in c^z$ then the re-ordering $1,3,2$ is suitable. And similarly when the roles of i_3 and j_3 are exchanged. □

As an immediate consequence of Fact 9.16.2, we have the following: Let $\varsigma\in\mathfrak{S}(2p)$ and $\lambda=\tau_1\cdots\tau_p\in\mathfrak{P}^{(2)}(2p)$. Define for each $1\leqslant q\leqslant p$, $\varsigma^{(q)}=\varsigma\,\tau_1\cdots\tau_q$, as well as $\varsigma^{(0)}=\varsigma$. Assume next that, for some $0\leqslant\delta\leqslant\lfloor(p+\sharp(\varsigma))/2\rfloor$,

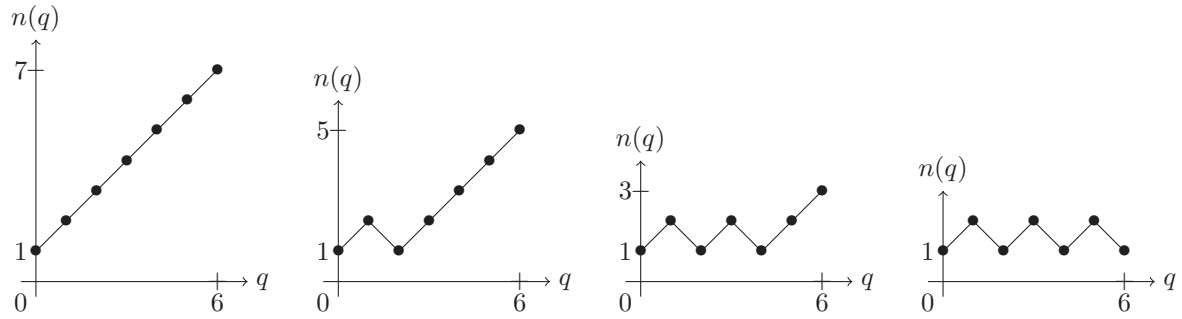
$$\sharp(\varsigma^{(p)})=\sharp(\varsigma)+p-2\delta.$$

Then, there exists a permutation π of the p indices $\{1,\dots,p\}$ such that, defining this time for each $1 \leqslant q \leqslant p$, $\varsigma_\pi^{(q)} = \varsigma\,\tau_{\pi(1)}\cdots\tau_{\pi(q)}$, as well as $\varsigma_\pi^{(0)} = \varsigma$, we have

$$\forall\ 1 \leqslant q \leqslant p,\ \begin{cases} \sharp(\varsigma_\pi^{(q)}) = \sharp(\varsigma_\pi^{(q-1)}) - 1 \text{ if } q \in \{2\epsilon\ :\ 1 \leqslant \epsilon \leqslant \delta\} \\ \sharp(\varsigma_\pi^{(q)}) = \sharp(\varsigma_\pi^{(q)}) + 1 \text{ if } q \notin \{2\epsilon\ :\ 1 \leqslant \epsilon \leqslant \delta\} \end{cases}. \tag{9.32}$$

Since $\lambda = \tau_1\cdots\tau_p = \tau_{\pi(1)}\cdots\tau_{\pi(p)}$, we see that we may always assume without loss of generality that, given ς, the transpositions $\tau_1,\dots,\tau_p$ in the decomposition of λ are ordered so that λ is under the canonical form (9.32). The behaviour of the function $q \in [p] \mapsto \sharp(\varsigma^{(q)})$ under this hypothesis, depending on the value of δ, is represented in Figure 9.7 (in the special case $p = 6$ and $\sharp(\varsigma) = 1$).

Figure 9.7: Case $p = 6$ and $\sharp(\varsigma) = 1$. From left to right: $\delta = 0$, $\delta = 1$, $\delta = 2$ and $\delta = 3$.



With this result in mind, let us now turn to the proof of Lemma 9.16.1.

Proof of Lemma 9.16.1. Given $\lambda = (i_1\,j_1)\cdots(i_p\,j_p) \in \mathfrak{P}^{(2)}(2p)$, we will always assume from now that the transpositions $(i_1\,j_1),\dots,(i_p\,j_p)$ in its decomposition are ordered so that λ is under the canonical form (9.32) for γ^{-1}. This means the following: defining, for each $1 \leqslant q \leqslant p$, the permutation $\widetilde{\lambda}^{(q)} = \gamma^{-1}(i_1\,j_1)\cdots(i_q\,j_q)$ and the integer $n(q) = \sharp(\widetilde{\lambda}^{(q)})$, as well as $\widetilde{\lambda}^{(0)} = \gamma^{-1}$ and $n(0) = \sharp(\widetilde{\lambda}^{(0)}) = 1$, we have, for any $0 \leqslant \delta \leqslant \lfloor p/2 \rfloor$,

$$\lambda \in \mathfrak{P}_\delta^{(2)}(2p) \ \Leftrightarrow\ \forall\ 1 \leqslant q \leqslant p,\ \begin{cases} n(q) = n(q-1) - 1 \text{ if } q \in \{2\epsilon\ :\ 1 \leqslant \epsilon \leqslant \delta\} \\ n(q) = n(q-1) + 1 \text{ if } q \notin \{2\epsilon\ :\ 1 \leqslant \epsilon \leqslant \delta\} \end{cases}.$$

In particular, $\lambda \in NC^{(2)}(2p) \ \Leftrightarrow\ \forall\ 1 \leqslant q \leqslant p,\ n(q) = n(q-1) + 1$, and we know that there are precisely Cat_p possibilities to build such pairing λ. This implies that, for each $1 \leqslant \delta \leqslant \lfloor p/2 \rfloor$, there are necessarily less than $\binom{2p}{2}\cdots\binom{2(p-\delta+1)}{2} \times \mathrm{Cat}_{p-2\delta}$ possibilities to build a $\lambda \in \mathfrak{P}_\delta^{(2)}(2p)$. Indeed, for the choice of the 2δ first disjoint transpositions we can use the trivial upper bound that would consist in picking them completely arbitrarily, while the $p - 2\delta$ last ones have to be chosen so that they form a non-crossing pairing of the $2p - 4\delta$ not yet selected indices. Now, we just have to observe that

$$\begin{aligned} \binom{2p}{2}\cdots\binom{2(p-2\delta+1)}{2} \times \mathrm{Cat}_{p-2\delta} &= \frac{2p\cdots(2p-4\delta+1)}{2^{2\delta}} \times \frac{(2p-4\delta)!}{(p-2\delta)!(p-2\delta+1)!} \\ &= \frac{1}{2^{2\delta}} \times \frac{p!(p+1)!}{(p-2\delta)!(p-2\delta+1)!} \times \frac{(2p)!}{p!(p+1)!} \\ &\leqslant \frac{p^{4\delta}}{2^{2\delta}} \times \mathrm{Cat}_p, \end{aligned}$$

which completes the proof. □

9.16.2 One needed generalization: bounding the number of pairings of $2p$ elements which are not on the geodesic path between the identity and a product of (few) cycles

The proof of Proposition 9.2.2 crucially relies at some point on a statement of the same kind as the one appearing in Lemma 9.16.1. Nevertheless, what we actually need there is a slight generalization of the latter. More specifically, we have to bound the number of pairings which have some defect of lying on the geodesics

between the identity and, not only a full cycle, but also a product of (few) cycles. So let us give the following extension of Lemma 9.16.1, which is really directed towards the application that we have in mind.

Lemma 9.16.3. *Let $p \in \mathbf{N}$. For any $f : [2p] \to [k]$ and any $0 \leqslant \delta \leqslant \lfloor (p + |\operatorname{Im}(f)|)/2 \rfloor$, define the set of pairings having a defect 2δ of being on the geodesics between* id *and γ_f (the product of the canonical full cycles on each of the $|\operatorname{Im}(f)|$ level sets of f) as*

$$\mathfrak{P}^{(2)}_{f,\delta}(2p) = \{\lambda \in \mathfrak{P}^{(2)}(2p) \ : \ \sharp(\gamma_f^{-1}\lambda) = p + |\operatorname{Im}(f)| - 2\delta\}.$$

Then, for any $0 \leqslant \delta \leqslant \lfloor p/2 \rfloor$, we have the upper bound

$$\left| \left\{ (f, \lambda) \ : \ \lambda \in \mathfrak{P}^{(2)}_{f,\delta}(2p) \right\} \right| \leqslant k^{p+2\delta} \mathrm{Cat}_p \left(\frac{p^4}{4} \right)^{\delta}.$$

Proof. We will follow the same strategy and employ the same notation as in the proof of Lemma 9.16.1. Given $\lambda = (i_1\, j_1) \cdots (i_p\, j_p) \in \mathfrak{P}^{(2)}(2p)$ and $f : [2p] \to [k]$, we will always assume that the transpositions $(i_1\, j_1), \ldots, (i_p\, j_p)$ in the decomposition of λ are ordered so that λ is under the canonical form (9.32) for γ_f^{-1}. This means the following: defining, for each $1 \leqslant q \leqslant p$, $\widetilde{\lambda}^{(q)} = \gamma_f^{-1}(i_1\, j_1) \cdots (i_q\, j_q)$ and $n(q) = \sharp(\widetilde{\lambda}^{(q)})$, as well as $\widetilde{\lambda}^{(0)} = \gamma_f^{-1}$ and $n(0) = \sharp(\widetilde{\lambda}^{(0)}) = |\operatorname{Im}(f)|$, we have, for any $0 \leqslant \delta \leqslant \lfloor (p + |\operatorname{Im}(f)|)/2 \rfloor$,

$$\lambda \in \mathfrak{P}^{(2)}_{f,\delta}(2p) \ \Leftrightarrow \ \forall\, 1 \leqslant q \leqslant p, \ \begin{cases} n(q) = n(q-1) - 1 \text{ if } q \in \{2\epsilon \ : \ 1 \leqslant \epsilon \leqslant \delta\} \\ n(q) = n(q-1) + 1 \text{ if } q \notin \{2\epsilon \ : \ 1 \leqslant \epsilon \leqslant \delta\} \end{cases} . \tag{9.33}$$

In particular, $\lambda \in \mathfrak{P}^{(2)}_{f,0}(2p) \ \Leftrightarrow \ \forall\ 1 \leqslant q \leqslant p, \ n(q) = n(q-1) + 1$, and we know that there are precisely $k^p\, \mathrm{Cat}_p$ possibilities to build a pair (f, λ) satisfying this condition (because the latter holds if and only if both constraints $\lambda \in NC^{(2)}(2p)$ and $f \circ \lambda = f$ are fulfilled).

In the case $1 \leqslant \delta \leqslant \lfloor p/2 \rfloor$, notice that after the 2δ first steps, we are left with a permutation $\overline{\varsigma}$ having $|\operatorname{Im}(f)|$ cycles, and we have to impose that the partial pairing $\overline{\lambda} = (i_{2\delta+1}\, j_{2\delta+1}) \cdots (i_{2p}\, j_{2p})$ lies on the geodesics between id and $\overline{\varsigma}$. Now, the number of such partial pairings is the same as the number of partial pairings lying on the geodesics between id and $\gamma_{\overline{f}}$, for any function $\overline{f} : [2p] \to [k]$ whose level sets are the supports of the cycles of $\overline{\varsigma}$. Hence, to build a pair $(\overline{f}, \lambda)$ meeting our requirements, we have at most $k^{4\delta} \binom{2p}{2} \cdots \binom{2(p-\delta+1)}{2} \times k^{p-2\delta}\, \mathrm{Cat}_{p-2\delta}$ possibilities. Indeed, for the 2δ first disjoint transpositions we can use the trivial upper bound that would consist in picking them, as well as the values of $\overline{f}$ on them, completely arbitrarily, while for the $p - 2\delta$ last ones we have to impose that they are non-crossing and that $\overline{f}$ takes only one value on a given transposition. Now, we know from the proof of Lemma 9.16.1 that $\binom{2p}{2} \cdots \binom{2(p-\delta+1)}{2} \mathrm{Cat}_{p-2\delta} \leqslant (p^4/4)^{\delta} \mathrm{Cat}_p$. So we get as announced that there are less than $k^{p+2\delta}\, \mathrm{Cat}_p (p^4/4)^{\delta}$ pairs (f, λ) satisfying condition (9.33). □

9.16.3 One needed adaptation: bounding the number of permutations of p elements which are not on the geodesic path between the identity and a product of (few) cycles

The proof of Proposition 9.6.1 requires a statement analogous to the one appearing in Lemma 9.16.3, but for permutations instead of pairings. In order to derive it, we need first to explicit a bit how an element of $\mathfrak{S}(p)$ can be put in one-to-one correspondence with an element of $\mathfrak{P}^{(2)}(2p)$ whose pairs are all composed of one even integer and one odd integer.

To a full cycle $c = (i_l \ \ldots \ i_1)$ on $\{1, \ldots, l\}$ we associate the pairing $\lambda_c = (2i_1\, 2i_2 - 1) \cdots (2i_l\, 2i_1 - 1)$ on $\{1, \ldots, 2l\}$. The reverse operation is obtained by collapsing the two elements $2i$ and $2i - 1$ to a single element i for each $1 \leqslant i \leqslant p$. Then as expected, we associate to a general permutation $\alpha = c_1 \cdots c_m \in \mathfrak{S}(p)$ the pairing $\lambda_\alpha = \lambda_{c_1} \cdots \lambda_{c_m} \in \mathfrak{P}^{(2)}(2p)$.

Observe that, denoting by γ the canonical full cycle either on $\{1, \ldots, p\}$ or on $\{1, \ldots, 2p\}$, we have

$$\forall\ \alpha \in \mathfrak{S}(p), \ \sharp(\alpha) + \sharp(\gamma^{-1}\alpha) = \sharp(\gamma^{-1}\lambda_\alpha). \tag{9.34}$$

Indeed, the cycles of $\gamma^{-1}\lambda_\alpha$ are precisely cycles of the form $(2i_i \ \ldots \ 2i_l)$ for $(i_l \ \ldots \ i_1)$ a cycle of α (supported on even integers) and of the form $(2i_{l'} - 1 \ \ldots \ 2i_1 - 1)$ for $(i_{l'} \ \ldots \ i_1)$ a cycle of $\gamma^{-1}\alpha$ (supported on odd integers). So what equation (9.34) shows is that the elements of $\mathfrak{S}(p)$ having a given geodesic defect are in bijection with the elements of $\mathfrak{P}^{(2)}(2p)$ with even-odd pairs only and having the same geodesic defect (between id and

γ in both cases). In particular, we recover the well-known bijection between $NC(p)$ and $NC^{(2)}(2p)$ (because a non-crossing pairing is necessarily composed of even-odd pairs only).

Next, for any function g, either from $[p]$ to $[k]$ or from $[2p]$ to $[k]$, we will denote by γ_g the permutation, either on $\{1,\dots,p\}$ or on $\{1,\dots,2p\}$, which is the product of the canonical full cycles on the level sets of g. For any function $f:[p]\to[k]$, we define the function $\widetilde{f}:[2p]\to[k]$ by $\widetilde{f}(2i)=\widetilde{f}(2i-1)=f(i)$ for each $1\leqslant i\leqslant p$. It is then easy to see that we have more generally

$$\forall\ f:[p]\to[k],\ \forall\ \alpha\in\mathfrak{S}(p),\ \sharp(\alpha)+\sharp(\gamma_f^{-1}\alpha)=\sharp(\gamma_{\widetilde{f}}^{-1}\lambda_\alpha).$$

This simple observation will allow us to derive, as a slight adaptation of Lemma 9.16.3, a corresponding estimate for permutations instead of pairings.

Lemma 9.16.4. *Let $p\in\mathbf{N}$. For any $f:[p]\to[k]$, any $0\leqslant\delta\leqslant\lfloor(p+|\operatorname{Im}(f)|)/2\rfloor$, and any $1\leqslant m\leqslant p-2\delta$, define the set of permutations which are composed of m disjoint cycles and which have a defect 2δ of being on the geodesics between* id *and γ_f (the product of the canonical full cycles on each of the $|\operatorname{Im}(f)|$ level sets of f) as*

$$\mathfrak{S}_{f,\delta,m}(p)=\{\alpha\in\mathfrak{S}(p)\ :\ \sharp(\alpha)=m\ \text{and}\ \sharp(\gamma_f^{-1}\alpha)+\sharp(\alpha)=p+|\operatorname{Im}(f)|-2\delta\}.$$

Then, for any $0\leqslant\delta\leqslant\lfloor p/2\rfloor$ and any $1\leqslant m\leqslant p-2\delta$, we have the upper bound

$$\left|\left\{(f,\alpha)\ :\ \alpha\in\mathfrak{S}_{f,\delta,m}(p)\right\}\right|\leqslant\left(4k^4p^4\right)^\delta\sum_{\epsilon=0}^{2\delta}k^{m-\epsilon}\mathrm{Nar}_p^{m-\epsilon}.$$

Proof. We just observed that, for any $0\leqslant\delta\leqslant\lfloor p/2\rfloor$ and $1\leqslant m\leqslant p-2\delta$, the following equivalence holds

$$\alpha\in\mathfrak{S}_{f,\delta,m}(p)\ \Leftrightarrow\ \sharp(\alpha)=m\ \text{and}\ \lambda_\alpha\in\mathfrak{P}^{(2)}_{\widetilde{f},\delta},\tag{9.35}$$

where $\mathfrak{P}^{(2)}_{\widetilde{f},\delta}$ denotes the set of pairings having a defect 2δ of lying on the geodesics between id and $\gamma_{\widetilde{f}}$, as defined in Lemma 9.16.3.

In particular, $\alpha\in\mathfrak{S}_{f,0,m}(p)\ \Leftrightarrow\ \sharp(\alpha)=m$ and $\lambda_\alpha\in\mathfrak{P}^{(2)}_{\widetilde{f},0}$, and we know that there are precisely $k^m\mathrm{Nar}_p^m$ possibilities to build a pair (f,α) satisfying this condition (because the latter holds if and only if the three constraints $\sharp(\alpha)=m$, $\alpha\in NC(p)$ and $f\circ\alpha=f$ are fulfilled).

For the case $1\leqslant\delta\leqslant\lfloor p/2\rfloor$, we will mimic the proof of Lemma 9.16.3. So let (f,α) be such that $\alpha\in\mathfrak{S}_{f,\delta,m}(p)$ and assume without loss of generality that the transpositions $(i_1\,j_1),\dots,(i_p\,j_p)$ in λ_α are ordered so that λ_α is under the canonical form (9.32) for $\gamma_{\widetilde{f}}$. This means that the partial pairing $(i_{2\delta+1}\,j_{2\delta+1})\cdots(i_p\,j_p)$ is on the geodesics between id and some $\overline{\varsigma}$ with $|\operatorname{Im}(f)|$ cycles, and the number of such partial pairings is the same as the number of partial pairings being on the geodesics between id and some $\gamma_{\overline{f}}$ with $|\operatorname{Im}(\overline{f})|=|\operatorname{Im}(f)|$. Hence, to count how many ways there are of constructing what happens on $\{i_1,j_1,\dots,i_{2\delta},j_{2\delta}\}$, we have the trivial upper bound that would arise if picking the 2δ first transpositions in λ_α, as well as the values of $\overline{f}$ on them, completely arbitrarily. This yields a number of possibilities of at most $k^{4\delta}\binom{2p}{2}\cdots\binom{2(p-\delta+1)}{2}$. While on $\{i_{2\delta+1},j_{2\delta+1},\dots,i_{2p},j_{2p}\}$, we have to impose that the $p-2\delta$ last transpositions in λ_α are non-crossing, and that, when collapsed into a permutation of $p-2\delta$ elements, the latter has between $m-2\delta$ and m cycles and the function $\overline{f}$ takes only one value on each of them. This leaves us with a number of possibilities of at most $\sum_{\epsilon=0}^{2\delta}k^{m-\epsilon}\mathrm{Nar}_{p-2\delta}^{m-\epsilon}$. Putting everything together, we see that the number of pairs (f,α) satisfying condition (9.35) is less than

$$k^{4\delta}\binom{2p}{2}\cdots\binom{2(p-\delta+1)}{2}\sum_{\epsilon=0}^{2\delta}k^{m-\epsilon}\mathrm{Nar}_{p-2\delta}^{m-\epsilon}\leqslant k^{4\delta}\left(2p^2\right)^{2\delta}\sum_{\epsilon=0}^{2\delta}k^{m-\epsilon}\mathrm{Nar}_p^{m-\epsilon},$$

which is exactly what we wanted to show. □

Remark 9.16.5. *The upper bound we established in Lemma 9.16.4 is probably far from optimal (e.g. it is likely that the exponent 4δ in the polynomial pre-factor in k and p can be improved). But this does not really matter for our specific goal. Nonetheless, in the special case of non-geodesic permutations between* id *and γ on $\{1,\dots,p\}$, it is in fact quite easy to obtain an upper bound which scales as $p^{3\delta}$ for the ratio between the number of 2δ non-geodesic permutations with a given number of cycles and the number of geodesic permutations with the same number of cycles. We present the result in Lemma 9.16.6 below, the problem being that the proof method*

does not seem to generalize so straightforwardly to the case that we truly need, that is the one of non-geodesic permutations between id *and* γ_f.

Note also that very similar looking upper bounds had previously been derived regarding the cardinality of the set $\mathfrak{S}_\delta(p) = \{\alpha \in \mathfrak{S}(p) \ : \ \sharp(\gamma^{-1}\alpha)+\sharp(\alpha) = p+1-2\delta\}$, *which is the union of the sets* $\mathfrak{S}_{\delta,m}(p)$ *defined in Lemma 9.16.6, for* $1 \leqslant m \leqslant p$. *In particular, it was established in [141], Lemma 12, that for any* $0 \leqslant \delta \leqslant \lfloor p/2 \rfloor$,

$$\left|\mathfrak{S}_\delta(p)\right| \leqslant \left|\mathfrak{S}_0(p)\right| p^{3\delta} = \mathrm{Cat}_p\, p^{3\delta}.$$

However, this is definitely even less enough for our purpose: the latter really requires an upper bound on the number of permutations which have a given defect and a given number of cycles in terms of the number of permutations which have no defect and the same (or a related) number of cycles.

On the other hand, one may have hoped for a stronger result than these simply counting ones. For instance something like

$$d(\mathrm{id},\alpha)+d(\alpha,\gamma) = d(\mathrm{id},\gamma)+2\delta \ \Rightarrow \ \exists\ \alpha' \ : \ d(\alpha,\alpha') = 2\delta' \text{ and } d(\mathrm{id},\alpha')+d(\alpha',\gamma) = d(\mathrm{id},\gamma),$$

with $\delta' \leqslant \theta\delta$ *and with the mapping* $\phi : \alpha \mapsto \alpha'$ *satisfying* $\left|\phi^{-1}(\alpha')\right| \leqslant p^{\kappa\delta}$, *for some coefficients* θ, κ. *However, determining whether this kind of statement holds or not seems to remain an open question.*

Lemma 9.16.6. *Let* $p \in \mathbf{N}$ *and denote by* γ *the canonical full cycle on* $\{1,\ldots,p\}$. *For any* $0 \leqslant \delta \leqslant \lfloor p/2 \rfloor$ *and* $1 \leqslant m \leqslant p-2\delta$, *define the set of permutations which are composed of* m *disjoint cycles and which are* 2δ*-away from the geodesics between* id *and* γ *as*

$$\mathfrak{S}_{\delta,m}(p) = \{\alpha \in \mathfrak{S}(p) \ : \ \sharp(\alpha) = m \text{ and } \sharp(\gamma^{-1}\alpha)+\sharp(\alpha) = p+1-2\delta\}.$$

Then, the cardinality of $\mathfrak{S}_{\delta,m}(p)$ *is upper bounded in terms of the cardinality of* $\mathfrak{S}_{0,m}(p)$ *as*

$$\left|\mathfrak{S}_{\delta,m}(p)\right| \leqslant \left|\mathfrak{S}_{0,m}(p)\right| \left(\frac{p^3}{2}\right)^\delta.$$

Proof. Let $p \in \mathbf{N}$ and $1 \leqslant m \leqslant p$. We know from [85], Theorems 4.1 and 4.2, that there exist polynomials P_q of degree q, for $0 \leqslant q \leqslant \lfloor p/2 \rfloor$, such that for any $0 \leqslant \delta \leqslant \lfloor (p-m)/2 \rfloor$,

$$\left|\mathfrak{S}_{\delta,m}(p)\right| = \frac{p!}{2^{2\delta}(2\delta)!}\binom{p+1-2\delta}{m}\sum_{\epsilon=0}^{\delta}\binom{p-1}{m-1+2\epsilon}P_\epsilon(m)P_{\delta-\epsilon}(p+1-m-2\delta). \tag{9.36}$$

What is more, one can check from the explicit expression provided there for the polynomials P_q that, for any $x \geqslant 0$, $P_q(x) \leqslant (2x)^q$. As a particular instance of equation (9.36), we have

$$\left|\mathfrak{S}_{0,m}(p)\right| = p!\binom{p+1}{m}\binom{p-1}{m-1}.$$

And as a consequence, we get by a brutal upper bounding that, for any $1 \leqslant \delta \leqslant \lfloor (p-m)/2 \rfloor$,

$$\begin{aligned}
\frac{\left|\mathfrak{S}_{\delta,m}(p)\right|}{\left|\mathfrak{S}_{0,m}(p)\right|} &= \frac{1}{2^{2\delta}(2\delta)!}\prod_{i=0}^{m-1}\frac{p+1-2\delta-i}{p+1-i}\sum_{\epsilon=0}^{\delta}\prod_{j=0}^{2\epsilon-1}\frac{p-m-j}{m+j}P_\epsilon(m)P_{\delta-\epsilon}(p+1-m-2\delta) \\
&\leqslant \frac{1}{2^{2\delta}(2\delta)!}\sum_{\epsilon=0}^{\delta}\left(\frac{p-m}{m}\right)^{2\epsilon} 2^\epsilon\, m^\epsilon\, 2^{\delta-\epsilon}\,(p+1-m-2\delta)^{\delta-\epsilon} \\
&\leqslant \frac{1}{2^{2\delta}(2\delta)!} \times (\delta+1)\,2^\delta\, p^{3\delta} \\
&\leqslant \left(\frac{p^3}{2}\right)^\delta.
\end{aligned}$$

And this is precisely the claimed upper bound. □

Remark 9.16.7. *There is a close link between the problem we are concerned with and the one of finding tractable expressions for the so-called connection coefficients of the symmetric group (the reader is referred e.g. to [84] for more on that topic). Closed formulas are actually known for the connection coefficients of* $\mathfrak{S}(p)$, *involving the characters of its irreducible representations. But unfortunately, they are not really handleable in there full*

generality. And it seems it is only in some specific cases that more manageable forms can been obtained (i.e. in the first place as a sum of positive terms, so that one can see more easily what its order of magnitude is). The two situations which are well-understood are, on the one hand when the function f is constant (which corresponds to the case where γ_f is the canonical full cycle, and hence has a particularly simple cycle type, that is treated e.g. in [85]), and on the other hand when the defect 2δ is 0 (which corresponds to the case of so-called top connection coefficients).

9.17 Appendix H: Extra remarks on the convergence of the studied random matrix ensembles

For any Hermitian M on $\mathbf{C}^n$, we shall denote by $\lambda_1(M), \ldots, \lambda_n(M) \in \mathbf{R}$ its eigenvalues, and by N_M its eigenvalue distribution, i.e. the probability measure on $\mathbf{R}$ defined by

$$N_M = \frac{1}{n} \sum_{i=1}^{n} \delta_{\lambda_i(M)}.$$

In words, for any $I \subset \mathbf{R}$, $N_M(I)$ is the proportion of eigenvalues of M which belong to I.

9.17.1 "Modified" Wishart ensemble

Fix $k \in \mathbf{N}$ and $c > 0$. Then, for each $d \in \mathbf{N}$, let $W \sim \mathscr{W}_{d^2, cd^2}$ and define the random positive semidefinite matrix W_d on $(\mathbf{C}^d)^{\otimes k+1}$ by

$$W_d = \frac{1}{d^2} \sum_{j=1}^{k} \widetilde{W}(j). \tag{9.37}$$

Proposition 9.6.2 establishes that when $d \to +\infty$, the eigenvalue distribution of W_d converges in moments towards a Marčenko-Pastur distribution of parameter ck. But a stronger result actually holds, namely that there is convergence in probability of N_{W_d} towards $\mu_{MP(ck)}$. What is meant is made precise in Theorem 9.17.1 below.

Theorem 9.17.1. *For any $I \in \mathbf{R}$ and any $\varepsilon > 0$,*

$$\lim_{d \to +\infty} \mathbf{P}_{W \sim \mathscr{W}_{d^2, cd^2}} \left(\left| N_{W_d}(I) - \mu_{MP(ck)}(I) \right| > \varepsilon \right) = 0,$$

where the matrix W_d is as defined in equation (9.37).

Theorem 9.17.1 is a direct consequence of the estimate on the p-order moments $\mathbf{E} \operatorname{Tr} \left[\left(\sum_{j=1}^{k} \widetilde{W}_{\mathrm{AB}^k}(j) \right)^p \right]$ from Proposition 9.6.2, combined with the estimate on the p-order variances $\mathbf{Var} \operatorname{Tr} \left[\left(\sum_{j=1}^{k} \widetilde{W}_{\mathrm{AB}^k}(j) \right)^p \right]$ from Proposition 9.17.2 below. The proof, which follows a quite standard procedure, may be found detailed in [4] and sketched in [10].

Proposition 9.17.2. *Let $p \in \mathbf{N}$. For any constant $c > 0$,*

$$\mathbf{Var}_{W_{\mathrm{AB}} \sim \mathscr{W}_{d^2, cd^2}} \operatorname{Tr} \left[\left(\sum_{j=1}^{k} \widetilde{W}_{\mathrm{AB}^k}(j) \right)^p \right] \underset{d \to +\infty}{=} o\left(d^{2p+k+1} \right).$$

Proof. Let $p \in \mathbf{N}$. We already know that $\left(\mathbf{E} \operatorname{Tr} \left[\left(\sum_{j=1}^{k} \widetilde{W}_{\mathrm{AB}^k}(j) \right)^p \right] \right)^2 \sim_{d \to +\infty} \left(\mathrm{M}^{(p)}_{MP(ck)} d^{2p+k+1} \right)^2$ thanks to Proposition 9.6.2. Consequently, the only thing that remains to be shown in order to establish Proposition 9.17.2 is that we also have $\mathbf{E} \left(\operatorname{Tr} \left[\left(\sum_{j=1}^{k} \widetilde{W}_{\mathrm{AB}^k}(j) \right)^p \right] \right)^2 \sim_{d \to +\infty} \left(\mathrm{M}^{(p)}_{MP(ck)} d^{2p+k+1} \right)^2$. The combinatorics involved in the proof of the latter estimate is very similar to the one already appearing in the proof of the former. We will therefore skip some of the details here.

To begin with, let us fix a few additional notation. We define $\gamma_1 = (p \ \ldots \ 1)$ and $\gamma_2 = (2p \ \ldots \ p+1)$ as the canonical full cycles on $\{1, \ldots, p\}$ and $\{p+1, \ldots, 2p\}$ respectively. Also, for each functions $f_1 : \{1, \ldots, p\} \to [k]$,

$f_2 : \{p+1, \ldots, 2p\} \to [k]$, we define the function $f_{1,2} : [2p] \to [k]$ by $f_{1,2} = f_1$ on $\{1, \ldots, p\}$ and $f_{1,2} = f_2$ on $\{p+1, \ldots, 2p\}$. Then, by a slight generalization of Proposition 9.12.6 we have that, for any $\alpha \in \mathfrak{S}(2p)$,

$$\sharp((\hat{\gamma}_1\hat{\gamma}_2)^{-1}\hat{\alpha}_{f_{1,2}}) = \sharp((\gamma_{1\,f_1}\gamma_{2\,f_2})^{-1}\alpha) + 2k - |\operatorname{Im}(f_1)| - |\operatorname{Im}(f_2)|.$$

We can thus derive from the graphical calculus for Wishart matrices (in complete analogy to the way formula (9.16) was obtained) that, for any $d \in \mathbf{N}$,

$$\mathbf{E}_{W_{\mathrm{AB}} \sim \mathscr{W}_{d^2,cd^2}} \left(\operatorname{Tr} \left[\left(\sum_{j=1}^{k} \widetilde{W}_{\mathrm{AB}^k}(j) \right)^p \right] \right)^2 = \sum_{\substack{f_1:\{1,\ldots,p\}\to[k] \\ f_2:\{p+1,\ldots,2p\}\to[k]}} \sum_{\alpha \in \mathfrak{S}(2p)} c^{\sharp(\alpha)} d^{n(\alpha,f_1,f_2)}, \tag{9.38}$$

where for each $\alpha \in \mathfrak{S}(2p)$ and $f_1 : \{1, \ldots, p\} \to [k]$, $f_2 : \{p+1, \ldots, 2p\} \to [k]$,

$$n(\alpha, f_1, f_2) = 2\sharp(\alpha) + \sharp((\gamma_1\gamma_2)^{-1}\alpha) + \sharp((\gamma_{1\,f_1}\gamma_{2\,f_2})^{-1}\alpha) + 2k - |\operatorname{Im}(f_1)| - |\operatorname{Im}(f_2)|.$$

Yet, by Lemma 9.10.1 and equation (9.28) in Lemma 9.10.5, we get: First, for any $\alpha \in \mathfrak{S}(2p)$,

$$\sharp(\alpha) + \sharp((\gamma_1\gamma_2)^{-1}\alpha) = 4p - \left(|\alpha| + |(\gamma_1\gamma_2)^{-1}\alpha|\right) \leqslant 4p - |\gamma_1\gamma_2| = 2p + \sharp(\gamma_1\gamma_2) = 2p + 2, \tag{9.39}$$

with equality if and only if $\alpha = \alpha_1\alpha_2$ where $\alpha_1 \in NC(\{1, \ldots, p\})$, $\alpha_2 \in NC(\{p+1, \ldots, 2p\})$. And second, for any $\alpha \in \mathfrak{S}(2p)$ and any $f_1 : \{1, \ldots, p\} \to [k]$, $f_2 : \{p+1, \ldots, 2p\} \to [k]$,

$$\sharp(\alpha) + \sharp((\gamma_{1\,f_1}\gamma_{2\,f_2})^{-1}\alpha) \leqslant 2p + \sharp(\gamma_{1\,f_1}\gamma_{2\,f_2}) = 2p + |\operatorname{Im}(f_1)| + |\operatorname{Im}(f_2)|, \tag{9.40}$$

with equality if and only if $\alpha = \alpha_1\alpha_2$ where $\alpha_1 \in NC(\{1, \ldots, p\})$ and $f_1 \circ \alpha_1 = f_1$, $\alpha_2 \in NC(\{p+1, \ldots, 2p\})$ and $f_2 \circ \alpha_2 = f_2$. So putting equations (9.39) and (9.40) together, we get in the end that for any $\alpha \in \mathfrak{S}(2p)$ and $f_1 : \{1, \ldots, p\} \to [k]$, $f_2 : \{p+1, \ldots, 2p\} \to [k]$,

$$n(\alpha, f_1, f_2) \leqslant 4p + 2k + 2,$$

with equality if and only if $\alpha = \alpha_1\alpha_2$ where $\alpha_1 \in NC(\{1, \ldots, p\})$ and $f_1 \circ \alpha_1 = f_1$, $\alpha_2 \in NC(\{p+1, \ldots, 2p\})$ and $f_2 \circ \alpha_2 = f_2$.

We thus get that, asymptotically, the dominant term in formula (9.38) factorizes as

$$\mathbf{E} \left(\operatorname{Tr} \left[\left(\sum_{j=1}^{k} \widetilde{W}_{\mathrm{AB}^k}(j) \right)^p \right] \right)^2 \underset{d\to+\infty}{\sim} d^{4p+2k+2} \sum_{\substack{\alpha_1 \in NC(\{1,\ldots,p\}) \\ \alpha_2 \in NC(\{p+1,\ldots,2p\})}} \sum_{\substack{f_1:\{1,\ldots,p\}\to[k],\, f_1\circ\alpha_1=f_1 \\ f_2:\{p+1,\ldots,2p\}\to[k],\, f_2\circ\alpha_2=f_2}} c^{\sharp(\alpha_1\alpha_2)}$$

$$\underset{d\to+\infty}{\sim} \left(d^{2p+k+1} \sum_{\alpha \in NC(p)} \sum_{\substack{f:[p]\to[k] \\ f\circ\alpha=f}} c^{\sharp(\alpha)} \right)^2,$$

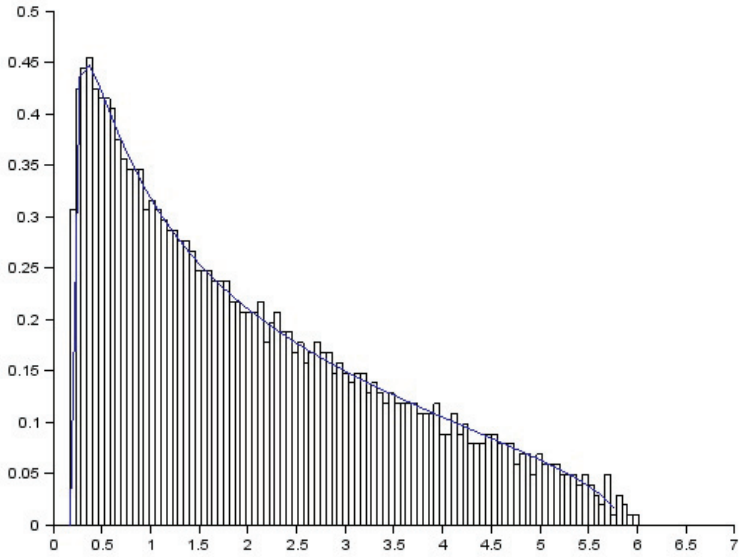
where the last equality is simply because $\sharp(\alpha_1\alpha_2) = \sharp(\alpha_1) + \sharp(\alpha_2)$. And hence,

$$\mathbf{E}_{W_{\mathrm{AB}} \sim \mathscr{W}_{d^2,cd^2}} \left(\operatorname{Tr} \left[\left(\sum_{j=1}^{k} \widetilde{W}_{\mathrm{AB}^k}(j) \right)^p \right] \right)^2 \underset{d\to+\infty}{\sim} \left(d^{2p+k+1} \sum_{\alpha \in NC(p)} (ck)^{\sharp(\alpha)} \right)^2 = \left(d^{2p+k+1} \mathrm{M}^{(p)}_{MP(ck)} \right)^2,$$

which is exactly what we needed to conclude the proof. □

Let us illustrate the result stated in Theorem 9.17.1 in the simplest case of 2-extendibility and uniformly distributed mixed states. In Figure 9.8, the spectral distribution of $W_d = \left(W_{\mathrm{AB}_1} \otimes \mathrm{Id}_{\mathrm{B}_2} + W_{\mathrm{AB}_2} \otimes \mathrm{Id}_{\mathrm{B}_1}\right)/d^2$, for $W_{\mathrm{AB}} \sim \mathscr{W}_{d^2,d^2}$, and a Marčenko-Pastur distribution of parameter 2 are plotted together. The empirical eigenvalue histogram is done in dimension $d = 12$, from 100 repetitions.

Figure 9.8: Spectral distribution of $(W_{\mathrm{AB}_1} \otimes \mathrm{Id}_{\mathrm{B}_2} + W_{\mathrm{AB}_2} \otimes \mathrm{Id}_{\mathrm{B}_1})/d^2$, for $W_{\mathrm{AB}} \sim \mathscr{W}_{d^2,d^2}$ vs Marčenko-Pastur distribution of parameter 2.



9.17.2 "Modified" GUE ensemble

Fix $k \in \mathbf{N}$. Then, for each $d \in \mathbf{N}$, let $G \sim GUE(d^2)$ and define the random Hermitian matrix G_d on $(\mathbf{C}^d)^{\otimes k+1}$ by

$$G_d = \frac{1}{d} \sum_{j=1}^{k} \widetilde{G}(j). \tag{9.41}$$

In complete analogy to what was explained in the case of Wishart matrices, Proposition 9.2.3 establishes that when $d \to +\infty$, the eigenvalue distribution of G_d converges in moments towards a centered semicircular distribution of parameter k. But here again, there is in fact convergence in probability of N_{G_d} towards $\mu_{SC(k)}$, which is made precise in Theorem 9.17.3 below.

Theorem 9.17.3. *For any $I \in \mathbf{R}$ and any $\varepsilon > 0$,*

$$\lim_{d \to +\infty} \mathbf{P}_{G \sim GUE(d^2)} \left(\left| N_{G_d}(I) - \mu_{SC(k)}(I) \right| > \varepsilon \right) = 0,$$

where the matrix G_d is as defined in equation (9.41).

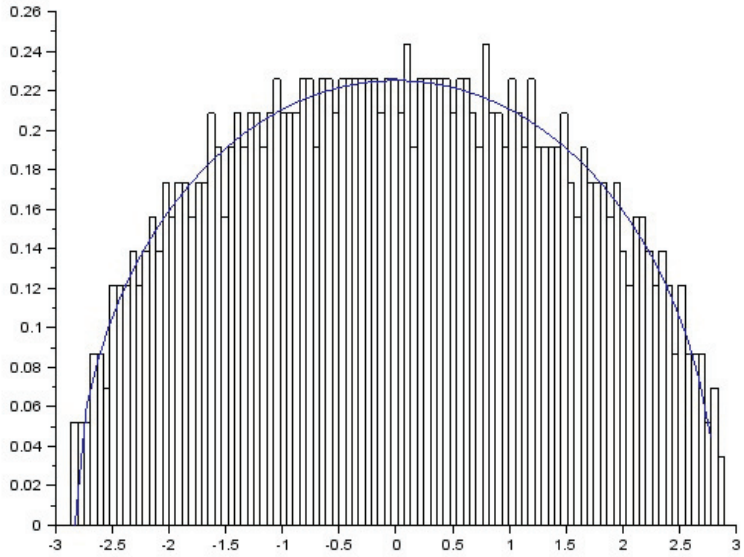
As already explained in the Wishart case, this follows directly from the moment's estimate in Proposition 9.2.3, together with the variance's estimate, for all $p \in \mathbf{N}$,

$$\mathbf{Var}_{G_{\mathrm{AB}} \sim GUE(d^2)} \mathrm{Tr} \left[\left(\sum_{j=1}^{k} \widetilde{G}_{\mathrm{AB}^k}(j) \right)^{2p} \right] \underset{d \to +\infty}{=} o\left(d^{2p+k+1} \right). \tag{9.42}$$

The proof follows the exact same lines as the one of Proposition 9.17.2 and is not repeated here.

Let us illustrate the result stated in Theorem 9.17.3 in the simplest case of 2-extendibility. In Figure 9.9, the spectral distribution of $G_d = (G_{\mathrm{AB}_1} \otimes \mathrm{Id}_{\mathrm{B}_2} + G_{\mathrm{AB}_2} \otimes \mathrm{Id}_{\mathrm{B}_1})/d^2$, for $G_{\mathrm{AB}} \sim GUE(d^2)$, and a centered semicircular distribution of parameter 2 are plotted together. The empirical eigenvalue histogram is done in dimension $d = 10$, from 100 repetitions.

Figure 9.9: Spectral distribution of $(G_{\mathrm{AB}_1} \otimes \mathrm{Id}_{\mathrm{B}_2} + G_{\mathrm{AB}_2} \otimes \mathrm{Id}_{\mathrm{B}_1})/d^2$, for $G_{\mathrm{AB}} \sim GUE(d^2)$ vs Centered semi-circular distribution of parameter 2.



Remark 9.17.4. *We may in fact say even more on the convergence of the random matrix sequences* $(W_d)_{d\in\mathbf{N}}$ *and* $(G_d)_{d\in\mathbf{N}}$ *defined by equations* (9.37) *and* (9.41) *respectively. Namely,*

$$N_{W_d} \overset{a.s.}{\underset{d\to+\infty}{\to}} \mu_{MP(ck)} \text{ and } N_{G_d} \overset{a.s.}{\underset{d\to+\infty}{\to}} \mu_{SC(k)}.$$

To establish this almost sure convergence result, the only thing that has to be verified is that, for any $p \in \mathbf{N}$*, the series of variances*

$$\sum_{d=1}^{+\infty} \mathbf{Var}\left[\frac{1}{d^{k+1}}\mathrm{Tr}\left(W_d^p\right)\right] \text{ and } \sum_{d=1}^{+\infty} \mathbf{Var}\left[\frac{1}{d^{k+1}}\mathrm{Tr}\left(G_d^{2p}\right)\right] \tag{9.43}$$

are summable. Indeed, almost sure convergence will then automatically follow from a standard application of the Chebyshev inequality and the Borel–Cantelli lemma. And condition (9.43) *actually holds, as a consequence of the fact that, for any* $p \in \mathbf{N}$*,*

$$\mathbf{Var}\left[\frac{1}{d^{k+1}}\mathrm{Tr}\left(W_d^p\right)\right] = O\left(d^{-2}\right) \text{ and } \mathbf{Var}\left[\frac{1}{d^{k+1}}\mathrm{Tr}\left(G_d^{2p}\right)\right] = O\left(d^{-2}\right).$$

9.17.3 Asymptotic freeness of certain Gaussian random matrices

Let us fix a few definitions and notation. Given $n \in \mathbf{N}$ and $[\Omega, P]$ a classical probability space, we define the free probability space $[\mathcal{M}_n(L^\infty[\Omega, P]), \varphi_n]$, where $\mathcal{M}_n(L^\infty[\Omega, P])$ is the set of $n \times n$ matrices with entries in $L^\infty[\Omega, P]$ and $\varphi_n(\cdot) = \mathbf{E}\,\mathrm{Tr}(\cdot)/n$ is the normalized trace function on $\mathcal{M}_n(L^\infty[\Omega, P])$. The two particular examples we shall focus on in the sequel are the ones we have already been extensively dealing with, namely GUE and Wishart matrices.

Lemma 9.17.5. *Given two finite-dimensional Hilbert spaces* $\mathrm{A} \equiv \mathbf{C}^{d_\mathrm{A}}$, $\mathrm{B} \equiv \mathbf{C}^{d_\mathrm{B}}$*, and* G *a random GUE matrix on* $\mathrm{A} \otimes \mathrm{B}$*, we define the following random matrices on* $\mathrm{A} \otimes \mathrm{B}_1 \otimes \mathrm{B}_2$*:*

$$\widetilde{G}_1 = \frac{1}{\sqrt{d_\mathrm{A} d_\mathrm{B}}} G_{\mathrm{AB}_1} \otimes \mathrm{Id}_{\mathrm{B}_2} \text{ and } \widetilde{G}_2 = \frac{1}{\sqrt{d_\mathrm{A} d_\mathrm{B}}} G_{\mathrm{AB}_2} \otimes \mathrm{Id}_{\mathrm{B}_1}.$$

Then, for any $p \in \mathbf{N}$ *and any function* $f : [2p] \to [2]$*,*

$$\lim_{d_\mathrm{A} \leqslant d_\mathrm{B} \to +\infty} \varphi_{d_\mathrm{A} d_\mathrm{B}^2}\left(\widetilde{G}_{f(1)} \ldots \widetilde{G}_{f(2p)}\right) = \left|\left\{\lambda \in NC^{(2)}(2p) \ : \ f \circ \lambda = f\right\}\right|. \tag{9.44}$$

Proof. We know from the proof of Proposition 2.3 (and using the same notation as those employed there) that

$$\varphi_{d_{\mathrm{A}}d_{\mathrm{B}}^2}\left(\widetilde{G}_{f(1)}\dots\widetilde{G}_{f(2p)}\right) = \sum_{\lambda\in\mathfrak{P}^{(2)}(2p)} d_{\mathrm{A}}^{\sharp(\gamma^{-1}\lambda)-p-1} d_B^{\sharp(\gamma_f^{-1}\lambda)-p-|\operatorname{Im}(f)|}.$$

Now, as explained there as well, for any $\lambda\in\mathfrak{P}^{(2)}(2p)$, on the one hand $\sharp(\gamma^{-1}\lambda)\leqslant p+1$ with equality iff $\lambda\in NC^{(2)}(2p)$, and on the other hand $\sharp(\gamma_f^{-1}\lambda)\leqslant p+|\operatorname{Im}(f)|$ with equality iff $\lambda\in NC^{(2)}(2p)$ and $f\circ\lambda=f$. The asymptotic estimate (9.44) therefore immediately follows. □

Theorem 9.17.6. *Let G_{AB} be a random GUE matrix on $\mathrm{A}\otimes\mathrm{B}$. Then, the random matrices $G_{\mathrm{AB}_1}\otimes\mathrm{Id}_{\mathrm{B}_2}$ and $G_{\mathrm{AB}_2}\otimes\mathrm{Id}_{\mathrm{B}_1}$ on $\mathrm{A}\otimes\mathrm{B}_1\otimes\mathrm{B}_2$ are asymptotically free.*

Proof. Theorem 9.17.6 is a direct consequence of Lemma 9.17.5 (see e.g. [144], proof of Proposition 22.22, for an entirely analogous argument). Indeed, as $d_{\mathrm{A}},d_{\mathrm{B}}\to+\infty$, the two empirical spectral distributions $\mu_{\widetilde{G}_1}$ and $\mu_{\widetilde{G}_2}$ both converge to the semicircular distribution with mean 0 and variance 1. And equation (9.44) is exactly the rule for computing mixed moments in two free such semicircular distributions (see e.g. [144], Lecture 12). □

Lemma 9.17.7. *Given two finite-dimensional Hilbert spaces $\mathrm{A}\equiv\mathbf{C}^{d_{\mathrm{A}}}$, $\mathrm{B}\equiv\mathbf{C}^{d_{\mathrm{B}}}$, and W a random Wishart matrix on $\mathrm{A}\otimes\mathrm{B}$ with parameter $cd_{\mathrm{A}}d_{\mathrm{B}}\in\mathbf{N}$, we define the following random matrices on $\mathrm{A}\otimes\mathrm{B}_1\otimes\mathrm{B}_2$:*

$$\widetilde{W}_1 = \frac{1}{d_{\mathrm{A}}d_{\mathrm{B}}}W_{\mathrm{AB}_1}\otimes\mathrm{Id}_{\mathrm{B}_2} \text{ and } \widetilde{W}_2 = \frac{1}{d_{\mathrm{A}}d_{\mathrm{B}}}W_{\mathrm{AB}_2}\otimes\mathrm{Id}_{\mathrm{B}_1}.$$

Then, for any $p\in\mathbf{N}$ and any function $f:[p]\to[2]$,

$$\lim_{d_{\mathrm{A}}\leqslant d_{\mathrm{B}}\to+\infty}\varphi_{d_{\mathrm{A}}d_{\mathrm{B}}^2}\left(\widetilde{W}_{f(1)}\dots\widetilde{W}_{f(p)}\right) = \sum_{\substack{\alpha\in NC(p)\\ f\circ\alpha=f}} c^{\sharp(\alpha)}. \tag{9.45}$$

Proof. We know from the proof of Proposition 6.2 (and using the same notation as those employed there) that

$$\varphi_{d_{\mathrm{A}}d_{\mathrm{B}}^2}\left(\widetilde{W}_{f(1)}\dots\widetilde{W}_{f(p)}\right) = \sum_{\alpha\in\mathfrak{S}(p)} c^{\sharp(\alpha)} d_{\mathrm{A}}^{\sharp(\alpha)+\sharp(\gamma^{-1}\alpha)-p-1} d_{\mathrm{B}}^{\sharp(\alpha)+\sharp(\gamma_f^{-1}\lambda)-p-|\operatorname{Im}(f)|}.$$

Now, as explained there as well, for any $\alpha\in\mathfrak{S}(p)$, on the one hand $\sharp(\alpha)+\sharp(\gamma^{-1}\alpha)\leqslant p+1$ with equality iff $\alpha\in NC(p)$, and on the other hand $\sharp(\alpha)+\sharp(\gamma_f^{-1}\alpha)\leqslant p+|\operatorname{Im}(f)|$ with equality iff $\alpha\in NC(p)$ and $f\circ\alpha=f$. The asymptotic estimate (9.45) therefore immediately follows. □

Theorem 9.17.8. *Let W_{AB} be a random Wishart matrix on $\mathrm{A}\otimes\mathrm{B}$ with parameter $cd_{\mathrm{A}}d_{\mathrm{B}}\in\mathbf{N}$. Then, the random matrices $W_{\mathrm{AB}_1}\otimes\mathrm{Id}_{\mathrm{B}_2}$ and $W_{\mathrm{AB}_2}\otimes\mathrm{Id}_{\mathrm{B}_1}$ on $\mathrm{A}\otimes\mathrm{B}_1\otimes\mathrm{B}_2$ are asymptotically free.*

Proof. Theorem 9.17.8 is a direct consequence of Lemma 9.17.7 (see e.g. [144], proof of Proposition 22.22, for an entirely analogous argument). Indeed, as $d_{\mathrm{A}},d_{\mathrm{B}}\to+\infty$, the two empirical spectral distributions $\mu_{\widetilde{W}_1}$ and $\mu_{\widetilde{W}_2}$ both converge to the Marčenko-Pastur distribution with parameter c. And equation (9.45) is exactly the rule for computing mixed moments in two free such Marčenko-Pastur distributions (see e.g. [144], Lectures 12 and 13). □

Part IV

Making use of permutation-symmetry to tackle multiplicativity issues

Chapter 9 in the previous part already made it quite clear that symmetries play an important role in quantum information theory. This part stands up much more ardently for such claim.

It is an ubiquitous issue in quantum information theory (but in fact in many other fields as well) to determine whether certain quantities have a multiplicative/additive behaviour. Indeed, a situation where perfect multiplicativity/additivity holds can usually be interpreted as follows: given a device that enables accomplishing a task with a certain performance, there is no way of combining two copies of this device which would allow performing better than when using them independently. In most cases though, this is not what happens: clever use of correlations does help. This assertion is actually even more true in the quantum than in the classical world: many quantities which are additive in classical information theory (and whose quantum analogues were therefore initially conjectured to be additive as well) have in fact been proved to fail additivity in quantum information theory. This is precisely what makes asymptotic performances much harder to quantify than single-copy ones. However, whenever the study of a multi-copy scenario can be reduced, in some manner or another, to that of an i.i.d. one, then the analysis becomes easy again. This is exactly the motivation behind de Finetti type statements, all of them having in common to make the most of the permutation-symmetry of the problem under consideration.

In Chapter 10 we first of all draw the outline of a very flexible de Finetti reduction. What we mean is that the latter has a broad scope of application (e.g. both quantum states and classical probability distributions) and that it allows keeping track of additional information that one may have on the considered object, apart from its permutation-symmetry. It is especially well-suited to the case where this extra knowledge consists of linear constraints that the object satisfies, one instance of which is extensively studied in Chapter 11. It may however still be useful in the case of convex constraints. In particular, it permits to derive the following connection: given any convex set of states $\mathcal{K}$, the exponential decay under tensoring of the maximum fidelity function $F(\cdot, \mathcal{K})$ implies that of the support function $h_{\mathcal{K}}(\cdot)$, and vice versa. We show-case this link on the example of the set of separable states. The latter is of prime importance because of its implications in a multitude of fields, most notably quantum computing and quantum Shannon theory. In this specific case, weak multiplicativity under tensoring does hold, but with a dependence on the ambient dimensions, and we do not know whether or not it could be removed. Finally, we examine how our techniques can be extended to the infinite-dimensional setting, in order to prove e.g. security of continuous variable quantum key distribution against general attacks, under assumptions as minimal as possible.

Chapter 11 entirely consists of one major application of the de Finetti reduction established in Chapter 10, namely to the study of the parallel repetition of multi-player non-local games. More precisely, we are interested in the following problem: if players sharing certain correlations are not able to win a given game with probability 1, does their probability of winning n instances of that game played in parallel decrease exponentially to 0 (with n), and if so at which rate? We solve this question in (almost) full generality in the case where the only limitation which is assumed on the physical power of the players is that they cannot signal information instantaneously from one another. For that, we introduce the notion of sub-no-signalling correlations between players. We then show that, if players sharing such correlations have a probability below $1 - \delta$ of winning one instance of a game, then they have an exponentially decaying probability of winning a fraction above $1 - \delta$ of n instances of this game played in parallel. This result translates into the analogous one for players sharing no-signalling correlations (a more commonly studied set of allowed strategies) in two cases: when the game only involves two players, or when the game is such that all potential queries to the players have a non-zero probability of being actually asked.

One chief question which remains open from that point is whether the flexible constrained de Finetti reduction of Chapter 10 could be used to study the parallel repetition problem, not only for no-signalling players as in Chapter 11, but also for quantum players (i.e. players whose strategy is dictated by the outcomes of local measurements which they perform on a shared entangled state). Let us enlarge a bit on this issue. The standard proof technique to tackle parallel repetition, in the quantum case but also, originally, in the classical and no-signalling cases, consists in iteratively assuming that the players have won a given instance of the game and then studying how this affects their winning probability in the others. Hence, if you can show that, conditioned on the event "the players have already won k instances of the game", the probability is high that they lose in at least 1, resp. most, of the $n - k$ remaining instances, you get exponential decay of the probability of winning all, resp. a fraction above the game value of, n instances of the game played in parallel. The main drawback of this approach is probably its "locality", which makes it not so straightforward and not so easily generalizable to more than 2 players as the more "global" de Finetti approach. That is why finding a way of attacking the quantum parallel repetition problem via de Finetti reductions would be desirable. One obstacle seems however

to be that there are certain steps from the standard route which remain unavoidable. One of them (which is actually crucial as well in Chapter 10, when studying exponential decay under tensoring of support functions of quantum state sets) is some kind of reconstruction step. Phrased informally: you have to be able to say that, if your strategy (or quantum state) almost satisfies the constraints defining your set of interest, then there must exist a strategy (or quantum state) which exactly satisfies them and which is not too far away from it. The hope that, more often than not, this quite natural expectancy will turn out to be true is at the heart of our whole constrained de Finetti reduction philosophy. Nonetheless, it is not always possible (or at least so easy) to get nice quantitative versions of this intuition.

Let us bring this discussion to a close by connecting it with more convex geometry orientated considerations. What we are always doing in this part is asking how a sequence of convex sets under examination $\{\mathcal{K}_n,\ n \in \mathbf{N}\}$ compares to the sequence of projective tensor power sets $\{\mathcal{K}_1^{\hat{\otimes} n},\ n \in \mathbf{N}\}$. Specifically, what we usually want to know is how much bigger than $\mathcal{K}_1^{\hat{\otimes} n}$ is our considered $\mathcal{K}_n$ in a given tensor power direction $u^{\otimes n}$. And in the cases where giving an answer valid for any such $u^{\otimes n}$ looks out of reach, attempting to still say what happens for most of them (in a way to be defined) could be something interesting to explore.

Part IV – Table of contents

Chapter 10 Flexible constrained de Finetti reductions and applications 161

10.1 Introduction 161

10.2 Flexible de Finetti reductions for finite-dimensional symmetric quantum systems 162

10.3 Exponential decay and concentration of h_S via de Finetti reduction approach 166

10.4 Exponential decay and concentration of h_S via entanglement measure approach 170

10.5 Equivalence between weak multiplicativity of support functions and of maximum fidelities . . 175

10.6 De Finetti reductions for infinite-dimensional symmetric quantum systems 178

10.7 Conclusion and outlook 180

Chapter 11 Parallel repetition and concentration for (sub-)no-signalling games 181

11.1 Non-local games and no-signalling strategies 181

11.2 Parallel repetition: definitions and main results 185

11.3 Constrained de Finetti reduction 186

11.4 Proofs of the main Theorems 188

11.5 Discussion 191

Chapter 10

Flexible constrained de Finetti reductions and applications

Based on "Flexible constrained de Finetti reductions and applications", in collaboration with A. Winter [128].

De Finetti theorems show how sufficiently exchangeable states are well-approximated by convex combinations of i.i.d. states. Recently, it was shown that in many quantum information applications a more relaxed *de Finetti reduction* (i.e. only a matrix inequality between the symmetric state and one of de Finetti form) is enough, and that it leads to more concise and elegant arguments.

Here we show several uses and general flexible applicability of a *constrained de Finetti reduction* in quantum information theory, which was recently discovered by Duan, Severini and Winter. In particular we show that the technique can accommodate other symmetries commuting with the permutation action, and permutation-invariant linear constraints. We then demonstrate that, in some cases, it is also fruitful with convex constraints, in particular separability in a bipartite setting. This is a constraint particularly interesting in the context of the complexity class QMA(2) of interactive quantum Merlin-Arthur games with unentangled provers, and our results relate to the soundness gap amplification of QMA(2) protocols by parallel repetition. It is also relevant for grasping the regularization of certain entropic channel parameters. Finally, we explore an extension to infinite-dimensional systems, which usually pose inherent problems to de Finetti techniques in the quantum case.

10.1 Introduction

The main motivation behind all de Finetti type theorems is to reduce the study of permutation-invariant scenarios to that of i.i.d. ones, which are often much easier to understand (see Chapter 3, Section 3.3, for a broader exposition). In many information theoretic situations, the problem is posed in such a way that one almost directly sees that the solution is (or is without loss of generality) permutation-invariant. Furthermore, in many scenarios one needs only to upper bound (and not to accurately approximate) a permutation-invariant object by i.i.d. ones. The seminal *de Finetti reduction* (aka *post-selection lemma*) of Christandl, König and Renner [49] was precisely designed for that: for any permutation-invariant state ρ on $\mathrm{H}^{\otimes n}$, with $d = |\mathrm{H}|$ the "local" Hilbert space dimension,

$$\rho \leqslant (n+1)^{d^2} \int_{\sigma \in \mathcal{D}(\mathrm{H})} \sigma^{\otimes n} \, \mathrm{d}\sigma, \tag{10.1}$$

where $\mathrm{d}\sigma$ denotes the uniform probability measure over the set of mixed states $\mathcal{D}(\mathrm{H})$ on H, and the inequality refers to the matrix ordering (for Hermitians A, B, $A \leqslant B$ means that $B - A$ is positive semidefinite). The beauty of this statement is that on the right hand side we have a universal object: one and the same convex combination provides the upper bound to all permutation-invariant states. At the same time, though, its very universality can be a drawback: every permutation-invariant state (quantum or classical) is upper bounded by the same convex combination of tensor power states, so that any other a priori information (apart from its permutation-invariance), that one may have on it, is lost. In [67], Appendix B, it was shown that at the sole cost of slightly increasing the polynomial pre-factor in front of the upper bounding de Finetti operator, it is actually possible to make it depend on the state of interest, or on some property that this state has, including in the

integral on the right hand side of equation (10.1) a fidelity term between ρ and the i.i.d. state $\sigma^{\otimes n}$. In [67], this *constrained de Finetti reduction* was applied to prove a coding theorem in a setting with adversarially chosen channel. In Chapter 11 another application to parallel repetition of no-signalling games is given.

In Section 10.2, we first review the constrained de Finetti reduction of [67] and give an alternative proof of it (Subsection 10.2.1). We then show that certain linear constraints lead to very simple and at the same time useful forms of the de Finetti reduction, such that certain "unwanted" contributions in the integral on the right hand side of equation (10.1) are either completely absent or exponentially suppressed (Subsection 10.2.2). Next, in Sections 10.3 and 10.4 we study in depth the case of separability, a convex constraint. In particular we show that there are several essentially equivalent ways of thinking about the exponential decay of the fidelity term. Inspired by separability, in Section 10.5 we present an axiomatic treatment of a wider class of convex constraints. Finally, in Section 10.6 we move to de Finetti reductions in the infinite-dimensional case.

10.2 Flexible de Finetti reductions for finite-dimensional symmetric quantum systems

10.2.1 A general constrained de Finetti reduction

Before getting into more specific statements, let us fix once and for all some notation that we shall use throughout the whole chapter (in addition to the general ones specified in Chapter 1, Section 1.3). Most of them, and many more related ones, are also introduced in Chapter 3, Section 3.1, to which the reader is referred for all the basics about the symmetric subspace. Given a Hilbert space H, we denote by $\mathrm{Sym}^n(\mathrm{H})$ the n-symmetric subspace of $\mathrm{H}^{\otimes n}$. In the case where $d = |\mathrm{H}| < +\infty$, $\dim\left(\mathrm{Sym}^n(\mathrm{H})\right) = \binom{n+d-1}{n}$ and the orthogonal projector onto $\mathrm{Sym}^n(\mathrm{H})$ may be written as

$$P_{\mathrm{Sym}^n(\mathrm{H})} = \binom{n+d-1}{n} \int_{|\psi\rangle \in S_{\mathrm{H}}} |\psi\rangle\langle\psi|^{\otimes n} \mathrm{d}\psi,$$

where $\mathrm{d}\psi$ stands for the uniform probability measure on the unit sphere S_{H} of H. A state ρ on $\mathrm{H}^{\otimes n}$ is then called permutation-invariant (or simply symmetric) if for any permutation $\pi \in \mathfrak{S}(n)$, it is invariant under the action of the associated permutation unitary $U(\pi) \in \mathcal{U}(\mathrm{H})$, i.e. $U(\pi)\rho U(\pi)^\dagger = \rho$. This can be expressed equivalently by saying that there exists a unit vector $|\psi\rangle \in \mathrm{Sym}^n(\mathrm{H} \otimes \mathrm{H}')$, where $|\mathrm{H}'| = d$, such that $\rho = \mathrm{Tr}_{\mathrm{H}'^{\otimes n}} |\psi\rangle\langle\psi|$.

Going from rigid to more flexible de Finetti reductions relies essentially on the so-called "pinching trick", which we state formally as Lemma 10.2.1 below. This is a generalization of results appearing in [97] and [98].

Lemma 10.2.1. *Let* H *be a Hilbert space and* $M_1, \ldots, M_r$ *be operators on* H. *Then, for any state* ρ *on* H,

$$\sum_{i,j=1}^{r} M_i \rho M_j^\dagger \leqslant r \sum_{i=1}^{r} M_i \rho M_i^\dagger.$$

Proof. To prove that Lemma 10.2.1 holds for any state on H, it is sufficient to prove that it holds for any pure state on H. Let therefore $|\psi\rangle$ be a unit vector in H. Then, for any unit vector $|\varphi\rangle$ in H, we have by the Cauchy-Schwarz inequality

$$\begin{aligned}
\langle\varphi| \left(\sum_{i,j=1}^{r} M_i |\psi\rangle\langle\psi| M_j^\dagger \right) |\varphi\rangle &= \left| \sum_{i=1}^{r} \langle\varphi|M_i|\psi\rangle \right|^2 \\
&\leqslant r \sum_{i=1}^{r} |\langle\varphi|M_i|\psi\rangle|^2 \\
&= \langle\varphi| \left(r \sum_{i=1}^{r} M_i |\psi\rangle\langle\psi| M_i^\dagger \right) |\varphi\rangle,
\end{aligned}$$

which concludes the proof. □

With this tool at hand, we are ready to get, first of all, the pure state version of the flexible de Finetti reduction.

Proposition 10.2.2. *Any unit vector* $|\theta\rangle \in \mathrm{Sym}^n(\mathrm{H})$ *satisfies*

$$|\theta\rangle\langle\theta| \leqslant \binom{n+d-1}{n}^3 \int_{|\psi\rangle\in S_{\mathrm{H}}} \left|\langle\theta|\psi^{\otimes n}\rangle\right|^2 |\psi\rangle\langle\psi|^{\otimes n}\mathrm{d}\psi.$$

Proof. Let $|\theta\rangle \in \mathrm{Sym}^n(\mathrm{H})$ be a unit vector. Then,

$$|\theta\rangle\langle\theta| = P_{\mathrm{Sym}^n(\mathrm{H})}|\theta\rangle\langle\theta|P^\dagger_{\mathrm{Sym}^n(\mathrm{H})} = \binom{n+d-1}{n}^2 \int_{|\psi\rangle,|\varphi\rangle\in S_{\mathrm{H}}} |\psi\rangle\langle\psi|^{\otimes n}|\theta\rangle\langle\theta||\varphi\rangle\langle\varphi|^{\otimes n}\,\mathrm{d}\psi\,\mathrm{d}\varphi.$$

Now observe, setting $r = \binom{n+d-1}{n}^2$, that the span of $\{|\psi\rangle\langle\psi|^{\otimes n},\ |\psi\rangle \in S_{\mathrm{H}}\}$, subject to the condition of having trace 1, has dimension $r-1$. So by Caratheodory's theorem, we know that there exist $\{p_1,\ldots,p_r\}$, a convex combination, and $\{\psi_1,\ldots,\psi_r\}$, a set of unit vectors in H, such that

$$\int_{|\psi\rangle\in S_{\mathrm{H}}} |\psi\rangle\langle\psi|^{\otimes n}\mathrm{d}\psi = \sum_{i=1}^r p_i|\psi_i\rangle\langle\psi_i|^{\otimes n}. \tag{10.2}$$

We can therefore rewrite

$$\begin{aligned}
|\theta\rangle\langle\theta| &= r\sum_{i,j=1}^r p_ip_j|\psi_i\rangle\langle\psi_i|^{\otimes n}|\theta\rangle\langle\theta||\psi_j\rangle\langle\psi_j|^{\otimes n} \\
&\leqslant r^2\sum_{i=1}^r p_i^2\left|\langle\theta|\psi_i^{\otimes n}\rangle\right|^2|\psi_i\rangle\langle\psi_i|^{\otimes n} \\
&\leqslant r^{3/2}\sum_{i=1}^r p_i\left|\langle\theta|\psi_i^{\otimes n}\rangle\right|^2|\psi_i\rangle\langle\psi_i|^{\otimes n},
\end{aligned}$$

where the next to last inequality is by Lemma 10.2.1, and the last inequality is because, for each $1 \leqslant i \leqslant r$, $p_i \leqslant 1/\sqrt{r}$ (which can be seen by contracting both sides of equation (10.2) with $\langle\psi_i^{\otimes n}|\cdot|\psi_i^{\otimes n}\rangle$). And consequently, since this holds for any ensemble $\{p_i,\,\psi_i\}_{1\leqslant i\leqslant r}$ satisfying equation (10.2), we have by convex combination

$$|\theta\rangle\langle\theta| \leqslant r^{3/2}\int_{|\psi\rangle\in S_{\mathrm{H}}} \left|\langle\theta|\psi^{\otimes n}\rangle\right|^2|\psi\rangle\langle\psi|^{\otimes n}\,\mathrm{d}\psi,$$

which is precisely the advertised result. □

From Proposition 10.2.2, we can now easily derive the general mixed state version of our flexible de Finetti reduction, which was also obtained in [67] by a slightly different route. In Theorem 10.2.3 below, as well as in the remainder of this chapter, $F(\rho,\sigma)$ stands for the fidelity between quantum states (or classical probability distributions) ρ and σ, defined as $F(\rho,\sigma) = \|\sqrt{\rho}\sqrt{\sigma}\|_1$.

Theorem 10.2.3 (cf. [67], Lemma 18)**.** *Any symmetric state* ρ *on* $\mathrm{H}^{\otimes n}$ *satisfies*

$$\rho \leqslant \binom{n+d^2-1}{n}^3 \int_{|\psi\rangle\in S_{\mathrm{H}\otimes\mathrm{H}'}} F\left(\rho,\sigma(\psi)^{\otimes n}\right)^2\sigma(\psi)^{\otimes n}\,\mathrm{d}\psi,$$

where for a unit vector $|\psi\rangle \in \mathrm{H}\otimes\mathrm{H}'$, $\sigma(\psi) = \mathrm{Tr}_{\mathrm{H}'}|\psi\rangle\langle\psi|$ *is the reduced state of* $|\psi\rangle\langle\psi|$ *on* H.

Proof. As noted before, there exists a unit vector $|\theta\rangle \in \mathrm{Sym}^n(\mathrm{H}\otimes\mathrm{H}')$ such that $\rho = \mathrm{Tr}_{\mathrm{H}'^{\otimes n}}|\theta\rangle\langle\theta|$. By Proposition 10.2.2, we have

$$|\theta\rangle\langle\theta| \leqslant \binom{n+d^2-1}{n}^3 \int_{|\psi\rangle\in S_{\mathrm{H}\otimes\mathrm{H}'}} \left|\langle\theta|\psi^{\otimes n}\rangle\right|^2|\psi\rangle\langle\psi|^{\otimes n}\,\mathrm{d}\psi.$$

Thus, after partial tracing over $\mathrm{H}'^{\otimes n}$, we obtain

$$\rho \leqslant \binom{n+d^2-1}{n}^3 \int_{|\psi\rangle\in S_{\mathrm{H}\otimes\mathrm{H}'}} \left|\langle\theta|\psi^{\otimes n}\rangle\right|^2\sigma(\psi)^{\otimes n}\,\mathrm{d}\psi.$$

To get the announced result, we then just have to notice that, by monotonicity of the fidelity under the CPTP map $\mathrm{Tr}_{\mathrm{H}'^{\otimes n}}$, we have for each $|\psi\rangle \in \mathrm{H}\otimes\mathrm{H}'$,

$$\left|\langle\theta|\psi^{\otimes n}\rangle\right| = F\left(|\theta\rangle\langle\theta|,|\psi\rangle\langle\psi|^{\otimes n}\right) \leqslant F\left(\rho,\sigma(\psi)^{\otimes n}\right).$$

□

10.2.2 Linear constraints

Let ρ be a symmetric state on $\mathrm{H}^{\otimes n}$. What Theorem 10.2.3 tells us is that there exists a probability measure μ over the set of states on H such that

$$\rho \leqslant \binom{n+d^2-1}{n}^3 \int_{\sigma\in\in\mathscr{D}(\mathrm{H})} F\left(\rho,\sigma^{\otimes n}\right)^2 \sigma^{\otimes n}\,\mathrm{d}\mu(\sigma) \leqslant (n+1)^{3d^2} \int_{\sigma\in\in\mathscr{D}(\mathrm{H})} F\left(\rho,\sigma^{\otimes n}\right)^2 \sigma^{\otimes n}\,\mathrm{d}\mu(\sigma). \quad (10.3)$$

It may be worth pointing out that μ is in fact the uniform probability measure over the set of mixed states on H (with respect to the Hilbert–Schmidt distance), since the latter is equivalently characterized as the partial trace over an environment H' having same dimension as H of uniformly distributed pure states on $\mathrm{H}\otimes\mathrm{H}'$ (see e.g. [166] for a proof).

Observe that, contrary to the original de Finetti reduction, where the upper bound is the same for every symmetric state, we here have a highly state-dependent upper bound, where only states which have a high fidelity with the state of interest ρ are given an important weight. This is especially useful when one knows that ρ satisfies some additional property. Indeed, one would then expect that, amongst states of the form $\sigma^{\otimes n}$, only those approximately satisfying this same property should have a non-negligible fidelity weight. There are at least two archetypical cases where this intuition can easily be seen to be true.

Corollary 10.2.4. *Let $\mathcal{N}:\mathscr{L}(\mathrm{H})\to\mathscr{L}(\mathrm{K})$ be a quantum channel, with $d=|\mathrm{H}|<+\infty$. Assume that ρ is a symmetric state on $\mathrm{H}^{\otimes n}$, which is additionally satisfying $\mathcal{N}^{\otimes n}(\rho)=\tau_0^{\otimes n}$, for some given state τ_0 on* K. *Then,*

$$\rho \leqslant (n+1)^{3d^2} \int_{\sigma\in\mathscr{D}(\mathrm{H})} F\left(\tau_0,\mathcal{N}(\sigma)\right)^{2n} \sigma^{\otimes n}\,\mathrm{d}\mu(\sigma).$$

Proof. This follows directly from inequality (10.3) by monotonicity of the fidelity under the CPTP map $\mathcal{N}$, and by multiplicativity of the fidelity on tensor products. □

This especially implies that, under the hypotheses of Corollary 10.2.4, we have: for any $0<\delta<1$, setting $\mathscr{K}_\delta=\{\sigma\in\mathscr{D}(\mathrm{H})\ :\ F\left(\tau_0,\mathcal{N}(\sigma)\right)\geqslant 1-\delta\}$,

$$\rho \leqslant (n+1)^{3d^2}\left(\int_{\sigma\in\mathscr{K}_\delta}\sigma^{\otimes n}\mathrm{d}\mu(\sigma)+(1-\delta)^{2n}\int_{\sigma\notin\mathscr{K}_\delta}\sigma^{\otimes n}\mathrm{d}\mu(\sigma)\right).$$

Such flexible de Finetti reduction, for states which satisfy the constraint of being sent to a certain tensor power state by a certain tensor power CPTP map, has already been fruitfully applied, for instance in the context of zero-error communication via quantum channels [67].

Another linear constraint is that of a fixed point equation.

Corollary 10.2.5. *Let $\mathcal{N}:\mathscr{L}(\mathrm{H})\to\mathscr{L}(\mathrm{H})$ be a quantum channel, with $d=|\mathrm{H}|<+\infty$. Assume that ρ is a symmetric state on $\mathrm{H}^{\otimes n}$, which is additionally satisfying $\mathcal{N}^{\otimes n}(\rho)=\rho$. Then,*

$$\rho \leqslant (n+1)^{3d^2} \int_{\sigma\in\mathscr{D}(\mathrm{H})} F\left(\rho,\mathcal{N}(\sigma)^{\otimes n}\right)^2 \mathcal{N}(\sigma)^{\otimes n}\mathrm{d}\mu(\sigma).$$

Proof. Apply $\mathcal{N}^{\otimes n}$ on both sides of inequality (10.3), and use once more the monotonicity of the fidelity under the CPTP map $\mathcal{N}$. □

This means that, under the assumptions of Corollary 10.2.5, there actually exists a probability measure $\widetilde{\mu}$ over the set of states on H which belong to the range of $\mathcal{N}$ such that

$$\rho \leqslant (n+1)^{3d^2} \int_{\sigma\in\mathrm{Range}(\mathcal{N})} F\left(\rho,\sigma^{\otimes n}\right)^2 \sigma^{\otimes n}\mathrm{d}\widetilde{\mu}(\sigma). \quad (10.4)$$

A case of particular interest for equation (10.4) is the following. Let G be a subgroup of the unitary group on H, equipped with its Haar measure μ_G (unique normalised left and right invariant measure over G). Its associated twirl is the quantum channel $\mathscr{T}_G:\mathscr{L}(\mathrm{H})\to\mathscr{L}(\mathrm{H})$ defined by

$$\mathscr{T}_G:\sigma\mapsto\int_{U\in G} U\sigma U^\dagger \mathrm{d}\mu_G(U).$$

The range of $\mathscr{T}_G$ is then precisely the set of states on H in the commutant of G, i.e.

$$\mathscr{K}_G=\{\sigma\in\mathscr{D}(\mathrm{H})\ :\ \forall\, U\in G,\ [\sigma,U]=0\}.$$

Hence, there exists a probability measure $\widetilde{\mu}$ over $\mathscr{K}_G$ such that, if ρ is a symmetric state on $\mathrm{H}^{\otimes n}$ satisfying $\mathcal{T}_G^{\otimes n}(\rho) = \rho$, then

$$\rho \leqslant (n+1)^{3d^2} \int_{\sigma \in \mathscr{K}_G} F\left(\rho, \sigma^{\otimes n}\right)^2 \sigma^{\otimes n} \mathrm{d}\widetilde{\mu}(\sigma).$$

Another situation where equation (10.4) might be especially useful is when $\mathcal{N}$ is a quantum-classical channel, so that its range can be identified with the set of classical probability distributions. We get in that case the corollary below.

Corollary 10.2.6. *Let $\mathfrak{X}$ be a finite alphabet and let $P_{\mathfrak{X}^n}$ be a symmetric probability distribution on $\mathfrak{X}^n$. There exists a universal probability measure $\mathrm{d}Q_{\mathfrak{X}}$ over the set of probability distributions on $\mathfrak{X}$ such that*

$$P_{\mathfrak{X}^n} \leqslant (n+1)^{3|\mathfrak{X}|^2} \int_{Q_{\mathfrak{X}}} F\left(P_{\mathfrak{X}^n}, Q_{\mathfrak{X}}^{\otimes n}\right)^2 Q_{\mathfrak{X}}^{\otimes n} \mathrm{d}Q_{\mathfrak{X}},$$

where the inequality sign signifies point-wise inequality between probability distributions on $\mathfrak{X}^n$.

Proof. This is a special case of Corollary 10.2.5. Indeed, we can make the identification $\mathfrak{X} \equiv \{1, \ldots, d\}$, where $d = |\mathfrak{X}|$. So let H be a d-dimensional Hilbert space, and denote by $\{|1\rangle, \ldots, |d\rangle\}$ an orthonormal basis of H. We can then define the "classical" state ρ on $\mathrm{H}^{\otimes n}$ by

$$\rho = \sum_{1 \leqslant x_1, \ldots, x_n \leqslant d} P(x_1, \ldots, x_n) |x_1 \otimes \cdots \otimes x_n\rangle\langle x_1 \otimes \cdots \otimes x_n|,$$

and the quantum-classical channel $\mathcal{N} : \mathscr{L}(\mathrm{H}) \to \mathscr{L}(\mathrm{H})$ by

$$\mathcal{N} : \sigma \mapsto \sum_{1 \leqslant x \leqslant d} Q_\sigma(x) |x\rangle\langle x| = \sum_{1 \leqslant x \leqslant d} |x\rangle\langle x|\sigma|x\rangle\langle x|.$$

By assumption on P, ρ is a symmetric state on $\mathrm{H}^{\otimes n}$, which is additionally, by construction, a fixed point of $\mathcal{N}^{\otimes n}$. Hence, by Corollary 10.2.5,

$$\rho \leqslant (n+1)^{3d^2} \int_{\sigma \in \mathscr{D}(\mathrm{H})} F\left(\rho, \mathcal{N}(\sigma)^{\otimes n}\right)^2 \mathcal{N}(\sigma)^{\otimes n} \mathrm{d}\mu(\sigma).$$

By the way ρ and $\mathcal{N}$ have been designed, this actually translates into the point-wise inequality

$$\forall\ 1 \leqslant x_1, \ldots, x_n \leqslant d,\ P(x_1, \ldots, x_n) \leqslant (n+1)^{3d^2} \int_{\sigma \in \mathscr{D}(\mathrm{H})} F\left(P, Q_\sigma^{\otimes n}\right)^2 Q_\sigma(x_1) \cdots Q_\sigma(x_n) \mathrm{d}\mu\left(Q_\sigma\right),$$

which is exactly the announced result. □

This flexible de Finetti reduction for probability distributions turns out to be especially useful when studying the parallel repetition of multi-player non-local games, as exemplified in Chapter 11.

Remark 10.2.7. *Note that all these results generalize to non-normalized permutation invariant positive semidefinite operators on finite-dimensional spaces (or positive distributions on finite alphabets). One just has to extend the usual definition of the fidelity by setting $F(M, N) = \|\sqrt{M}\sqrt{N}\|_1$ for any positive semidefinite operators (or positive distributions) M, N.*

10.2.3 On to convex constraints?

We just saw that, in the case where the symmetric state ρ under consideration is additionally known to satisfy certain linear constraints, it is possible to upper bound it by a de Finetti operator where either no or exponentially small weight is given to tensor power states which do not satisfy this same constraint. But what about the case where the a priori information on ρ is that it belongs or not to a certain convex subset of states? This is the question we investigate in the sequel, focussing first in Sections 10.3 and 10.4 on the paradigmatic example of the set of separable states, and then describing in Section 10.5 the general setting in which similar conclusions hold.

10.3 Exponential decay and concentration of $h_{\mathcal{S}}$ via de Finetti reduction approach

As we just mentioned, we will now be interested for a while in the case where the underlying Hilbert space is a tensor product Hilbert space $\mathrm{H} = \mathrm{A} \otimes \mathrm{B}$, and the kind of symmetric states on $\mathrm{H}^{\otimes n}$ that we will look at are those which additionally satisfy the (convex but non-linear) constraint of being separable across the bipartite cut $\mathrm{A}^{\otimes n}$:$\mathrm{B}^{\otimes n}$. For such a state ρ, one can of course still write down a de Finetti reduction of the form

$$\rho \leqslant (n+1)^{3|\mathrm{A}|^2|\mathrm{B}|^2} \int_{\sigma \in \mathcal{D}(\mathrm{A}\otimes\mathrm{B})} F\left(\rho, \sigma^{\otimes n}\right)^2 \sigma^{\otimes n} \, \mathrm{d}\mu(\sigma).$$

And what we would like to understand is whether it is possible to argue that only the states $\sigma^{\otimes n}$ which are such that σ is separable across the bipartite cut A:B are given a non exponentially small weight in this integral representation. As we shall see, this question is especially relevant when analysing the multiplicative behaviour of the support function of the set of biseparable states.

So let us specify a bit what we have in mind. Given a positive operator M on $\mathrm{A} \otimes \mathrm{B}$, its maximum overlap with states which are separable across the bipartite cut A:B, which we denote by $\mathcal{S}(\mathrm{A}{:}\mathrm{B})$, is defined as

$$h_{\mathcal{S}(\mathrm{A}:\mathrm{B})}(M) = \sup_{\sigma \in \mathcal{S}(\mathrm{A}:\mathrm{B})} \mathrm{Tr}(M\sigma).$$

Our interest is in understanding how this quantity behaves under tensoring. Concretely, this means that we want to know, for any $n \in \mathbf{N}$, how $h_{\mathcal{S}(\mathrm{A}^n:\mathrm{B}^n)}(M^{\otimes n})$ relates to $h_{\mathcal{S}(\mathrm{A}:\mathrm{B})}(M)$ (where the former quantity is defined as the maximum overlap of $M^{\otimes n}$ with states which are separable across the bipartite cut $\mathrm{A}^{\otimes n}$:$\mathrm{B}^{\otimes n}$). Note that here, and in the remainder of this chapter, we use the shorthand notation $\mathrm{A}^n, \mathrm{B}^n$ for $\mathrm{A}^{\otimes n}, \mathrm{B}^{\otimes n}$ whenever they appear subscript. Because $h_{\mathcal{S}(\mathrm{A}:\mathrm{B})}$ is linear homogeneous in its argument, we can always rescale M by a positive constant such that $0 \leqslant M \leqslant \mathrm{Id}$, meaning that M can be interpreted as a POVM element of the binary test with operators $(M, \mathrm{Id} - M)$. We shall make this assumption throughout from now on. Then, it is easy to see that, for any $n \in \mathbf{N}$, we have the inequalities

$$\left(h_{\mathcal{S}(\mathrm{A}:\mathrm{B})}(M)\right)^n \leqslant h_{\mathcal{S}(\mathrm{A}^n:\mathrm{B}^n)}(M^{\otimes n}) \leqslant h_{\mathcal{S}(\mathrm{A}:\mathrm{B})}(M) \leqslant 1. \tag{10.5}$$

But in the case where $h_{\mathcal{S}(\mathrm{A}:\mathrm{B})}(M) < 1$, the gap between the lower and upper bounds in equation (10.5) grows exponentially with n, making these inequalities very little informative.

This problem is interesting in itself, but also because it connects to plethora of others, some of them even outside the purely quantum information range of applications. The reader is referred to [94] for a full list of problems which are exactly or approximately equivalent to estimating $h_{\mathcal{S}}$. Two notable applications of $h_{\mathcal{S}}$ arise in quantum computing and in quantum Shannon theory.

The first is to QMA(2), the class of quantum Merlin-Arthur interactive proof systems with two unentangled provers. The setting is that a verifier requires states α and β from separate provers which are assumed to be computationally unlimited, and then performs a binary test with POVM $(M, \mathrm{Id} - M)$ on the separable state $\alpha \otimes \beta$. The maximum probability of passing the test that the provers can achieve evidently equals precisely $h_{\mathcal{S}}(M)$. For complexity theoretic considerations (in particular the so-called soundness gap amplification) it is important to understand how well many instances of the same test, performed in parallel, can be passed, either all n, leading to $h_{\mathcal{S}}(M^{\otimes n})$, or t out of n, where $t > n\, h_{\mathcal{S}}(M)$.

The second application appears in the problem of minimum output entropies of quantum channels, and their asymptotic behaviour. Namely, a quantum channel $\mathcal{N} : \mathcal{L}(\mathrm{A}) \to \mathcal{L}(\mathrm{B})$ can be represented in Stinespring form by $\mathcal{N}(X) = \mathrm{Tr}_{\mathrm{E}}(VXV^\dagger)$, with an isometry $V : \mathrm{A} \hookrightarrow \mathrm{B} \otimes \mathrm{E}$. Its minimum output Rényi p-entropy is given by

$$S_p^{\min}(\mathcal{N}) = \min_{\rho \in \mathcal{D}(\mathrm{A})} S_p(\mathcal{N}(\rho)), \text{ where } \forall\ \sigma \in \mathcal{D}(\mathrm{B}),\ S_p(\sigma) = -\frac{p}{p-1} \log \|\sigma\|_p.$$

Taking the limit, we recover the von Neumann entropy $S(\sigma)$ for $p = 1$ and $-\log \lambda_{\max}(\sigma)$ for $p = \infty$ (see Chapter 5, Sections 5.1 and 5.5, for further details). From this, it is not hard to see that, with $M = VV^\dagger$ the projector onto the range of V, i.e. the subspace $V(\mathrm{A}) \subset \mathrm{B} \otimes \mathrm{E}$, we have $S_\infty^{\min}(\mathcal{N}) = -\log h_{\mathcal{S}(\mathrm{B}:\mathrm{E})}(M)$. In quantum Shannon theory, the behaviour of $S_p^{\min}(\mathcal{N}^{\otimes n})$, and in particular of $S_\infty^{\min}(\mathcal{N}^{\otimes n}) = -\log h_{\mathcal{S}(\mathrm{B}^n:\mathrm{E}^n)}(M^{\otimes n})$ as n grows is of great interest.

10.3.1 Some general facts about "filtered by measurements" distance measures

We need to introduce first a few definitions and properties regarding "filtered by measurements" distance measures. The reader is e.g. referred to Chapter 6, Section 6.2, for more precise definitions.

Let H be a Hilbert space and let $\mathbf{M}$ be a set of POVMs on H. For any states ρ, σ on H, we define their measured by $\mathbf{M}$ trace-norm distance as

$$d_{\mathbf{M}}(\rho,\sigma) = \sup_{\mathscr{M}\in\mathbf{M}} \frac{1}{2} \left\| \mathscr{M}(\rho) - \mathscr{M}(\sigma) \right\|_1 ,$$

and their measured by $\mathbf{M}$ fidelity distance as

$$F_{\mathbf{M}}(\rho,\sigma) = \inf_{\mathscr{M}\in\mathbf{M}} F\left(\mathscr{M}(\rho), \mathscr{M}(\sigma)\right).$$

We have the well-known relations between these two distances (see e.g. [51], Chapter 9)

$$1 - F_{\mathbf{M}} \leqslant d_{\mathbf{M}} \leqslant \left(1 - F_{\mathbf{M}}^2\right)^{1/2}. \tag{10.6}$$

We further define, for any set of states $\mathscr{K}$ on H, the measured by $\mathbf{M}$ trace-norm distance of ρ to $\mathscr{K}$ as

$$d_{\mathbf{M}}\left(\rho, \mathscr{K}\right) = \inf_{\sigma\in\mathscr{K}} d_{\mathbf{M}}(\rho,\sigma),$$

and the measured by $\mathbf{M}$ fidelity distance of ρ to $\mathscr{K}$ as

$$F_{\mathbf{M}}\left(\rho, \mathscr{K}\right) = \sup_{\sigma\in\mathscr{K}} F_{\mathbf{M}}(\rho,\sigma),$$

In the sequel, we shall consider the case where $\mathrm{H} = \mathrm{A}\otimes\mathrm{B}$ is a bipartite Hilbert space, with $|\mathrm{A}|, |\mathrm{B}| < +\infty$. In this setting, we will be especially interested in the set of separable states $\mathcal{S}$ and in the set of separable POVMs **SEP** on H (in the bipartite cut A:B).

Lemma 10.3.1. *Let $\mathrm{A}_1, \mathrm{B}_1, \mathrm{A}_2, \mathrm{B}_2$ be Hilbert spaces. Then, for any states ρ_1 on $\mathrm{A}_1\otimes\mathrm{B}_1$ and ρ_2 on $\mathrm{A}_2\otimes\mathrm{B}_2$, we have*

$$F\big(\rho_1\otimes\rho_2, \mathcal{S}(\mathrm{A}_1\mathrm{A}_2{:}\mathrm{B}_1\mathrm{B}_2)\big) \leqslant F_{\mathbf{SEP}}\big(\rho_1, \mathcal{S}(\mathrm{A}_1{:}\mathrm{B}_1)\big) F\big(\rho_2, \mathcal{S}(\mathrm{A}_2{:}\mathrm{B}_2)\big).$$

Proof. The proof is directly inspired from [147], adapted here to the case of fidelities rather than relative entropies.

Let $\mathscr{M}_1 \equiv \left(M_1^{(i)}\right)_{i\in I} \in \mathbf{SEP}(\mathrm{A}_1{:}\mathrm{B}_1)$. Then, by monotonicity of the fidelity under the CPTP map $\mathscr{M}_1\otimes\mathcal{I}d_2$, we have

$$\begin{aligned}\sup_{\sigma_{12}\in\mathcal{S}(\mathrm{A}_1\mathrm{A}_2:\mathrm{B}_1\mathrm{B}_2)} F\left(\rho_1\otimes\rho_2, \sigma_{12}\right) &\leqslant \sup_{\sigma_{12}\in\mathcal{S}(\mathrm{A}_1\mathrm{A}_2:\mathrm{B}_1\mathrm{B}_2)} F\left(\mathscr{M}_1\otimes\mathcal{I}d_2(\rho_1\otimes\rho_2), \mathscr{M}_1\otimes\mathcal{I}d_2(\sigma_{12})\right) \\ &= F\left(\mathscr{M}_1\otimes\mathcal{I}d_2(\rho_1\otimes\rho_2), \mathscr{M}_1\otimes\mathcal{I}d_2(\widetilde{\sigma}_{12})\right),\end{aligned}$$

for some $\widetilde{\sigma}_{12} \in \mathcal{S}(\mathrm{A}_1\mathrm{A}_2{:}\mathrm{B}_1\mathrm{B}_2)$. Next,

$$F\left(\mathscr{M}_1\otimes\mathcal{I}d_2(\rho_1\otimes\rho_2), \mathscr{M}_1\otimes\mathcal{I}d_2(\widetilde{\sigma}_{12})\right) = \sum_{i\in I} \sqrt{\mathrm{Tr}\left(M_1^{(i)}\rho_1\right)}\sqrt{\mathrm{Tr}\left(M_1^{(i)}\widetilde{\sigma}_1\right)} F\left(\rho_2, \widetilde{\sigma}_2^{(i)}\right),$$

where $\widetilde{\sigma}_1 = \mathrm{Tr}_{\mathrm{A}_2\mathrm{B}_2}\left(\widetilde{\sigma}_{12}\right)$ and for each $i\in I$, $\widetilde{\sigma}_2^{(i)} = \mathrm{Tr}_{\mathrm{A}_1\mathrm{B}_1}\left(M_1^{(i)}\otimes\mathrm{Id}_2\widetilde{\sigma}_{12}\right)/\mathrm{Tr}_{\mathrm{A}_1\mathrm{B}_1}\left(M_1^{(i)}\widetilde{\sigma}_1\right)$, so $\widetilde{\sigma}_1 \in \mathcal{S}(\mathrm{A}_1{:}\mathrm{B}_1)$ and for each $i\in I$, $\widetilde{\sigma}_2^{(i)} \in \mathcal{S}(\mathrm{A}_2{:}\mathrm{B}_2)$. Hence, for each $i\in I$, $F\big(\rho_2, \widetilde{\sigma}_2^{(i)}\big) \leqslant \sup_{\sigma_2\in\mathcal{S}(\mathrm{A}_2:\mathrm{B}_2)} F\left(\rho_2,\sigma_2\right)$, and subsequently

$$\begin{aligned}\sum_{i\in I} \sqrt{\mathrm{Tr}\left(M_1^{(i)}\rho_1\right)}\sqrt{\mathrm{Tr}\left(M_1^{(i)}\widetilde{\sigma}_1\right)} F\big(\rho_2, \widetilde{\sigma}_2^{(i)}\big) &\leqslant \left(\sup_{\sigma_2\in\mathcal{S}_{\mathrm{A}_2:\mathrm{B}_2}} F\left(\rho_2,\sigma_2\right)\right) F\left(\mathscr{M}_1(\rho_1), \mathscr{M}_1(\widetilde{\sigma}_1)\right) \\ &\leqslant \left(\sup_{\sigma_2\in\mathcal{S}(\mathrm{A}_2:\mathrm{B}_2)} F\left(\rho_2,\sigma_2\right)\right)\left(\sup_{\sigma_1\in\mathcal{S}(\mathrm{A}_1:\mathrm{B}_1)} F\left(\mathscr{M}_1(\rho_1), \mathscr{M}_1(\sigma_1)\right)\right).\end{aligned}$$

We thus have shown that, for any $\mathscr{M}_1 \in \mathbf{SEP}(\mathrm{A}_1{:}\mathrm{B}_1)$,

$$F\big(\rho_1\otimes\rho_2, \mathcal{S}(\mathrm{A}_1\mathrm{A}_2{:}\mathrm{B}_1\mathrm{B}_2)\big) \leqslant \left(\sup_{\sigma_1\in\mathcal{S}(\mathrm{A}_1:\mathrm{B}_1)} F\left(\mathscr{M}_1(\rho_1), \mathscr{M}_1(\sigma_1)\right)\right) F\left(\rho_2, \mathcal{S}(\mathrm{A}_2{:}\mathrm{B}_2)\right).$$

Taking the infimum over $\mathscr{M}_1 \in \mathbf{SEP}(\mathrm{A}_1{:}\mathrm{B}_1)$, we get precisely the statement in Lemma 10.3.1. □

Theorem 10.3.2. *Let* A, B *be Hilbert spaces, and let* ρ *be a state on* $\mathrm{A} \otimes \mathrm{B}$*. Then, for any* $n \in \mathbf{N}$*,*

$$F\big(\rho^{\otimes n}, \mathcal{S}(\mathrm{A}^n{:}\mathrm{B}^n)\big) \leqslant F_{\mathbf{SEP}(\mathrm{A}:\mathrm{B})}\big(\rho, \mathcal{S}(\mathrm{A}{:}\mathrm{B})\big)^n.$$

Proof. Theorem 10.3.2 is a direct corollary of Lemma 10.3.1, obtained by iterating the latter. □

10.3.2 Weak multiplicativity of $h_{\mathcal{S}}$

With these facts prepared, we can now derive our main theorem.

Theorem 10.3.3. *Let* M *be an operator on the tensor product Hilbert space* $\mathrm{A} \otimes \mathrm{B}$*, satisfying* $0 \leqslant M \leqslant \mathrm{Id}$*, and set* $r = \|M\|_2$*. If* $h_{\mathcal{S}(\mathrm{A}:\mathrm{B})}(M) \leqslant 1 - \delta$*, for some* $0 < \delta < 1$*, then for any* $n \in \mathbf{N}$*,*

$$h_{\mathcal{S}(\mathrm{A}^n:\mathrm{B}^n)}(M^{\otimes n}) \leqslant \left(1 - \frac{\delta^2}{5r^2}\right)^n.$$

Proof. Let $\rho \in \mathcal{S}(\mathrm{A}^n{:}\mathrm{B}^n)$. Our goal will be first of all to show that $\mathrm{Tr}\,(M^{\otimes n}\rho) \leqslant 2(n+1)^{3|\mathrm{A}|^2|\mathrm{B}|^2}\left(1 - \delta^2/2r^2\right)^n$. Now observe that, denoting, for each permutation $\pi \in \mathfrak{S}(n)$, by $U(\pi) \in \mathcal{U}\big((\mathrm{A} \otimes \mathrm{B})^{\otimes n}\big)$ the associated permutation unitary,

$$\mathrm{Tr}\left(M^{\otimes n}\rho\right) = \mathrm{Tr}\left(\left(\frac{1}{n!}\sum_{\pi\in\mathfrak{S}(n)} U(\pi)M^{\otimes n}U(\pi)^\dagger\right)\rho\right) = \mathrm{Tr}\left(M^{\otimes n}\left(\frac{1}{n!}\sum_{\pi\in\mathfrak{S}(n)} U(\pi)^\dagger\rho U(\pi)\right)\right).$$

The first equality is by n-symmetry of $M^{\otimes n}$ and the second one is by cyclicity of the trace. Hence, for our purposes, we may actually assume without loss of generality that $\rho \in \mathcal{S}(\mathrm{A}^n{:}\mathrm{B}^n)$ is n-symmetric.

Yet, if ρ is an n-symmetric state on $(\mathrm{A} \otimes \mathrm{B})^{\otimes n}$, we know by Theorem 10.2.3 that there exists a probability measure μ on the set of states on $\mathrm{A} \otimes \mathrm{B}$ such that

$$\rho \leqslant (n+1)^{3|\mathrm{A}|^2|\mathrm{B}|^2}\int_{\sigma\in\mathcal{D}(\mathrm{A}\otimes\mathrm{B})} F\left(\rho, \sigma^{\otimes n}\right)^2 \sigma^{\otimes n}\,\mathrm{d}\mu(\sigma).$$

So, by multiplicativity of the trace on tensor products, we get in that case

$$\mathrm{Tr}\left(M^{\otimes n}\rho\right) \leqslant (n+1)^{3|\mathrm{A}|^2|\mathrm{B}|^2}\int_{\sigma\in\mathcal{D}(\mathrm{A}\otimes\mathrm{B})} F\left(\rho, \sigma^{\otimes n}\right)^2 \mathrm{Tr}\,(M\sigma)^n\,\mathrm{d}\mu(\sigma).$$

Consequently, for any $0 < \epsilon < 1$, setting $\mathscr{K}_\epsilon = \{\sigma \in \mathcal{D}(\mathrm{A} \otimes \mathrm{B}) : \|\sigma - \mathcal{S}(\mathrm{A}{:}\mathrm{B})\|_2 \leqslant \epsilon/r\}$, we have, upper bounding either $F\left(\rho, \sigma^{\otimes n}\right)$ or $\mathrm{Tr}\,(M\sigma)$ by 1,

$$\mathrm{Tr}\left(M^{\otimes n}\rho\right) \leqslant (n+1)^{3|\mathrm{A}|^2|\mathrm{B}|^2}\left(\int_{\sigma\in\mathscr{K}_\epsilon}\mathrm{Tr}\,(M\sigma)^n\,\mathrm{d}\mu(\sigma) + \int_{\sigma\notin\mathscr{K}_\epsilon} F\left(\rho, \sigma^{\otimes n}\right)^2\mathrm{d}\mu(\sigma)\right).$$

Now, if $\sigma \in \mathscr{K}_\epsilon$, this means that there exists $\tau \in \mathcal{S}(\mathrm{A}{:}\mathrm{B})$ such that $\|\sigma - \tau\|_2 \leqslant \epsilon/r$, so that

$$\mathrm{Tr}(M\sigma) = \mathrm{Tr}(M\tau) + \mathrm{Tr}(M(\sigma - \tau)) \leqslant \mathrm{Tr}(M\tau) + \|M\|_2\|\sigma - \tau\|_2 \leqslant 1 - \delta + \epsilon.$$

The next to last inequality is simply by the Cauchy–Schwarz inequality, while the last one is by assumption on M, τ, σ. And if $\sigma \notin \mathscr{K}_\epsilon$, then

$$F\left(\rho, \sigma^{\otimes n}\right) \leqslant F\left(\sigma^{\otimes n}, \mathcal{S}(\mathrm{A}^n{:}\mathrm{B}^n)\right) \leqslant F_{\mathbf{SEP}(\mathrm{A}:\mathrm{B})}\left(\sigma, \mathcal{S}(\mathrm{A}{:}\mathrm{B})\right)^n \leqslant \left(1 - \frac{\epsilon^2}{4r^2}\right)^{n/2}.$$

The first inequality is because $\rho \in \mathcal{S}(\mathrm{A}^n{:}\mathrm{B}^n)$, the second one is by Theorem 10.3.2, and the third one is obtained by combining equation (10.6) with the known lower bound $\|\sigma - \mathcal{S}(\mathrm{A}{:}\mathrm{B})\|_{\mathbf{SEP}(\mathrm{A}:\mathrm{B})} \geqslant \|\sigma - \mathcal{S}(\mathrm{A}{:}\mathrm{B})\|_2$ (see e.g. [130]).

Putting everything together, we obtain in the end that for any $0 < \epsilon < 1$,

$$\mathrm{Tr}\left(M^{\otimes n}\rho\right) \leqslant (n+1)^{3|\mathrm{A}|^2|\mathrm{B}|^2}\left((1 - \delta + \epsilon)^n + \left(1 - \frac{\epsilon^2}{4r^2}\right)^n\right). \tag{10.7}$$

In particular, choosing $\epsilon = 2r^2\left((1 + \delta/r^2)^{1/2} - 1\right)$ in equation (10.7), so that $\epsilon^2/4r^2 = \delta - \epsilon \geqslant \delta^2/5r^2$, we get $\mathrm{Tr}\,(M^{\otimes n}\rho) \leqslant 2(n+1)^{3|\mathrm{A}|^2|\mathrm{B}|^2}\left(1 - \delta^2/5r^2\right)^n$. And consequently

$$h_{\mathcal{S}(\mathrm{A}^n:\mathrm{B}^n)}(M^{\otimes n}) \leqslant 2(n+1)^{3|\mathrm{A}|^2|\mathrm{B}|^2}\left(1 - \frac{\delta^2}{5r^2}\right)^n. \tag{10.8}$$

In order to conclude, we just need to remove the polynomial pre-factor in equation (10.8). Assume that there exists a constant $C > 0$ such that $h_{\mathcal{S}(\mathrm{A}^N:\mathrm{B}^N)}(M^{\otimes N}) \geqslant C\left(1-\delta^2/5r^2\right)^N$ for some $N \in \mathbf{N}$. Then, we would have for any $n \in \mathbf{N}$,

$$h_{\mathcal{S}(\mathrm{A}^{Nn}:\mathrm{B}^{Nn})}\left(M^{\otimes Nn}\right) \geqslant \left(h_{\mathcal{S}(\mathrm{A}^N:\mathrm{B}^N)}\left(M^{\otimes N}\right)\right)^n \geqslant C^n\left(1-\frac{\delta^2}{5r^2}\right)^{Nn}.$$

On the other hand, equation (10.8) says that we also have

$$h_{\mathcal{S}(\mathrm{A}^{Nn}:\mathrm{B}^{Nn})}\left(M^{\otimes Nn}\right) \leqslant 2(Nn+1)^{3|\mathrm{A}|^2|\mathrm{B}|^2}\left(1-\frac{\delta^2}{5r^2}\right)^{Nn}.$$

Letting n grow, we see that the only option to make these two inequalities compatible is to have $C \leqslant 1$, which is precisely what we wanted to show. □

The conclusion of Theorem 10.3.3 had already been obtained via completely different techniques than the one presented here (and even with slightly better constants). However, the good thing about the de Finetti reduction approach is that it gives, almost for free, not only this exponential decay result for the behaviour of $h_{\mathcal{S}}$ under tensoring, but also some kind of concentration statement. To be precise, assume that M is an operator on $\mathrm{A} \otimes \mathrm{B}$, satisfying $0 \leqslant M \leqslant \mathrm{Id}$ and $h_{\mathcal{S}(\mathrm{A}:\mathrm{B})}(M) \leqslant 1-\delta$ for some $0 < \delta < 1$. Then, M can be identified with a binary test that a separable state is guaranteed to pass only with probability $h_{\mathcal{S}(\mathrm{A}:\mathrm{B})}(M) \leqslant 1-\delta$, while there exists some (entangled) state that would pass it with probability $h_{\mathcal{D}(\mathrm{A}\otimes\mathrm{B})}(M) = \|M\|_\infty$, which may be 1. Hence, a natural question would be: performing this test n times in parallel, what is the probability that a separable state passes a certain fraction t/n of them? Such maximum probability is nothing else than $h_{\mathcal{S}(\mathrm{A}^n:\mathrm{B}^n)}\left(M^{(t/n)}\right)$, where the operator $M^{(t/n)}$ on $(\mathrm{A}\otimes\mathrm{B})^{\otimes n}$ is defined as

$$M^{(t/n)} := \sum_{I\subset[n],\,|I|\geqslant t} M^{\otimes I} \otimes \mathrm{Id}^{\otimes I^c}.$$

Above (and several times later on as well) we use the following shorthand notation: for any $1 \leqslant k \leqslant n$ and any bipartition $\{i_1,\ldots,i_k\} \sqcup \{j_1,\ldots,j_{n-k}\} = \{1,\ldots,n\}$,

$$M_{\mathrm{AB}}^{\otimes\{i_1,\ldots,i_k\}} \otimes \mathrm{Id}_{\mathrm{AB}}^{\otimes\{j_1,\ldots,j_{n-k}\}} := M_{\mathrm{A}_{i_1}\mathrm{B}_{i_1}} \otimes \cdots \otimes M_{\mathrm{A}_{i_k}\mathrm{B}_{i_k}} \otimes \mathrm{Id}_{\mathrm{A}_{j_1}\mathrm{B}_{j_1}} \otimes \cdots \otimes \mathrm{Id}_{\mathrm{A}_{j_{n-k}}\mathrm{B}_{j_{n-k}}}.$$

Obviously, if $t < (1-\delta)n$ then the answer is asymptotically 1, whereas for $t = n$ the answer is $h_{\mathcal{S}(\mathrm{A}^n:\mathrm{B}^n)}(M^{\otimes n})$, which decays exponentially fast with n as established in Theorem 10.3.3. But is such exponential amplification of the failing probability already true for t just slightly above $(1-\delta)n$? Theorem 10.3.4 answers this question affirmatively.

Theorem 10.3.4. *Let M be an operator on the tensor product Hilbert space $\mathrm{A}\otimes\mathrm{B}$, satisfying $0 \leqslant M \leqslant \mathrm{Id}$, and set $r = \|M\|_2$. If $h_{\mathcal{S}(\mathrm{A}:\mathrm{B})}(M) \leqslant 1-\delta$ for some $0 < \delta < 1$, then for any $n, t \in \mathbf{N}$ with $t \geqslant (1-\delta+\alpha)n$ for some $0 < \alpha \leqslant \delta$, we have*

$$h_{\mathcal{S}(\mathrm{A}^n:\mathrm{B}^n)}\left(M^{(t/n)}\right) \leqslant \exp\left(-n\frac{\alpha^2}{5r^2}\right).$$

Proof. Following the exact same lines as in the proof of Theorem 10.3.3, we now have in place of equation (10.7)

$$h_{\mathcal{S}(\mathrm{A}^n:\mathrm{B}^n)}\left(M^{(t/n)}\right) \leqslant (n+1)^{3|\mathrm{A}|^2|\mathrm{B}|^2}\left(\exp\left[-2n(\alpha-\epsilon)^2\right] + \exp\left[-n\frac{\epsilon^2}{4r^2}\right]\right). \tag{10.9}$$

This is indeed a consequence of Hoeffding's inequality (and of the fact that $e^{-x} \geqslant 1-x$ for any $x > 0$). So in particular, choosing $\epsilon = \alpha\left(1-(\sqrt{2}-1)/(8r^2-1)\right)$ in equation (10.9), so that $\epsilon^2/4r^2 = 2(\alpha-\epsilon)^2 \geqslant \alpha^2/5r^2$, and removing the polynomial pre-factor by the same trick as in the proof of Theorem 10.3.3, we get as announced

$$h_{\mathcal{S}(\mathrm{A}^n:\mathrm{B}^n)}\left(M^{(t/n)}\right) \leqslant \exp\left(-n\frac{\alpha^2}{5r^2}\right).$$ □

10.4 Exponential decay and concentration of h_S via entanglement measure approach

10.4.1 Quantifying the disturbance induced by measurements

Let us state first a few technical lemmas that we will need later on to establish our main result. The reader is referred to Chapter 2, Section 2.5, for all quantum Shannon theory definitions which are used in this section (entropy and relative entropy, mutual information etc.).

Lemma 10.4.1. *Let ρ be a state on $\mathrm{U}\otimes\mathrm{V}$ and T be an operator on U, satisfying $0\leqslant T\leqslant \mathrm{Id}$. Define next $p=\mathrm{Tr}_{\mathrm{UV}}\left[(T_{\mathrm{U}}\otimes\mathrm{Id}_{\mathrm{V}})\,\rho_{\mathrm{UV}}\right]$ as the probability of obtaining the first outcome when the two-outcome POVM $(T_{\mathrm{U}}\otimes\mathrm{Id}_{\mathrm{V}},(\mathrm{Id}_{\mathrm{U}}-T_{\mathrm{U}})\otimes\mathrm{Id}_{\mathrm{V}})$ is performed on ρ_{UV}, and $\tau_{\mathrm{V}}=\mathrm{Tr}_{\mathrm{U}}\left[(T_{\mathrm{U}}\otimes\mathrm{Id}_{\mathrm{V}})\,\rho_{\mathrm{UV}}\right]/p$ as the corresponding post-measurement state on V. Also, denote by $\rho_{\mathrm{V}}=\mathrm{Tr}_{\mathrm{U}}\left[\rho_{\mathrm{UV}}\right]$ the reduced state of ρ_{UV} on V. Then,*

$$D\left(\tau_{\mathrm{V}}\middle\|\rho_{\mathrm{V}}\right)\leqslant-\log p.$$

Proof. Note that $\rho_{\mathrm{V}}=p\tau_{\mathrm{V}}+(1-p)\sigma_{\mathrm{V}}$, where $\sigma_{\mathrm{V}}=\mathrm{Tr}_{\mathrm{U}}\left[((\mathrm{Id}_{\mathrm{U}}-T_{\mathrm{U}})\otimes\mathrm{Id}_{\mathrm{V}})\,\rho_{\mathrm{UV}}\right]/(1-p)$. We therefore have the operator inequality $p\tau_{\mathrm{V}}\leqslant\rho_{\mathrm{V}}$. And hence,

$$D\left(\tau_{\mathrm{V}}\middle\|\rho_{\mathrm{V}}\right)=\mathrm{Tr}\left[\tau_{\mathrm{V}}\left(\log\tau_{\mathrm{V}}-\log\rho_{\mathrm{V}}\right)\right]\leqslant\mathrm{Tr}\left[\tau_{\mathrm{V}}\left(\log\tau_{\mathrm{V}}-\log(p\tau_{\mathrm{V}})\right)\right]=-\log p,$$

the next to last inequality being because log is an operator monotone function. □

Let us recall the definition of the squashed entanglement E_{sq}, introduced in [44]:

$$E_{sq}\left(\rho_{\mathrm{AB}}\right)=\inf\left\{\frac{1}{2}I(\mathrm{A}{:}\mathrm{B}|\mathrm{E})\left(\rho_{\mathrm{ABE}}\right)\ :\ \mathrm{Tr}_{\mathrm{E}}\left(\rho_{\mathrm{ABE}}\right)=\rho_{\mathrm{AB}}\right\}.$$

Lemma 10.4.2. *Let M_{AB} be an operator on the tensor product Hilbert space $\mathrm{A}\otimes\mathrm{B}$, satisfying $0\leqslant M_{\mathrm{AB}}\leqslant\mathrm{Id}$, and let $\alpha_{\mathrm{A}^n},\beta_{\mathrm{B}^n}$ be states on $\mathrm{A}^{\otimes n},\mathrm{B}^{\otimes n}$ respectively. Next, fix $1\leqslant k\leqslant n-1$, and define*

$$p_k=\mathrm{Tr}_{\mathrm{A}^n\mathrm{B}^n}\left[\left(M_{\mathrm{AB}}^{\otimes k}\otimes\mathrm{Id}_{\mathrm{AB}}^{\otimes n-k}\right)\alpha_{\mathrm{A}^n}\otimes\beta_{\mathrm{B}^n}\right]$$
$$\tau(k)_{\mathrm{A}^{n-k}\mathrm{B}^{n-k}}=\frac{1}{p_k}\mathrm{Tr}_{\mathrm{A}^k\mathrm{B}^k}\left[\left(M_{\mathrm{AB}}^{\otimes k}\otimes\mathrm{Id}_{\mathrm{AB}}^{\otimes n-k}\right)\alpha_{\mathrm{A}^n}\otimes\beta_{\mathrm{B}^n}\right].$$

Then, it holds that

$$\sum_{j=k+1}^{n}E_{sq}\left(\tau(k)_{\mathrm{A}_j\mathrm{B}_j}\right)\leqslant\frac{1}{2}\log\frac{1}{p_k}.$$

Proof. By Lemma 10.4.1, with $\mathrm{U}=\mathrm{A}^{\otimes k}\otimes\mathrm{B}^{\otimes k}$, $\mathrm{V}=\mathrm{A}^{\otimes n-k}\otimes\mathrm{B}^{\otimes n-k}$, $T_{\mathrm{U}}=M_{\mathrm{AB}}^{\otimes k}$ and $\rho_{\mathrm{UV}}=\alpha_{\mathrm{A}^n}\otimes\beta_{\mathrm{B}^n}$, we know that

$$D\left(\tau(k)_{\mathrm{A}^{n-k}\mathrm{B}^{n-k}}\middle\|\alpha_{\mathrm{A}^{n-k}}\otimes\beta_{\mathrm{B}^{n-k}}\right)\leqslant\log\frac{1}{p_k}.$$

Now, observe that

$$\begin{aligned}
D\left(\tau(k)_{\mathrm{A}^{n-k}\mathrm{B}^{n-k}}\middle\|\alpha_{\mathrm{A}^{n-k}}\otimes\beta_{\mathrm{B}^{n-k}}\right)&\geqslant D\left(\tau(k)_{\mathrm{A}^{n-k}\mathrm{B}^{n-k}}\middle\|\tau(k)_{\mathrm{A}^{n-k}}\otimes\tau(k)_{\mathrm{B}^{n-k}}\right)\\
&=I\left(\mathrm{A}_{k+1}\ldots\mathrm{A}_n{:}\mathrm{B}_{k+1}\ldots\mathrm{B}_n\right)\left(\tau(k)_{\mathrm{A}^{n-k}\mathrm{B}^{n-k}}\right)\\
&=\sum_{j=k+1}^{n}I\left(\mathrm{A}_j{:}\mathrm{B}_{k+1}\ldots\mathrm{B}_n|\mathrm{A}_{k+1}\ldots\mathrm{A}_{j-1}\right)\left(\tau(k)_{\mathrm{A}^{n-k}\mathrm{B}^{n-k}}\right)\\
&\geqslant\sum_{j=k+1}^{n}I\left(\mathrm{A}_j{:}\mathrm{B}_j|\mathrm{A}_{k+1}\ldots\mathrm{A}_{j-1}\right)\left(\tau(k)_{\mathrm{A}^{n-k}\mathrm{B}^{n-k}}\right)\\
&\geqslant\sum_{j=k+1}^{n}2\,E_{sq}\left(\tau(k)_{\mathrm{A}_j\mathrm{B}_j}\right).
\end{aligned}$$

The first inequality is due to the fact that, given a bipartite state τ_{UV} on $\mathrm{U}\otimes\mathrm{V}$, for any states ρ_{U}, ρ_{V} on U, V respectively, $D(\tau_{\mathrm{UV}}\|\rho_{\mathrm{U}}\otimes\rho_{\mathrm{V}})\geqslant D(\tau_{\mathrm{UV}}\|\tau_{\mathrm{U}}\otimes\tau_{\mathrm{V}})$. The third equality and the fourth inequality follow from the chain rule and the monotonicity under discarding of subsystems, respectively, for the quantum mutual information. And the last inequality is by definition of the squashed entanglement. □

Remark 10.4.3. *Observe that under the assumptions of Lemma 10.4.2, we actually have the stronger conclusion*

$$\sum_{j=k+1}^{n} E_I\left(\tau(k)_{\mathrm{A}_j\mathrm{B}_j}\right) \leqslant \frac{1}{2}\log\frac{1}{p_k},$$

where E_I is the conditional entanglement of mutual information (CEMI) *introduced in [109]:*

$$E_I\left(\rho_{\mathrm{AB}}\right) = \inf\left\{\frac{1}{2}\left[I(\mathrm{AA}'{:}\mathrm{BB}')\left(\rho_{\mathrm{AA}'\mathrm{BB}'}\right) - I(\mathrm{A}'{:}\mathrm{B}')\left(\rho_{\mathrm{AA}'\mathrm{BB}'}\right)\right] \;:\; \mathrm{Tr}_{\mathrm{A}'\mathrm{B}'}\left(\rho_{\mathrm{AA}'\mathrm{BB}'}\right) = \rho_{\mathrm{AB}}\right\}.$$

CEMI is always at least as large as squashed entanglement: for any state ρ_{AB}, $E_I(\rho_{\mathrm{AB}}) \geqslant E_{sq}(\rho_{\mathrm{AB}})$. But the precise relation between these two entanglement measures is unknown. In [109] it was furthermore shown that, like squashed entanglement, CEMI is additive, and more generally super-additive in the sense that

$$E_I(\rho_{\mathrm{A}_1\mathrm{A}_2:\mathrm{B}_1\mathrm{B}_2}) \geqslant E_I(\rho_{\mathrm{A}_1:\mathrm{B}_1}) + E_I(\rho_{\mathrm{A}_2:\mathrm{B}_2}).$$

However, unlike squashed entanglement, there is no simple proof of monogamy of CEMI, and it may well not hold in general.

Lemma 10.4.4 (cf. [104], Lemma 8.6)**.** *Let $0 < \nu < 1$ and $c > 0$. Let also $n \in \mathbf{N}$ and assume that $(p_k)_{1\leqslant k\leqslant n}$ is a sequence of numbers satisfying $1 > p_1 \geqslant \cdots \geqslant p_n > 0$ and*

$$\forall\ 1 \leqslant k \leqslant n-1,\ p_{k+1} \leqslant p_k\left(\sqrt{\frac{c}{n-k}\log\frac{1}{p_k}} + \nu\right).$$

Then, for any $0 < \gamma < 1-\nu$ such that $p_1 \leqslant \nu+\gamma$, we have

$$\forall\ 1 \leqslant k \leqslant n,\ p_k \leqslant (\nu+\gamma)^{\min(k,k_0)} \text{ for } k_0 = \frac{\gamma^2}{c\log[1/(\nu+\gamma)] + \gamma^2}(n+1).$$

Proof. To prove Lemma 10.4.4, we only have to show that

$$\forall\ 1 \leqslant k \leqslant k_0,\ p_k \leqslant (\nu+\gamma)^k. \tag{10.10}$$

Indeed, the case $k > k_0$ then directly follows from the assumption that the sequence $(p_k)_{1\leqslant k\leqslant n}$ is non-increasing, so that $p_k \leqslant p_{k_0} \leqslant (\nu+\gamma)^{k_0}$.

Let us establish (10.10) by recursivity. The statement obviously holds for $k = 1$ since $p_1 \leqslant \nu+\gamma$ by hypothesis. So assume next that it holds for some $k \leqslant k_0 - 1$. If $p_k \leqslant (\nu+\gamma)^{k+1}$, then clearly $p_{k+1} \leqslant p_k \leqslant (\nu+\gamma)^{k+1}$. Otherwise, by the way p_{k+1} is related to p_k, we then have

$$p_{k+1} \leqslant (\nu+\gamma)^k\left(\sqrt{\frac{c(k+1)\log[1/(\nu+\gamma)]}{(n-k)}} + \nu\right),$$

and the latter quantity is smaller than $(\nu+\gamma)^{k+1}$ if

$$\frac{k+1}{n-k} \leqslant \frac{\gamma^2}{c\log[1/(\nu+\gamma)]},$$

which can be checked to be equivalent to $k+1 \leqslant k_0$. Hence in both cases, the statement holds for $k+1$. □

Corollary 10.4.5 (cf. [104], Lemma 8.6)**.** *Let $(p_k)_{1\leqslant k\leqslant n}$ be a sequence of numbers satisfying the assumptions of Lemma 10.4.4, and the additional condition $p_1 \leqslant 1-(1-\nu)/2$. Then,*

$$p_n \leqslant \left(1 - \frac{(1-\nu)^2}{8c}\right)^n.$$

Proof. Corollary 10.4.5 follows from applying Lemma 10.4.4 in the particular case $\gamma = (1-\nu)/2$. Indeed, we then have $\nu+\gamma = 1-(1-\nu)/2 = (1+\nu)/2$, so that

$$k_0 = \frac{(1-\nu)^2}{4c\log[2/(1+\nu)] + (1-\nu)^2}(n+1) \geqslant \frac{(1-\nu)^2}{4c\log[2^{1-\nu}] + 1}(n+1) \geqslant \frac{1-\nu}{4c}n,$$

And consequently,

$$p_n \leqslant \left(1 - \frac{1-\nu}{2}\right)^{n(1-\nu)/(4c)} \leqslant \left(1 - \frac{(1-\nu)^2}{8c}\right)^n,$$

which is exactly the announced upper bound for p_n. □

10.4.2 Weak multiplicativity of $h_{\mathcal{S}}$

Our approach in this section, to prove the multiplicative behaviour of $h_{\mathcal{S}}$, is directly inspired from the seminal angle of attack to the parallel repetition problem for classical non-local games: our Theorem 10.4.6 is an analogue of the exponential decay results by Raz [151] and Holenstein [104], while our Theorem 10.4.8 is an analogue of the concentration bound result by Rao [150]. Indeed, here in the same spirit as theirs, we want to make precise the following intuition: if the initial state of a system on $(\mathrm{A}\otimes\mathrm{B})^{\otimes n}$ is product across the cut $\mathrm{A}^n{:}\mathrm{B}^n$, then performing a measurement $(M, \mathrm{Id}-M)$ on only a few subsystems $\mathrm{A}\otimes\mathrm{B}$ should not create too much correlations in the post-measurement state on the remaining subsystems.

Before we prove the main result of this section, we need to recall one last definition: for any $q\in\mathbf{N}$, a state ρ_{AB} on a bipartite Hilbert space $\mathrm{A}\otimes\mathrm{B}$ is said to be q-extendible with respect to B if there exists a state ρ_{AB^q} on $\mathrm{A}\otimes\mathrm{B}^{\otimes q}$ that is invariant under any permutation of the B-subsystems and such that $\rho_{\mathrm{AB}} = \mathrm{Tr}_{\mathrm{B}^{q-1}}\rho_{\mathrm{AB}^q}$ (cf. Chapter 9). We shall denote by $\mathcal{E}_q(\mathrm{A}{:}\mathrm{B})$ the set of q-extendible states with respect to B on $\mathrm{A}\otimes\mathrm{B}$, and by $h_{\mathcal{E}_q(\mathrm{A}:\mathrm{B})}$ its associated support function.

Theorem 10.4.6. *Let M be an operator on the tensor product Hilbert space $\mathrm{A}\otimes\mathrm{B}$, satisfying $0\leqslant M\leqslant\mathrm{Id}$. Then, for any $q\in\mathbf{N}$,*

$$h_{\mathcal{S}(\mathrm{A}^n:\mathrm{B}^n)}\left(M^{\otimes n}\right)\leqslant\left(1-\frac{\left(1-h_{\mathcal{E}_q(\mathrm{A}:\mathrm{B})}(M)\right)^2}{8\ln 2\,q^2}\right)^n. \tag{10.11}$$

And consequently, if $h_{\mathcal{S}(\mathrm{A}:\mathrm{B})}(M)\leqslant 1-\delta$ for some $0<\delta<1$, then

$$h_{\mathcal{S}(\mathrm{A}^n:\mathrm{B}^n)}\left(M^{\otimes n}\right)\leqslant\left(1-\frac{\delta^4}{512\ln 2\,d^4}\right)^n, \tag{10.12}$$

assuming $|\mathrm{A}|=|\mathrm{B}|=d$.

Proof. To establish the first statement (10.11), we have to show that,

$$\forall\ \rho_{\mathrm{A}^n\mathrm{B}^n}\in\mathcal{S}(\mathrm{A}^n{:}\mathrm{B}^n),\ \mathrm{Tr}\left(M_{\mathrm{AB}}^{\otimes n}\rho_{\mathrm{A}^n\mathrm{B}^n}\right)\leqslant\left(1-\frac{\left(1-h_{\mathcal{E}_q(\mathrm{A}:\mathrm{B})}(M_{\mathrm{AB}})\right)^2}{8\ln 2\,q^2}\right)^n.$$

Note that, with this aim in view, we can without loss of generality focuss only on states which are extremal in $\mathcal{S}(\mathrm{A}^n{:}\mathrm{B}^n)$, namely on states which are product across the cut $\mathrm{A}^{\otimes n}{:}\mathrm{B}^{\otimes n}$. So let $\alpha_{\mathrm{A}^n}\otimes\beta_{\mathrm{B}^n}$ be such a state, and set $p_0=1$, $\tau(0)_{\mathrm{A}^n\mathrm{B}^n}=\alpha_{\mathrm{A}^n}\otimes\beta_{\mathrm{B}^n}$. In the sequel, we will use the following notation: given $I_k\subset[n]$ with $|I_k|=k$, define $M^{(I_k)}_{\mathrm{A}^n\mathrm{B}^n}$ as

$$M^{(I_k)}_{\mathrm{A}^n\mathrm{B}^n}=M_{\mathrm{AB}}^{\otimes I_k}\otimes\mathrm{Id}_{\mathrm{AB}}^{\otimes I_k^c}.$$

Then, for each $1\leqslant k\leqslant n$, construct recursively

$$p_k=\mathrm{Tr}_{\mathrm{A}^n\mathrm{B}^n}\left[M^{(I_k)}_{\mathrm{A}^n\mathrm{B}^n}\alpha_{\mathrm{A}^n}\otimes\beta_{\mathrm{B}^n}\right],$$
$$\tau(k)_{\mathrm{A}_{I_k^c}\mathrm{B}_{I_k^c}}=\frac{1}{p_k}\mathrm{Tr}_{\mathrm{A}_{I_k}\mathrm{B}_{I_k}}\left[M^{(I_k)}_{\mathrm{A}^n\mathrm{B}^n}\alpha_{A^n}\otimes\beta_{B^n}\right],$$

with i_k chosen in I_{k-1}^c such that

$$E_{sq}\left(\tau(k-1)_{\mathrm{A}_{i_k}\mathrm{B}_{i_k}}\right)\leqslant\frac{1}{n-k+1}\frac{1}{2}\log\frac{1}{p_{k-1}}.$$

We know that this is possible. Indeed, assuming that p_{k-1}, $\tau(k-1)_{\mathrm{A}_{I_{k-1}^c}\mathrm{B}_{I_{k-1}^c}}$ have been constructed, Lemma 10.4.2 guarantees that

$$\frac{1}{n-k+1}\sum_{j=1}^{n-k+1}E_{sq}\left(\tau(k-1)_{\mathrm{A}_{I_{k-1}^c}\mathrm{B}_{I_{k-1}^c}}\right)\leqslant\frac{1}{n-k+1}\frac{1}{2}\log\frac{1}{p_{k-1}},$$

so that there necessarily exists an index $i\in I_{k-1}^c$ such that $E_{sq}\big(\tau(k-1)_{\mathrm{A}_i\mathrm{B}_i}\big)$ is smaller than the quantity on the right-hand-side of the average upper bound above.

Now, notice that the p_k, $0 \leqslant k \leqslant n$, are related by the recursion formula

$$\forall\ 0 \leqslant k \leqslant n-1,\ p_{k+1} = p_k \operatorname{Tr}_{\mathrm{A}_{i_{k+1}}\mathrm{B}_{i_{k+1}}} \left(M_{\mathrm{A}_{i_{k+1}}\mathrm{B}_{i_{k+1}}} \tau(k)_{\mathrm{A}_{i_{k+1}}\mathrm{B}_{i_{k+1}}} \right),$$

where, by the way the $\tau(k)$, $0 \leqslant k \leqslant n$, are built

$$E_{sq}\left(\tau(k)_{\mathrm{A}_{i_{k+1}}\mathrm{B}_{i_{k+1}}}\right) \leqslant \frac{1}{n-k}\frac{1}{2}\log\frac{1}{p_k}.$$

Yet, we know from [133] that this implies that

$$\exists\ \sigma_{\mathrm{AB}} \in \mathscr{C}_q(\mathrm{A{:}B}) :\ \left\|\tau(k)_{\mathrm{AB}} - \sigma_{\mathrm{AB}}\right\|_1 \leqslant \sqrt{2\ln 2}(q-1)\sqrt{\frac{1}{n-k}\frac{1}{2}\log\frac{1}{p_k}} \leqslant \sqrt{\frac{\ln 2\, q^2}{n-k}\log\frac{1}{p_k}}. \tag{10.13}$$

And therefore,

$$p_{k+1} \leqslant p_k \left(\|M_{\mathrm{AB}}\|_\infty \left\|\tau(k)_{\mathrm{AB}} - \sigma_{\mathrm{AB}}\right\|_1 + \operatorname{Tr}\left(M_{\mathrm{AB}}\sigma_{\mathrm{AB}}\right)\right) \leqslant p_k \left(\sqrt{\frac{\ln 2\, q^2}{n-k}\log\frac{1}{p_k}} + h_{\mathscr{C}_q(\mathrm{A:B})}\left(M_{\mathrm{AB}}\right)\right).$$

With this upper bound, and because we also clearly have $1 > p_1 \geqslant \cdots \geqslant p_n > 0$ as well as the requirement $p_1 \leqslant h_{\mathcal{S}(\mathrm{A:B})}(M) \leqslant h_{\mathscr{C}_q(\mathrm{A:B})}(M) \leqslant 1 - (1 - h_{\mathscr{C}_q(\mathrm{A:B})})/2$, it follows from Corollary 10.4.5 that

$$\operatorname{Tr}\left(M_{\mathrm{AB}}^{\otimes n}\alpha_{\mathrm{A}^n} \otimes \beta_{\mathrm{B}^n}\right) = p_n \leqslant \left(1 - \frac{\left(1 - h_{\mathscr{C}_q(\mathrm{A:B})}(M_{\mathrm{AB}})\right)^2}{8\ln 2\, q^2}\right)^n,$$

which is precisely what we wanted to prove.

From there, the second statement (10.12) easily follows. Indeed, in the case where $|\mathrm{A}| = |\mathrm{B}| = d$, we know from [48] that, for any $q \in \mathbf{N}$, if $\rho_{\mathrm{AB}} \in \mathscr{C}_q(\mathrm{A{:}B})$ then there exists $\sigma_{\mathrm{AB}} \in \mathcal{S}(\mathrm{A{:}B})$ such that $\|\rho_{\mathrm{AB}} - \sigma_{\mathrm{AB}}\|_1 \leqslant 2d^2/q$, so that $h_{\mathscr{C}_q(\mathrm{A:B})} \leqslant h_{\mathcal{S}(\mathrm{A:B})} + 2d^2/q$. Hence, if $h_{\mathcal{S}(\mathrm{A:B})}(M_{\mathrm{AB}}) \leqslant 1 - \delta$, making the choice $q = 4d^2/\delta$, in order to have $h_{\mathscr{C}_q(\mathrm{A:B})}(M_{\mathrm{AB}}) \leqslant 1 - \delta/2$, yields, after a straightforward computation, exactly the announced exponential decay result. □

The scaling as $(\delta/d)^4$ in the upper bound provided by equation (10.12) of Theorem 10.4.6 is much worse than the scaling as $(\delta/d)^2$ in the upper bound provided by Theorem 10.3.3. However, equation (10.11) of Theorem 10.4.6, which relates $h_{\mathcal{S}(\mathrm{A}^n:\mathrm{B}^n)}(M^{\otimes n})$ to $h_{\mathscr{C}_q(\mathrm{A:B})}(M)$, may be of interest in some specific cases, namely when M has a maximum overlap with q-extendible states which is already of the same order as its maximum overlap with separable states for $q \ll d^2$.

Remark 10.4.7. *By Remark 10.4.3, we see that we could also have done the recursive construction described in the proof of Theorem 10.4.6 by imposing instead that, for each* $0 \leqslant k \leqslant n-1$,

$$p_{k+1} = p_k \operatorname{Tr}_{\mathrm{A}_{i_{k+1}}\mathrm{B}_{i_{k+1}}} \left(M_{\mathrm{A}_{i_{k+1}}\mathrm{B}_{i_{k+1}}} \tau(k)_{\mathrm{A}_{i_{k+1}}\mathrm{B}_{i_{k+1}}} \right),\ \textit{with}\ E_I\left(\tau(k)_{\mathrm{A}_{i_{k+1}}\mathrm{B}_{i_{k+1}}}\right) \leqslant \frac{1}{n-k}\frac{1}{2}\log\frac{1}{p_k}.$$

Now, it is an open question to determine whether there exists a dimension-independent constant $C > 0$ *such that an implication of the following form holds true:*

$$E_I\left(\rho_{\mathrm{AB}}\right) \leqslant \epsilon\ \Rightarrow\ \exists\ \sigma_{\mathrm{AB}} \in \mathcal{S}(\mathrm{A{:}B}) :\ \left\|\rho_{\mathrm{AB}} - \sigma_{\mathrm{AB}}\right\|_1 \leqslant C\sqrt{\epsilon}. \tag{10.14}$$

If Conjecture (10.14) *indeed held, this would imply that the* $(p_k)_{1 \leqslant k \leqslant n}$ *satisfy*

$$\forall\ 0 \leqslant k \leqslant n-1,\ p_{k+1} \leqslant p_k \left(\sqrt{\frac{C^2/2}{n-k}\log\frac{1}{p_k}} + h_{\mathcal{S}(\mathrm{A:B})}\left(M_{\mathrm{AB}}\right)\right).$$

And hence eventually, the following dimension-free exponential decay result for $h_{\mathcal{S}(\mathrm{A:B})}$*:*

$$h_{\mathcal{S}(\mathrm{A:B})}(M) \leqslant 1 - \delta\ \Rightarrow\ h_{\mathcal{S}(\mathrm{A}^n:\mathrm{B}^n)}\left(M^{\otimes n}\right) \leqslant \left(1 - \frac{\delta^2}{4C^2}\right)^n.$$

In fact, if a more general variant of Conjecture (10.14) *held, with* $C\sqrt{\epsilon}$ *replaced by* $\varphi(\epsilon)$ *for* φ *a (universal) non-decreasing function such that* $\varphi(0) = 0$*, then one could prove analogously that*

$$h_{\mathcal{S}(\mathrm{A:B})}(M) \leqslant 1 - \delta\ \Rightarrow\ h_{\mathcal{S}(\mathrm{A}^n:\mathrm{B}^n)}\left(M^{\otimes n}\right) \leqslant \left(1 - \frac{\varphi^{-1}(\delta)}{4}\right)^n.$$

The way property (10.13) *of strong faithfulness of squashed entanglement with respect to q-extendible states, is proved in [133] is relying on the breakthrough result by Fawzi and Renner [72] that small conditional mutual information does imply approximate recoverability. Now, in an even stronger manner than* $E_{sq}(\rho)$ *being small means that the conditional mutual information of any extension of* ρ *is small,* $E_I(\rho)$ *being small is a condition that is expressible as a bunch of conditional mutual information of extensions of* ρ *being simultaneously small. So it could be that recoverability results (in particular the best one up-to-date [120], which carries the advantage over the original one [72] of being universal and explicit) would help in an attempt to prove a strong faithfulness property of CEMI with respect to separable states such as* (10.14).

Theorem 10.4.8. *Let* M *be an operator on the tensor product Hilbert space* $\mathrm{A}\otimes\mathrm{B}$*, satisfying* $0\leqslant M\leqslant \mathrm{Id}$*. If* $h_{\mathcal{S}(\mathrm{A}:\mathrm{B})}(M)\leqslant 1-\delta$ *for some* $0<\delta<1$*, then for any* $n,t\in\mathbf{N}$ *with* $t\geqslant(1-\delta+\alpha)n$ *for some* $0<\alpha\leqslant\delta$*, we have*

$$h_{\mathcal{S}(\mathrm{A}^n:\mathrm{B}^n)}\left(M^{(t/n)}\right)\leqslant\left(1-\frac{\alpha^5}{2048\ln 2\,d^4\,(2\delta-\alpha)}\right)^n,$$

assuming $|\mathrm{A}|=|\mathrm{B}|=d$*.*

Proof. The proof of this theorem follows a very similar route to that of Theorem 10.4.6: for any given state $\alpha_{\mathrm{A}^n}\otimes\beta_{\mathrm{B}^n}$ which is product across the cut $\mathrm{A}^{\otimes n}$:$\mathrm{B}^{\otimes n}$, we want to show that the probability that it passes at least t amongst n tests defined by M_{AB} is upper bounded as

$$P_t(\alpha_{\mathrm{A}^n}\otimes\beta_{\mathrm{B}^n})\leqslant\left(1-\frac{\alpha^5}{2048\ln 2\,d^4\,(2\delta-\alpha)}\right)^n.$$

In that aim, we start by defining the following deterministic set, number and state: $I_0=\emptyset$, $p_{I_0}=1$ and $\tau(I_0)_{\mathrm{A}^n\mathrm{B}^n}=\alpha_{\mathrm{A}^n}\otimes\beta_{\mathrm{B}^n}$. Then, for each $1\leqslant k\leqslant n$, we construct recursively the following random set, number and state: pick i_k uniformly at random in I_{k-1}^c, and define

$$\begin{aligned}
&I_k=I_{k-1}\cup\{i_k\},\\
&p_{I_k}=\mathrm{Tr}_{\mathrm{A}^n\mathrm{B}^n}\left[M^{(I_k)}_{\mathrm{A}^n\mathrm{B}^n}\alpha_{\mathrm{A}^n}\otimes\beta_{\mathrm{B}^n}\right],\\
&\tau(I_k)_{\mathrm{A}_{I_k^c}\mathrm{B}_{I_k^c}}=\frac{1}{p_{I_k}}\mathrm{Tr}_{\mathrm{A}_{I_k}\mathrm{B}_{I_k}}\left[M^{(I_k)}_{\mathrm{A}^n\mathrm{B}^n}\alpha_{\mathrm{A}^n}\otimes\beta_{\mathrm{B}^n}\right].
\end{aligned}$$

Lemma 10.4.2 guarantees that, on average, for each $0\leqslant k\leqslant n-1$,

$$E_{sq}\left(\overline{\tau}(I_k)_{\mathrm{A}_{i_{k+1}}\mathrm{B}_{i_{k+1}}}\right)\leqslant\frac{1}{n-k}\frac{1}{2}\log\frac{1}{\overline{p}_{I_k}},$$

so that, on average, for any $q\in\mathbf{N}$,

$$\overline{p}_{I_{k+1}}\leqslant\overline{p}_{I_k}\left(\sqrt{\frac{\ln 2\,q^2}{n-k}\log\frac{1}{\overline{p}_{I_k}}}+h_{\mathcal{E}_q(\mathrm{A}:\mathrm{B})}\left(M_{\mathrm{AB}}\right)\right).$$

In particular, we can make the choice $q=8d^2/\alpha$, in order to have $h_{\mathcal{E}_q(\mathrm{A}:\mathrm{B})}(M_{\mathrm{AB}})\leqslant 1-\delta+\alpha/4$. And we thus get from Lemma 10.4.4, after computation, that on average,

$$\overline{p}_{I_{k_0}}\leqslant\left(1-\delta+\frac{\alpha}{2}\right)^{k_0}\text{ for }k_0=\frac{\alpha^4}{1024\ln 2\,d^4\,\log[1/(1-\delta+\alpha/2)]+\alpha^4}(n+1)\geqslant\frac{\alpha^4}{1024\ln 2\,d^4\,(2\delta-\alpha)}n.$$

To finish off the proof, we just have to observe (cf. [150], Section 8) that

$$\begin{aligned}
P_t(\alpha_{\mathrm{A}^n}\otimes\beta_{\mathrm{B}^n})&\leqslant\sum_{I_{k_0}\subset[n],\,|I_{k_0}|=k_0}\frac{1}{\binom{(1-\delta+\alpha)n}{k_0}}p_{I_{k_0}}\\
&\leqslant\frac{\binom{n}{k_0}}{\binom{(1-\delta+\alpha)n}{k_0}}\overline{p}_{I_{k_0}}\\
&\leqslant\left(\frac{n-k_0+1}{(1-\delta+\alpha)n-k_0+1}\right)^{k_0}\left(1-\delta+\frac{\alpha}{2}\right)^{k_0},
\end{aligned}$$

where the last inequality follows from the fact that $\prod_{i=0}^{l-1}(a+i)/(b+i) \leqslant (a/b)^l$, combined with the upper bound on $\overline{p}_{I_{k_0}}$. In the end, we can therefore conclude that

$$P_t(\alpha_{\mathrm{A}^n} \otimes \beta_{\mathrm{B}^n}) \leqslant \left(1 - \frac{\alpha}{2}\right)^{k_0} \leqslant \left(1 - \frac{\alpha^5}{2048 \ln 2\, d^4\,(2\delta - \alpha)}\right)^n,$$

where the first inequality follows from the fact that $(1 - \delta + \alpha/2)/(1 - \delta + \alpha) \leqslant 1 - \alpha/2$, while the second inequality is a consequence of the lower bound on k_0. □

Remark 10.4.9. *Here again, we see by Remark 10.4.3 that, if Conjecture* (10.14) *held, then we could have obtained in the proof of Theorem 10.4.8 that, on average*

$$\overline{p}_{I_{k_0}} \leqslant \left(1 - \delta + \frac{\alpha}{2}\right)^{k_0} \text{ for } k_0 = \frac{\alpha^2}{8\log[1/(1-\delta+\alpha/2)] + \alpha^2}(n+1) \geqslant \frac{\alpha^2}{8(2\delta - \alpha)}\, n.$$

And hence eventually, the following dimension-free concentration result for $h_{\mathcal{S}(\mathrm{A:B})}$*:*

$$h_{\mathcal{S}(\mathrm{A:B})}(M) \leqslant 1 - \delta \;\Rightarrow\; \forall\, 0 < \alpha < \delta,\ \forall\, t \geqslant (1 - \delta + \alpha)n,\ h_{\mathcal{S}(\mathrm{A}^n:\mathrm{B}^n)}\left(M^{(t/n)}\right) \leqslant \left(1 - \frac{\alpha^3}{16C^2(2\delta - \alpha)}\right)^n.$$

10.5 Equivalence between weak multiplicativity of support functions and of maximum fidelities

In the previous Sections 10.3 and 10.4, we studied in great depth one particular example of convex constraint on quantum states, namely the separability one. We showed in this specific case that there is a strong connection between the (weakly) multiplicative behaviour under tensoring of either the support function $h_{\mathcal{S}}$ or the maximum fidelity $F(\cdot, \mathcal{S})$. We would now like to describe, more generally, which kind of convex sets of states exhibit a similar feature.

So let us fix $d \in \mathbf{N}$, H a d-dimensional Hilbert space, and assume that we have a sequence of convex sets of states $\mathcal{K}^{(n)}$ on $\mathrm{H}^{\otimes n}$, $n \in \mathbf{N}$, with the following stability properties (under permutation and partial trace):

$$\rho \in \mathcal{K}^{(n)} \;\Rightarrow\; \forall\, \pi \in \mathfrak{S}(n),\ U(\pi)\rho U(\pi)^\dagger \in \mathcal{K}^{(n)} \text{ and } \mathrm{Tr}_{\mathrm{H}}\, \rho \in \mathcal{K}^{(n-1)}. \tag{10.15}$$

Note that requirement (10.15) implies in particular that, if $\rho^{\otimes n} \in \mathcal{K}^{(n)}$, then $\rho \in \mathcal{K}^{(1)}$. In view of our subsequent discussion, it would be meaningless not to impose that the opposite holds as well, i.e. that, if $\rho \in \mathcal{K}^{(1)}$, then $\rho^{\otimes n} \in \mathcal{K}^{(n)}$. This means in other words that, for each $n \in \mathbf{N}$, $\mathcal{K}^{(n)}$ is assumed to contain the so-called n^{th} projective tensor power of $\mathcal{K}^{(1)}$, which is defined as

$$\left(\mathcal{K}^{(1)}\right)^{\hat{\otimes} n} := \mathrm{conv}\left\{\rho_1 \otimes \cdots \otimes \rho_n,\ \rho_1, \ldots, \rho_n \in \mathcal{K}^{(1)}\right\}.$$

10.5.1 Exponential decay and concentration of $h_{\mathcal{K}}$ from multiplicativity of $F(\cdot, \mathcal{K})$

Given an operator M on H, satisfying $0 \leqslant M \leqslant \mathrm{Id}$, define the support function of $\mathcal{K}^{(n)}$ at $M^{\otimes n}$ as

$$h_{\mathcal{K}^{(n)}}\left(M^{\otimes n}\right) = \sup_{\sigma \in \mathcal{K}^{(n)}} \mathrm{Tr}\left(M^{\otimes n}\sigma\right).$$

Define also more generally, for any $0 \leqslant t \leqslant n$, $h_{\mathcal{K}^{(n)}}\left(M^{(t/n)}\right)$ as the maximum probability for a state in $\mathcal{K}^{(n)}$ to pass a fraction t/n of n binary tests $(M, \mathrm{Id} - M)$ performed in parallel. The question we are next interested in is to understand how $h_{\mathcal{K}^{(n)}}(M^{\otimes n})$ and $h_{\mathcal{K}^{(n)}}\left(M^{(t/n)}\right)$ relate to $h_{\mathcal{K}^{(1)}}(M)$.

Hence, assume also that these sets $\mathcal{K}^{(n)}$ satisfy the following condition: there exists a non-decreasing function $f : \epsilon \in]0,1[\mapsto f(\epsilon) \in]0,1[$ such that, for any state ρ on $\mathbf{C}^d$ and any $0 < \epsilon < 1$,

$$\left\|\rho - \mathcal{K}^{(1)}\right\|_2 \geqslant \epsilon \;\Rightarrow\; F\left(\rho^{\otimes n}, \mathcal{K}^{(n)}\right)^2 \leqslant (1 - f(\epsilon))^n. \tag{10.16}$$

Then, under assumption (10.16) for the sets $\mathcal{K}^{(n)}$, the following holds: for any operator M on H, satisfying $0 \leqslant M \leqslant \mathrm{Id}$, and $\|M\|_2 = r$ for some $0 \leqslant r \leqslant \sqrt{d}$, and any $0 < \alpha \leqslant \delta < 1$,

$$h_{\mathcal{K}^{(1)}}(M) \leqslant 1 - \delta \;\Rightarrow\; h_{\mathcal{K}^{(n)}}(M^{\otimes n}) \leqslant (1 - g(\delta, r))^n,$$
$$\forall\, t \geqslant (1 - \delta + \alpha)n,\ h_{\mathcal{K}^{(n)}}\left(M^{(t/n)}\right) \leqslant e^{-ng'(\alpha, r)},$$

where $g(\delta, r) = f(\epsilon(\delta, r))$ for $0 < \epsilon(\delta, r) < 1$ the solution of the equation $f(\epsilon) = \delta - r\epsilon$ and $g'(\alpha, r) = f(\epsilon(\alpha, r))$ for $0 < \epsilon(\alpha, r) < 1$ the solution of the equation $f(\epsilon) = 2(\alpha - r\epsilon)^2$.

To come to these statements, the strategy is entirely analogous to the one adopted in the proofs of Theorems 10.3.3 and 10.3.4. It is therefore only sketched below. First of all, when looking for a state $\rho \in \mathscr{K}^{(n)}$ maximizing $\mathrm{Tr}\,(M^{\otimes n}\rho)$, one can in fact assume without loss of generality that ρ is n-symmetric. And for such state ρ, reasoning as in the proof of Theorem 10.3.3, we know that we have

$$\mathrm{Tr}\left(M^{\otimes n}\rho\right) \leqslant (n+1)^{3d^2} \int_{\sigma \in \mathbb{D}(\mathrm{H})} F\left(\rho, \sigma^{\otimes n}\right)^2 \mathrm{Tr}\,(M\sigma)^n \,\mathrm{d}\mu(\sigma).$$

Hence, we get as a consequence of hypothesis (10.16) that, for any $0 < \epsilon < 1$,

$$\mathrm{Tr}\left(M^{\otimes n}\rho\right) \leqslant (n+1)^{3d^2} \left((1 - \delta + r\epsilon)^n + (1 - f(\epsilon))^n\right).$$

So choosing ϵ such that $f(\epsilon) = \delta - r\epsilon$ and ρ such that $\mathrm{Tr}\,(M^{\otimes n}\rho) = h_{\mathscr{K}^{(n)}}(M^{\otimes n})$ yields in particular

$$h_{\mathscr{K}^{(n)}}(M^{\otimes n}) \leqslant 2(n+1)^{3d^2} \left(1 - g(\delta, r)\right)^n.$$

Similarly, it follows from hypothesis (10.16) as well that, for any $0 < \epsilon < 1$,

$$h_{\mathscr{K}^{(n)}}\left(M^{(t/n)}\right) \leqslant (n+1)^{3d^2} \left(\exp\left[-2n(\alpha - r\epsilon)^2\right] + \exp\left[-nf(\epsilon)\right]\right),$$

so that choosing ϵ such that $f(\epsilon) = 2(\alpha - r\epsilon)^2$ gives

$$h_{\mathscr{K}^{(n)}}\left(M^{(t/n)}\right) \leqslant 2(n+1)^{3d^2} e^{-ng'(\alpha, r)}.$$

In both cases the polynomial pre-factor $2(n+1)^{3d^2}$ can then be removed by the exact same argument as in the proof of Theorem 10.3.3.

10.5.2 Weak multiplicativity of $F(\cdot, \mathscr{K})$ from exponential decay and concentration of $h_{\mathscr{K}}$

We would now like to go in the other direction. Namely, let us assume this time that these sets $\mathscr{K}^{(n)}$ satisfy the following condition: there exists a function $f : (\alpha, d) \in]0,1[\times\mathbf{N} \mapsto f(\alpha, d) \in]0,1[$, non-decreasing in α and non-increasing in d, such that, for any operator M on H, satisfying $0 \leqslant M \leqslant \mathrm{Id}$, and any $0 < \alpha \leqslant \delta < 1$,

$$h_{\mathscr{K}^{(1)}}(M) \leqslant 1 - \delta \;\Rightarrow\; \forall\, t \geqslant (1 - \delta + \alpha)n,\; h_{\mathscr{K}^{(n)}}\left(M^{(t/n)}\right) \leqslant e^{-nf(\alpha, d)}. \tag{10.17}$$

Then, under assumption (10.17) for the sets $\mathscr{K}^{(n)}$, the following holds: for any state ρ on H and any $0 < \epsilon < 1$,

$$\frac{1}{2}\left\|\rho - \mathscr{K}^{(1)}\right\|_1 \geqslant \epsilon \;\Rightarrow\; F\left(\rho^{\otimes n}, \mathscr{K}^{(n)}\right) \leqslant 2e^{-ng(\epsilon, d)},$$

where $g(\epsilon, d) = f(\alpha(\epsilon, d), d)/2$ for $0 < \alpha(\epsilon, d) < 1$ the solution of the equation $f(\alpha, d)/2 = (\alpha - \epsilon)^2$.

Here is the strategy to derive such result: Imagine you are given a state on $\mathrm{H}^{\otimes n}$, which you know is either $\rho^{\otimes n}$ or in $\mathscr{K}^{(n)}$, and you want to decide between these two hypotheses. For that, you can design a binary test (T_+, T_-) such that outcome $+$ is obtained with a high probability p if the state was $\rho^{\otimes n}$ and outcome $-$ is obtained with a high probability q if the state was in $\mathscr{K}^{(n)}$. Then, clearly

$$F\left(\rho^{\otimes n}, \mathscr{K}^{(n)}\right) \leqslant F\left((p, 1-p), (q, 1-q)\right) \leqslant \sqrt{1-p} + \sqrt{1-q}.$$

Therefore, if both error probabilities $1 - p$ and $1 - q$ are exponentially small, the conclusion follows.

In the present case, the fact that $\left\|\rho - \mathscr{K}^{(1)}\right\|_1 = 2\epsilon$, implies that there exist $0 \leqslant M \leqslant \mathrm{Id}$ and $\epsilon < \eta < 1$ such that $\mathrm{Tr}(M\rho) = 1 - \eta + \epsilon$ whereas $h_{\mathscr{K}^{(1)}}(M) = 1 - \eta$. So consider the binary POVM $(M_0, M_1) = (M, \mathrm{Id} - M)$ performed n times in parallel, and the corresponding binary test (T_+, T_-) with $+$ being the event "outcome 0 is obtained more than $(1 - \eta + \alpha)n$ times" and $-$ being the event "outcome 0 is obtained less than $(1 - \eta + \alpha)n$ times", for some $0 < \alpha < \epsilon$ to be chosen later. Define next, for each $1 \leqslant i \leqslant n$, the random variable X_i,

respectively Y_i, as the outcome of measurement number i given that the state was $\rho^{\otimes n}$, respectively in $\mathscr{K}^{(n)}$. We then have

$$1-p = \mathbf{P}\left(-\middle|\rho^{\otimes n}\right) = \mathbf{P}\left(\sum_{i=1}^{n} X_i < (1-\eta+\alpha)n\right),$$
$$1-q = \mathbf{P}\left(+\middle|\mathscr{K}^{(n)}\right) = \mathbf{P}\left(\sum_{i=1}^{n} Y_i > (1-\eta+\alpha)n\right).$$

Yet on the one hand, $X_1, \ldots, X_n$ are independent Bernoulli random variables with expectation $1-\eta+\epsilon$, so by Hoeffding's inequality

$$\mathbf{P}\left(\sum_{i=1}^{n} X_i < (1-\eta+\alpha)n\right) \leqslant e^{-2n(\epsilon-\alpha)^2}.$$

While on the other hand, for any $0 \leqslant t \leqslant n$, $\mathbf{P}\left(\sum_{i=1}^{n} Y_i > t\right) = h_{\mathscr{K}^{(n)}}\left(M^{(t/n)}\right)$, so assumption (10.17) guarantees that

$$\mathbf{P}\left(\sum_{i=1}^{n} Y_i > (1-\eta+\alpha)n\right) \leqslant e^{-nf(\alpha,d)}.$$

Hence, putting everything together, we eventually obtain that, for any $0 < \alpha < \epsilon$,

$$F\left(\rho^{\otimes n}, \mathscr{K}^{(n)}\right) \leqslant e^{-n(\epsilon-\alpha)^2} + e^{-nf(\alpha,d)/2},$$

which yields the wanted result after choosing α such that $f(\alpha,d)/2 = (\epsilon-\alpha)^2$.

Remark 10.5.1. *Note that requirement* (10.15) *is clearly fulfilled by the sets* $\mathcal{S}_{\mathrm{A}^n:\mathrm{B}^n}$ *of biseparable states on* $(\mathrm{A}\otimes\mathrm{B})^{\otimes n}$. *Furthermore, they satisfy requirements* (10.16) *and* (10.17) *as well, with* $f(\epsilon) = \epsilon^2/4$ *and* $f(\alpha, d^2) = \alpha^2/5d^2$.

It may also be worth emphasizing that conditions (10.16) *and* (10.17) *are just strengthened and quantitative versions of the following stability property for the sets* $\mathscr{K}^{(n)}$*:* $\rho \notin \mathscr{K}^{(1)} \;\Rightarrow\; \rho^{\otimes n} \notin \mathscr{K}^{(n)}$, *i.e. equivalently* $\rho^{\otimes n} \in \mathscr{K}^{(n)} \;\Rightarrow\; \rho \in \mathscr{K}^{(1)}$.

10.5.3 One simple example

Let us look at what the previous discussion becomes in the case of the simplest possible sequence $\{\mathscr{K}^{(n)},\ n \in \mathbf{N}\}$ satisfying requirement (10.15), namely when there exists a set of states $\mathscr{K}$ on H such that, for each $n \in \mathbf{N}$, $\mathscr{K}^{(n)}$ is exactly the nth projective tensor power of $\mathscr{K}$, i.e.

$$\mathscr{K}^{(n)} = \mathscr{K}^{\hat{\otimes} n} := \mathrm{conv}\left\{\rho_1 \otimes \cdots \otimes \rho_n,\ \rho_1, \ldots, \rho_n \in \mathscr{K}\right\}.$$

Then, assumption (10.17) is clearly satisfied, in the following way: for any operator M on H, satisfying $0 \leqslant M \leqslant \mathrm{Id}$, and any $0 < \alpha \leqslant \delta < 1$,

$$h_{\mathscr{K}^{(1)}}(M) \leqslant 1-\delta \;\Rightarrow\; \forall\, t \geqslant (1-\delta+\alpha)n,\ h_{\mathscr{K}^{(n)}}\left(M^{(t/n)}\right) \leqslant e^{-n2\alpha^2}.$$

This is a consequence of Hoeffding's inequality, following an argument similar to the one detailed in the previous subsection. And by the result established in the latter, this implies that: for any state ρ on H and any $0 < \epsilon < 1$,

$$F\left(\rho, \mathscr{K}^{(1)}\right) \leqslant e^{-\epsilon} \;\Rightarrow\; F\left(\rho^{\otimes n}, \mathscr{K}^{(n)}\right) \leqslant 2e^{-n\epsilon^2/8}.$$

This is because $F(\rho, \mathscr{K}^{(1)}) \leqslant e^{-\epsilon} \;\Rightarrow\; \|\rho - \mathscr{K}^{(1)}\|_1/2 \geqslant 1 - e^{-\epsilon} \geqslant \epsilon/2$.

In connection with the discussion developed in Sections 10.3 and 10.4, we see that we are actually facing the following interesting open question: how differently do $\mathcal{S}(\mathrm{A}^n{:}\mathrm{B}^n)$ and $\mathcal{S}(\mathrm{A}{:}\mathrm{B})^{\hat{\otimes} n}$ behave, from the (more or less equivalent) points of view of support functions and maximum fidelity functions?

10.6 De Finetti reductions for infinite-dimensional symmetric quantum systems

All quantum de Finetti theorems and reductions require a bound on the dimension of the involved Hilbert spaces. So what can be said about symmetric states on $\mathrm{H}^{\otimes n}$ when H is an infinite-dimensional Hilbert space? What extra assumptions do we need on them in order to be able to reduce their study to that of states in some de Finetti form? The original de Finetti reduction of [49] was especially designed to prove the security of QKD protocols against general attacks. Yet, showing security of continuous variable QKD is also a major issue. This was the motivation behind the infinite-dimensional de Finetti type theorem of [53]. Our ultimate goal here is the same, which we rather try to achieve via a de Finetti reduction under constraints.

10.6.1 Infinite-dimensional post-selection lemma

Let H be an infinite-dimensional Hilbert space, and let $\bar{\mathrm{H}} \subset \mathrm{H}$ be a finite d-dimensional subspace of H. Denote by $\{|j\rangle\}_{j\in\mathbf{N}}$ an orthonormal basis of H, chosen such that $\{|j\rangle\}_{1\leqslant j\leqslant d}$ is an orthonormal basis of $\bar{\mathrm{H}}$. Then, for any $n, k \in \mathbf{N}$, the $(n+k)$-symmetric subspace of $\bar{\mathrm{H}}^{\otimes n} \otimes \mathrm{H}^{\otimes k}$ is defined as

$$\mathrm{Sym}^{n+k}\left(\bar{\mathrm{H}}, \mathrm{H}\right) := \mathrm{Span}\left\{ \sum_{\pi\in\mathfrak{S}(n+k)} |j_{\pi(1)}\rangle \otimes \cdots \otimes |j_{\pi(n+k)}\rangle \ : \ j_1 \leqslant \cdots \leqslant j_{n+k},\ \forall\, 1 \leqslant q \leqslant n,\ j_q \leqslant d \right\}.$$

Note that, denoting by $\bar{\mathrm{H}}_\perp \subset \mathrm{H}$ the orthogonal complement of $\bar{\mathrm{H}}$, i.e. $\mathrm{H} = \bar{\mathrm{H}} \oplus \bar{\mathrm{H}}_\perp$, we have

$$\mathrm{Sym}^{n+k}\left(\bar{\mathrm{H}}, \mathrm{H}\right) \subset \mathrm{V}^{n+k}\left(\bar{\mathrm{H}}, \mathrm{H}\right) := \bigoplus_{\substack{I\subset[n+k]\\ |I|\geqslant n}} \bar{\mathrm{H}}_\perp^{\otimes I^c} \otimes \mathrm{Sym}\left(\bar{\mathrm{H}}^{\otimes I}\right).$$

Lemma 10.6.1. *Let* H *be an infinite-dimensional Hilbert space, and let* $\bar{\mathrm{H}} \subset \mathrm{H}$ *be a finite* d*-dimensional subspace of* H*. Let also* $n, k \in \mathbf{N}$*. Then, any unit vector* $|\theta\rangle \in \mathrm{Sym}^{n+k}\left(\bar{\mathrm{H}}, \mathrm{H}\right)$ *satisfies*

$$|\theta\rangle\langle\theta| \leqslant \left[\sum_{q=0}^{k} \binom{n+k}{q} \binom{n+d-1}{n}^3\right] \sum_{\substack{I\subset[n+k]\\ |I|\geqslant n}} \int_{|x\rangle\in S_{\bar{\mathrm{H}}}} \epsilon(\theta_x)_{\bar{\mathrm{H}}_\perp^{I^c}} \otimes |x\rangle\langle x|_{\bar{\mathrm{H}}}^{\otimes I} \mathrm{d}x,$$

where for all $0 \leqslant q \leqslant k$ *and all unit vector* $|x\rangle \in \bar{\mathrm{H}}$, $\epsilon(\theta_x)_{\bar{\mathrm{H}}_\perp^{k-q}} = \mathrm{Tr}_{\bar{\mathrm{H}}^{n+q}}\left[\left(\mathrm{Id}_{\bar{\mathrm{H}}_\perp}^{\otimes k-q} \otimes |x\rangle\langle x|_{\bar{\mathrm{H}}}^{\otimes n+q}\right)|\theta\rangle\langle\theta|\right]$ *is a sub-normalized state on* $\bar{\mathrm{H}}_\perp^{\otimes k-q}$.

Proof. Since $\mathrm{Sym}^{n+k}\left(\bar{\mathrm{H}}, \mathrm{H}\right) \subset \mathrm{V}^{n+k}\left(\bar{\mathrm{H}}, \mathrm{H}\right)$, any unit vector $|\theta\rangle \in \mathrm{Sym}^{n+k}\left(\bar{\mathrm{H}}, \mathrm{H}\right)$ satisfies

$$\begin{aligned}
|\theta\rangle\langle\theta| =& P_{\mathrm{V}^{n+k}\left(\bar{\mathrm{H}},\mathrm{H}\right)} |\theta\rangle\langle\theta| P^\dagger_{\mathrm{V}^{n+k}\left(\bar{\mathrm{H}},\mathrm{H}\right)} \\
=& \binom{n+d-1}{n}^2 \sum_{\substack{I,J\subset[n+k]\\ |I|,|J|\geqslant n}} \int_{|x\rangle,|y\rangle\in S_{\bar{\mathrm{H}}}} \left(\mathrm{Id}_{\bar{\mathrm{H}}_\perp}^{\otimes I^c} \otimes |x\rangle\langle x|_{\bar{\mathrm{H}}}^{\otimes I}\right) |\theta\rangle\langle\theta| \left(\mathrm{Id}_{\bar{\mathrm{H}}_\perp}^{\otimes J^c} \otimes |y\rangle\langle y|_{\bar{\mathrm{H}}}^{\otimes J}\right)^\dagger \mathrm{d}x\mathrm{d}y.
\end{aligned}$$

Now, by Lemma 10.2.1 (and using the same Caratheodory argument as in the proof of Proposition 10.2.2), we have

$$\begin{aligned}
&\sum_{\substack{I,J\subset[n+k]\\ |I|,|J|\geqslant n}} \int_{|x\rangle,|y\rangle\in S_{\bar{\mathrm{H}}}} \left(\mathrm{Id}_{\bar{\mathrm{H}}_\perp}^{\otimes I^c} \otimes |x\rangle\langle x|_{\bar{\mathrm{H}}}^{\otimes I}\right) |\theta\rangle\langle\theta| \left(\mathrm{Id}_{\bar{\mathrm{H}}_\perp}^{\otimes J^c} \otimes |y\rangle\langle y|_{\bar{\mathrm{H}}}^{\otimes J}\right)^\dagger \mathrm{d}x\mathrm{d}y \\
&\leqslant \left[\sum_{q=0}^{k} \binom{n+k}{q} \binom{n+d-1}{n}\right] \sum_{\substack{I\subset[n+k]\\ |I|\geqslant n}} \int_{|x\rangle\in S_{\bar{\mathrm{H}}}} \left(\mathrm{Id}_{\bar{\mathrm{H}}_\perp}^{\otimes I^c} \otimes |x\rangle\langle x|_{\bar{\mathrm{H}}}^{\otimes I}\right) |\theta\rangle\langle\theta| \left(\mathrm{Id}_{\bar{\mathrm{H}}_\perp}^{\otimes I^c} \otimes |x\rangle\langle x|_{\bar{\mathrm{H}}}^{\otimes I}\right)^\dagger \mathrm{d}x.
\end{aligned}$$

We then just have to notice that, for any $0 \leqslant q \leqslant k$ and any unit vector $|x\rangle \in \bar{\mathrm{H}}$,

$$\left(\mathrm{Id}_{\bar{\mathrm{H}}_\perp}^{\otimes k-q} \otimes |x\rangle\langle x|_{\bar{\mathrm{H}}}^{\otimes n+q}\right)|\theta\rangle\langle\theta|\left(\mathrm{Id}_{\bar{\mathrm{H}}_\perp}^{\otimes k-q} \otimes |x\rangle\langle x|_{\bar{\mathrm{H}}}^{\otimes n+q}\right)^\dagger = \mathrm{Tr}_{\bar{\mathrm{H}}^{n+q}}\left[\left(\mathrm{Id}_{\bar{\mathrm{H}}_\perp}^{\otimes k-q} \otimes |x\rangle\langle x|_{\bar{\mathrm{H}}}^{\otimes n+q}\right)|\theta\rangle\langle\theta|\right] \otimes |x\rangle\langle x|_{\bar{\mathrm{H}}}^{\otimes n+q},$$

in order to actually get the advertised result. □

10.6.2 An application

One application of Lemma 10.6.1 is to the security analysis of quantum cryptographic schemes, when there is no a priori bound on the dimension of the information carriers. This problem was originally investigated in [53] via a de Finetti theorem specifically designed for it. It was shown there that, under experimentally verifiable conditions, it is possible to ensure the security of quantum key distribution (QKD) protocols with continuous variables against general attacks. We show here that similar conclusions can be reached using the de Finetti reduction of Lemma 10.6.1.

We look at things from the exact same point of view as the one adopted in [53]. Let H be an infinite-dimensional Hilbert space and let X, Y be two canonical operators on H. Then, denote by $\Lambda = X^2 + Y^2$ the corresponding Hamiltonian, fix $\lambda_0 > 0$, and define

$$\bar{\mathrm{H}} := \{|\theta\rangle \in \mathrm{H} \ : \ \Lambda|\theta\rangle = \lambda|\theta\rangle, \ \lambda \leqslant \lambda_0\},$$

finite-dimensional subspace of H spanned by the eigenvectors of Λ with associated eigenvalue at most λ_0.

Let $n, k \in \mathbf{N}$, with $n \geqslant 2k$, and let $\rho^{(n+2k)}$ be a $(n+2k)$-symmetric state on $\mathrm{H}^{\otimes n+2k}$. Next, define the two events $\mathscr{A}$ and $\mathscr{B}$ as

$$\begin{aligned}
\mathscr{A} &= \text{“ } \forall \ 1 \leqslant q \leqslant k, \ \mathrm{Tr}\left(\Lambda \rho_q^{(1)}\right) \leqslant \lambda_0 \text{ ”} \\
\mathscr{B} &= \text{“ } \exists \ |\theta^{(n+k)}\rangle \in \mathrm{Sym}^{n+k}\left(\bar{\mathrm{H}} \otimes \bar{\mathrm{H}}', \mathrm{H} \otimes \mathrm{H}'\right) : \ \rho^{(n+k)} = \mathrm{Tr}_{\mathrm{H}'^{n+k}} |\theta^{(n+k)}\rangle\langle\theta^{(n+k)}| \text{ ”},
\end{aligned}$$

where for all $1 \leqslant q \leqslant k$, $\rho_q^{(1)} = \mathrm{Tr}_{\mathrm{H}^{n+2k} \setminus \mathrm{H}_q} \rho^{(n+2k)}$, and $\rho^{(n+k)} = \mathrm{Tr}_{\mathrm{H}_{k+1} \cdots \mathrm{H}_{2k}} \rho^{(n+2k)}$. We know by Lemma III.3 in [53] that there exist universal constants $C_0, c > 0$ such that, whenever $\lambda_0 \geqslant C_0 \log(n/k)$, we have

$$\mathbf{P}\left(\mathscr{A} \wedge \neg \mathscr{B}\right) \leqslant e^{-ck^3/n^2}.$$

In words, this means the following. Fix a threshold $\lambda_0 \geqslant C_0 \log(n/k)$, and assume that when measuring the energy Λ on the k first subsystems of $\rho^{(n+2k)}$, only values below λ_0 are obtained. Then, with probability greater than $1 - e^{-ck^3/n^2}$, the remaining $n+k$ subsystems of $\rho^{(n+2k)}$ have a purification which is the symmetrization of a state with more than n subsystems supported in $\bar{\mathrm{H}} \otimes \bar{\mathrm{H}}'$.

Now, let $\widetilde{\rho}$ be a state on $\mathrm{H}^{\otimes n+k}$ such that $\widetilde{\rho} = \mathrm{Tr}_{\mathrm{H}'^{n+k}} |\widetilde{\theta}\rangle\langle\widetilde{\theta}|$ for some $|\widetilde{\theta}\rangle \in \mathrm{Sym}^{n+k}\left(\bar{\mathrm{H}} \otimes \bar{\mathrm{H}}', \mathrm{H} \otimes \mathrm{H}'\right)$. Denoting by d the dimension of $\bar{\mathrm{H}}$, we have by Lemma 10.6.1 that $|\widetilde{\theta}\rangle$ satisfies

$$|\widetilde{\theta}\rangle\langle\widetilde{\theta}| \leqslant \left[\sum_{q=0}^{k} \binom{n+k}{q} \binom{n+d^2-1}{n}^3\right] \sum_{\substack{I \subset [n+k] \\ |I| \geqslant n}} \int_{|x\rangle \in S_{\bar{\mathrm{H}} \otimes \bar{\mathrm{H}}'}} \epsilon(\widetilde{\theta}_x)_{(\mathrm{HH}')^{I^c}} \otimes |x\rangle\langle x|_{\bar{\mathrm{H}}\bar{\mathrm{H}}'}^{\otimes I} \mathrm{d}x.$$

And hence, after partial tracing over $\mathrm{H}'^{\otimes n+k}$, we finally get

$$\widetilde{\rho} \leqslant \left[\sum_{q=0}^{k} \binom{n+k}{q} \binom{n+d^2-1}{n}^3\right] \sum_{\substack{I \subset [n+k] \\ |I| \geqslant n}} \int_{|x\rangle \in S_{\bar{\mathrm{H}} \otimes \bar{\mathrm{H}}'}} \varepsilon(\widetilde{\theta}_x)_{\mathrm{H}^{I^c}} \otimes \sigma(x)_{\bar{\mathrm{H}}}^{\otimes I} \mathrm{d}x,$$

where for all $0 \leqslant q \leqslant k$ and all unit vector $|x\rangle \in \bar{\mathrm{H}} \otimes \bar{\mathrm{H}}'$, $\sigma(x)_{\bar{\mathrm{H}}} = \mathrm{Tr}_{\bar{\mathrm{H}}'} |x\rangle\langle x|_{\bar{\mathrm{H}}\bar{\mathrm{H}}'}$ is the reduced state of $|x\rangle\langle x|$ on $\bar{\mathrm{H}}$, and $\varepsilon(\widetilde{\theta}_x)_{\mathrm{H}^{k-q}} = \mathrm{Tr}_{\mathrm{H}'^{k-q}} \epsilon(\widetilde{\theta}_x)_{(\mathrm{HH}')^{k-q}}$ is the reduced sub-normalized state of $\epsilon(\widetilde{\theta}_x)$ on $\mathrm{H}^{\otimes k-q}$.

Putting everything together, we can eventually get the following: Let $n \in \mathbf{N}$ and $k = \lfloor n^\alpha \rfloor$ for a given α fulfilling $2/3 < \alpha < 1$. Let also λ_0 be a threshold such that on the one hand $\lambda_0 \geqslant C_0 \log n$, where $C_0 > 0$ is a universal constant, and on the other hand $d \leqslant n^\beta$ for a given β fulfilling $0 < \beta < 1/2$. Suppose next that $\rho^{(n+2k)}$ is a $(n+2k)$-symmetric state on $\mathrm{H}^{\otimes n+2k}$ such that event $\mathscr{A}$ holds. Then, with probability greater than $1 - e^{-cn^{3\alpha-2}}$, where $c > 0$ is a universal constant, the reduced state $\rho^{(n+k)}$ of $\rho^{(n+2k)}$ on $\mathrm{H}^{\otimes n+k}$ satisfies

$$\rho^{(n+k)} \leqslant (Cn)^{n^\alpha + n^{2\beta}} \sum_{\substack{I \subset [n+k] \\ |I| \geqslant n}} \int_{\sigma_{\bar{\mathrm{H}}} \in \mathscr{D}(\bar{\mathrm{H}})} \varepsilon(\rho, \sigma)_{\mathrm{H}^{I^c}} \otimes \sigma_{\bar{\mathrm{H}}}^{\otimes I} \mathrm{d}\mu(\sigma_{\bar{\mathrm{H}}}),$$

where $C > 0$ is a universal constant, μ is a probability measure on the set of states on $\bar{\mathrm{H}}$, and for each $0 \leqslant q \leqslant k$ and each state σ on $\bar{\mathrm{H}}$, $\varepsilon(\rho, \sigma)_{\mathrm{H}^{k-q}}$ is a sub-normalized state on $\mathrm{H}^{\otimes k-q}$.

Now, let $\mathcal{N} : \mathscr{L}(\mathrm{H}) \to \mathscr{L}(\mathrm{K})$ be a quantum channel, and assume that there exists some $0 < \delta < 1$ such that

$$\sup\left\{\|\mathcal{N}(\sigma)\|_1 \ : \ \sigma \in \mathscr{D}(\bar{\mathrm{H}})\right\} \leqslant \delta.$$

This implies in particular that, for any $0 \leqslant q \leqslant k$, and any states ε on $\mathrm{H}^{\otimes k-q}$, σ on $\bar{\mathrm{H}}$, we have

$$\left\|\mathcal{N}^{\otimes n+k}\left(\varepsilon \otimes \sigma^{\otimes n+q}\right)\right\|_1 = \left\|\mathcal{N}^{\otimes k-q}\left(\varepsilon\right)\right\|_1 \left\|\mathcal{N}(\sigma)\right\|_1^{n+q} \leqslant \delta^{n+q}.$$

And subsequently, by what precedes, for any state $\rho^{(n+2k)}$ on $\mathrm{H}^{\otimes n+2k}$ such that event $\mathscr{A}$ holds, denoting by $\rho^{(n+k)}$ its reduced state on $\mathrm{H}^{\otimes n+k}$, it holds with probability greater than $1 - e^{-cn^{3\alpha-2}}$ that

$$\begin{aligned}
\left\|\mathcal{N}^{\otimes n+k}\left(\rho^{(n+k)}\right)\right\|_1 &\leqslant (Cn)^{n^\alpha+n^{2\beta}} \sum_{q=0}^{k} \binom{n+k}{q} \sup\left\{\left\|\mathcal{N}^{\otimes n+k}\left(\varepsilon \otimes \sigma^{\otimes n+q}\right)\right\|_1 \ : \ \varepsilon \in \mathscr{D}(\mathrm{H}^{\otimes k-q}),\ \sigma \in \mathscr{D}(\bar{\mathrm{H}})\right\} \\
&\leqslant (Cn)^{n^\alpha+n^{2\beta}} \sum_{q=0}^{k} \binom{n+k}{q} \delta^{n+q} \\
&\leqslant (C'n)^{n^\alpha+n^{2\beta}}\, \delta^n,
\end{aligned}$$

where $C' > 0$ is a universal constant. By the way α, β have been chosen, this means that, for any $\tilde{\delta} > \delta$ and $n \geqslant n_{\tilde{\delta}}$, we have with high probability

$$\sup\left\{\left\|\mathcal{N}^{\otimes n+k}\left(\rho^{(n+k)}\right)\right\|_1 \ : \ \rho^{(n+k)} = \mathrm{Tr}_{\mathrm{H}^{\otimes k}} \rho^{(n+2k)} \text{ with } \rho^{(n+2k)} \in \mathscr{D}(\mathrm{H}^{\otimes n+2k}) \text{ s.t. } \mathscr{A} \text{ holds}\right\} \leqslant \tilde{\delta}^n.$$

10.7 Conclusion and outlook

We have reviewed (and given a new proof of) the constrained de Finetti reduction of [67]. We have demonstrated its adaptability to various situations where one would like to impart a known constraint satisfied by a permutation symmetric state onto the i.i.d. states occurring in the operator with which to compare it. We have seen that our technique works especially well in the case of linear constraints (see [67] and Chapter 11 for two developed such applications).

We have then spent considerable effort on a particularly interesting convex constraint, separability. Apart from the obvious relevance to entanglement theory, the constrained de Finetti reduction provides a very natural framework in which to derive bounds on the success probability of parallel repetitions of tests, and has immediate applications in the parallel repetition of QMA(2), quantum Merlin-Arthur interactive proof systems with two unentangled provers (see [94] for further details). Conversely, we showed that certain progress in entanglement theory (on the conjectured faithfulness properties of the CEMI entanglement measure for instance) would imply even stronger, dimension-independent bounds, which would show in particular that the soundness gap of QMA(2) can be amplified exponentially by parallel repetition, without any other devices. It is curious to see that the progress on questions like this can depend on the properties of a simple, but little-understood entanglement measure such as CEMI, and we would like to recommend its study to the reader's attention. Indeed, it seems to be the best candidate so far for a *magical*, or even *supercalifragilistic* entanglement measure. The latter is defined as one which has the post-selection property with respect to an initial product state and measurement on a separate subsystem (cf. Lemma 10.4.1), is super-additive, and satisfies a universal faithfulness bound with respect to the trace-norm distance (cf. Conjecture (10.14)).

We have also presented a more abstract framework of convex constraints, that allows us to demonstrate in greater generality the interplay between the multiplicative behaviour of (*i*) the support function and (*ii*) the maximum fidelity function. The way (*i*) is derived from (*ii*) is via our de Finetti reduction with fidelity weight in the upper bounding operator. And (*ii*) is obtained from (*i*) by constructing a test whose failure probability decays exponentially under parallel repetition.

Finally, seeing that the de Finetti reductions had been so far always limited by the finite dimensionality of the system involved, we have made first steps towards an extension of the main technical tool to infinite-dimensional systems under suitable constraints. It remains to be seen how widely it or a variation can be applied to quantum cryptography in continuous variable systems [53, 41], or similar problems.

Chapter 11

Parallel repetition and concentration for (sub-)no-signalling games

Based on "Parallel repetition and concentration for (sub-)no-signalling games via a flexible constrained de Finetti reduction", in collaboration with A. Winter [129].

We use a recently discovered constrained de Finetti reduction (aka "Post-Selection Lemma") to study the parallel repetition of multi-player non-local games under no-signalling strategies. Since the technique allows us to reduce general strategies to independent plays, we obtain parallel repetition (corresponding to winning all rounds) in the same way as exponential concentration of the probability to win a fraction larger than the value of the game.

Our proof technique leads us naturally to a relaxation of no-signalling (NS) strategies, which we dub *sub-no-signalling (SNOS)*. While for two players the two concepts coincide, they differ for three or more players. Our results are most complete and satisfying for arbitrary number of sub-no-signalling players, where we get universal parallel repetition and concentration for any game, while the no-signalling case is obtained as a corollary, but only for games with "full support".

11.1 Non-local games and no-signalling strategies

A multi-player non-local game is played between cooperating but non-communicating players. Each player receives an input from some input alphabet and has to produce an output in some output alphabet. The common goal of the players is to satisfy some pre-defined predicate on their inputs and outputs. For that, they may agree on a strategy before the game starts, but are then not allowed to communicate anymore. Such games are especially relevant in theoretical physics in the context of the foundations of quantum mechanics and quantum information, and in computer science where they arise in multi-prover interactive proof systems. Indeed, they may provide an intuitive and quantitative understanding of the role played by various degrees of correlations in global systems which are composed of several local subsystems. These games also arise in complexity theory, under the formulation of multi-provers with some shared resources producing a protocol that should convince a referee, or in cryptography as attacks from malicious parties having a more or less restricted physical power.

The *value* of a game is the maximum winning probability of the players, over all allowed joint strategies, using possibly some prescribed correlation resource such as shared randomness, quantum entanglement or no-signalling correlations. It has been a subject of considerable study how the availability of different resources affects the values of certain games [26, 54, 172, 149, 21].

In this context, a natural question is how the value of a game behaves when n independent instances of the game are played simultaneously, i.e. each player gets n independent inputs and has to provide n outputs such that each game instance is won (or a large fraction of them). This is the parallel repetition problem (which obviously has the exact same flavor as the question of how support functions behave under tensoring, that was at the heart of Chapter 10). Playing independently the optimal single-game strategy on all n game instances will result in an exponentially decreasing winning probability. But although that was found paradoxical at first, this is in general not optimal [74, 75]. For classical two-player games, Raz [151], later simplified and improved by Holenstein [104], established the first general parallel repetition theorem, showing that the value of n repetitions

decreases exponentially for every game. Holenstein [104] also proved an analogous parallel repetition theorem for the no-signalling value of general two-player games. Only recently, parallel repetition theorems were proved for the entangled value of two-player games: for general games, nothing better than a polynomial decay result is known up to now [123], while exponential decay results have been established in several special cases (perfect parallel repetition for XOR games [55], exponential decrease under parallel repetition for unique games [122], projection games [61], free games [43, 116]).

Multi-player games have received little attention until recently. And apart from the result in [52] (containing both classical and quantum statements), only in the no-signalling setting [40, 7]. The present work has the same focus on multiple no-signalling players, albeit we will find that the theory becomes much more satisfying for *sub-no-signalling* players.

Specifically, we will consider here ℓ-player games G with input alphabets $\mathcal{X}_1, \ldots, \mathcal{X}_\ell$ and output alphabets $\mathcal{A}_1, \ldots, \mathcal{A}_\ell$. By way of notation,

$$\underline{\mathcal{X}} := \bigtimes_{i=1}^{\ell} \mathcal{X}_i \text{ and } \underline{\mathcal{A}} := \bigtimes_{i=1}^{\ell} \mathcal{A}_i.$$

Furthermore, for any subset $I \subset [\ell]$ of indices,

$$\mathcal{X}_I := \bigtimes_{i \in I} \mathcal{X}_i \text{ and } \mathcal{A}_I := \bigtimes_{i \in I} \mathcal{A}_i.$$

For any $I, J \subset [\ell]$, given T a probability distribution on $\mathcal{X}_I$, resp. P a conditional probability distribution on $\mathcal{A}_J|\mathcal{X}_I$, we may denote it by $T_{\mathcal{X}_I}$, resp. $P_{\mathcal{A}_J|\mathcal{X}_I}$, when confusion on the considered alphabets is at risk. We may also sometimes use the abbreviation "p.d." for "probability distribution".

From now on, we will be interested in making minimal a priori assumptions on how powerful the ℓ players may be. This will naturally lead us to considering that their common strategy to win the game G could be any no-signalling (or even sub-no-signalling) strategy, which we define now.

Definition 11.1.1. *The sets of* no-signalling *and* sub-no-signalling *correlations, denoted respectively* $NS(\underline{\mathcal{A}}|\underline{\mathcal{X}})$ *and* $SNOS(\underline{\mathcal{A}}|\underline{\mathcal{X}})$*, consist of non-negative densities* $P(\underline{a}|\underline{x}) \geqslant 0$ *defined as follows:*

$$P \in NS(\underline{\mathcal{A}}|\underline{\mathcal{X}}) :\Leftrightarrow \forall\ I \subsetneq [\ell],\ \exists\ Q(\cdot|x_I)\ p.d.\text{'s on } \mathcal{A}_I\ s.t.\ \forall\ \underline{x}, a_I,\ P(a_I|\underline{x}) = Q(a_I|x_I), \tag{11.1}$$

$$P \in SNOS(\underline{\mathcal{A}}|\underline{\mathcal{X}}) :\Leftrightarrow \forall\ I \subsetneq [\ell],\ \exists\ Q(\cdot|x_I)\ p.d.\text{'s on } \mathcal{A}_I\ s.t.\ \forall\ \underline{x}, a_I,\ P(a_I|\underline{x}) \leqslant Q(a_I|x_I). \tag{11.2}$$

Here, $P(a_I|\underline{x})$ *denotes the marginal density,*

$$P(a_I|\underline{x}) = \sum_{a_{I^c} \in \mathcal{A}_{I^c}} P(\underline{a} = a_I a_{I^c}|\underline{x}).$$

Remark 11.1.2. *Note that under this definition,* $NS(\underline{\mathcal{A}}|\underline{\mathcal{X}}) \subset SNOS(\underline{\mathcal{A}}|\underline{\mathcal{X}})$*, but the latter is a strictly larger set (e.g. it always contains the all-zero density). Furthermore,* $P \in NS(\underline{\mathcal{A}}|\underline{\mathcal{X}})$ *iff* $P \in SNOS(\underline{\mathcal{A}}|\underline{\mathcal{X}})$ *and* P *is* normalized *in the sense that for all* $\underline{x} \in \underline{\mathcal{X}}$, $\sum_{\underline{a}} P(\underline{a}|\underline{x}) = 1$*. Indeed, NS consists of conditional probability distributions, while SNOS allows, given each input, a total "probability" of less than or equal to* 1*.*

Also, it can be shown that in equation (11.1), only sets of the form $I = [\ell] \setminus i$ *need to be considered. This is because the no-signalling conditions take the form of equations and this subset spans the set of all equations required (cf. [92], Lemma 2.7). The analogous statement for sub-no-signalling is not known and likely false. Nevertheless, one might in other contexts consider to relax the conditions of equation (11.2) to hold only for a selected family of subsets* $I \subset [\ell]$*.*

An ℓ-player game G is characterized by a probability distribution $T(\underline{x})$ on the queries $\underline{\mathcal{X}}$, and a binary predicate $V(\underline{a}, \underline{x}) \in \{0, 1\}$ on the answers and queries $\underline{\mathcal{A}} \times \underline{\mathcal{X}}$, as illustrated in Figure 11.1. The no-signalling, resp. sub-no-signalling, value of the game, denoted $\omega_{\mathrm{NS}}(G)$, resp. $\omega_{\mathrm{SNOS}}(G)$, is the maximum of the winning probability

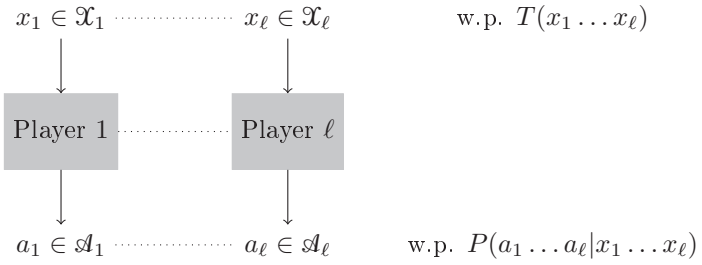
$$\mathbf{P}\,(\mathrm{win}) = \mathbf{E}\, V(\underline{A}, \underline{X}) = \sum_{\underline{a}, \underline{x}} T(\underline{x}) V(\underline{a}, \underline{x}) P(\underline{a}|\underline{x})$$

over all $P \in \mathrm{NS}(\underline{\mathcal{A}}|\underline{\mathcal{X}})$, resp. $P \in \mathrm{SNOS}(\underline{\mathcal{A}}|\underline{\mathcal{X}})$, where the distribution of $\underline{X} = X_1 \dots X_\ell$ and $\underline{A} = A_1 \dots A_\ell$ is as expected,

$$\forall\ \underline{x}, \underline{a},\ \mathbf{P}\left(\underline{X} = \underline{x},\ \underline{A} = \underline{a}\right) = T(\underline{x})P(\underline{a}|\underline{x}).$$

In words, the (sub-)no-signalling value of a game is the maximal probability of winning it when no limitation is assumed on the power of the players, apart from the fact that they cannot signal information instantaneously from one another. In the sub-no-signalling case, constraints are relaxed even more: players are not forced to always produce an output, and it is only required that their strategy "looks as if it were no-signalling" (even though they may have "hidden" in their abstentions the fact that it is signalling). In Section 11.5 we briefly discuss other kinds of restrictions that one may put on the players' physical power, such as shared randomness or shared quantum entanglement only.

Figure 11.1: An ℓ-player non local game



The players win iff $V(a_1 \dots a_\ell, x_1 \dots x_\ell) = 1$

11.1.1 Two-player SNOS $\equiv$ NS

Not surprisingly, the no-signalling and sub-no-signalling values of games are related. We start by showing that for any two-player game G, they are identical, i.e. $\omega_{\mathrm{NS}}(G) = \omega_{\mathrm{SNOS}}(G)$. As NS $\subset$ SNOS, the inequality "$\leqslant$" is evident, and we only need to prove the opposite inequality "$\geqslant$". This follows from the following structural lemma.

Lemma 11.1.3 (cf. [115])**.** *Let $P \in SNOS(\mathcal{A} \times \mathcal{B}|\mathcal{X} \times \mathcal{Y})$ be a two-player sub-no-signalling correlation. Then there exists a no-signalling correlation $P' \in NS(\mathcal{A} \times \mathcal{B}|\mathcal{X} \times \mathcal{Y})$ with $P \leqslant P'$ pointwise, i.e. $P(ab|xy) \leqslant P'(ab|xy)$ for all a, b, x, y.*

Proof. If P is normalized, i.e. if for all x, y, $\sum_{ab} P(ab|xy) = 1$, there is nothing to prove because P is already no-signalling.

Otherwise, there exist x, y with weight $\sum_{ab} P(ab|xy) =: w < 1$. By sub-no-signalling assumption, we have distributions $Q(a|x)$ and $Q(b|y)$ dominating the marginals:

$$\forall\ a, b,\ P(a|xy) \leqslant Q(a|x),\ P(b|xy) \leqslant Q(b|y).$$

As the total weight of both marginals of $P(\cdot|xy)$ is $w < 1$, we can find a and b such that

$$P(a|xy) < Q(a|x), \quad P(b|xy) < Q(b|y),$$

so we can increase $P(ab|xy)$ by some $\epsilon > 0$ to $P'(ab|xy) = P(ab|xy) + \epsilon$ and still satisfy the sub-no-signalling conditions. By choosing ϵ maximally so, we can reduce the total number of strict inequality signs in the SNOS conditions. Iterating this procedure we arrive at a sub-no-signalling correlation P' with all inequalities met with equality, i.e. a no-signalling correlation.

Another presentation of this argument appeals to compactness. Consider the following set of correlations:

$$\mathrm{X}_{P,Q} := \left\{P' \ :\ \forall\ a, b, x, y,\ P'(ab|xy) \geqslant P(ab|xy),\ P'(a|x) \leqslant Q(a|x),\ P'(b|y) \leqslant Q(b|y)\right\}.$$

$\mathrm{X}_{P,Q}$ being compact and $P' \mapsto \sum_{xy}\sum_{ab} P'(ab|xy)$ being continuous,

$$\sup\left\{\sum_{xy}\sum_{ab} P'(ab|xy) \;:\; P' \in \mathrm{X}_{P,Q}\right\}$$

is actually attained. If it were less than $|\mathcal{X}\times\mathcal{Y}|$, we could use the procedure above to increase the objective function, contradicting that it is a maximum. □

Note that the "bumping up" procedure described above, in order to transform any two-player sub-no-signalling strategy into a no-signalling one dominating it pointwise, may fail for more players. The two-player case is indeed special, due to non-overlapping of the two SNOS or NS constraints. However, already in the case of three players, even just the three inequalities $P_{\mathcal{A}_i\mathcal{A}_j|\mathcal{X}_i\mathcal{X}_j\mathcal{X}_k} \leqslant Q_{\mathcal{A}_i\mathcal{A}_j|\mathcal{X}_i\mathcal{X}_j}$ may be impossible to bring simultaneously to equalities by pointwise increment.

11.1.2 Multi-player SNOS vs NS

Clearly, $\omega_{\mathrm{NS}}(G) \leqslant \omega_{\mathrm{SNOS}}(G)$ for every game, and there are examples of games (with game distribution T having strictly smaller than full support) where $\omega_{\mathrm{NS}}(G) < 1$ but $\omega_{\mathrm{SNOS}}(G) = 1$, for instance the *anticorrelation game* (and likewise a number of games where frustration prohibits extension of a winning sub-no-signalling strategy to a normalized, no-signalling one).

Example 11.1.4. (cf. [7], Appendix A) Consider the three-player *anti-correlation game* A_3, which has binary input and output for all players and game distribution T supported on $\{0,1\}^3 \setminus \{111\}$, i.e. 111 does not occur as a triple of questions. The winning predicate is that if any two inputs are 1, say $x_i = x_j = 1$, then the corresponding outputs must be different, $a_i \neq a_j$. While if there are zero or only a single 1 amongst the inputs, outputs may be arbitrary.

It is straightforward to verify that the following correlation is in SNOS $\left(\{0,1\}^3|\{0,1\}^3\right)$ and wins the game with certainty:

$$P(a_1a_2a_3|x_1x_2x_3) = \begin{cases} 0 & \text{if } x_1x_2x_3 = 111, \\ 1/8 & \text{if } \exists\ 1 \leqslant i \neq j \leqslant 3 :\ x_i = x_j = 0, \\ \delta_{a_i,1-a_j}/4 & \text{if } \exists\ 1 \leqslant i \neq j \leqslant 3 :\ x_i = x_j = 1 \text{ and } x_1x_2x_3 \neq 111. \end{cases}$$

So $\omega_{\mathrm{SNOS}}(A_3) = 1$. On the other hand, for, say, T uniform on $\{011, 101, 110\}$, one can check by elementary means that $\omega_{\mathrm{NS}}(A_3) = 2/3$.

However, for a game distribution T having full support, a simple reasoning shows that $\omega_{\mathrm{NS}}(G) < 1$ implies $\omega_{\mathrm{SNOS}}(G) < 1$. Indeed, we show the contrapositive, assuming that $\omega_{\mathrm{SNOS}}(G) = 1$. Because of the full support of T, this implies that for the optimal sub-no-signalling strategy P and every $\underline{x}$,

$$1 = \sum_{\underline{a}} V(\underline{a},\underline{x})P(\underline{a}|\underline{x}) \leqslant \sum_{\underline{a}} P(\underline{a}|\underline{x}) \leqslant 1,$$

hence equality (i.e. normalization) holds for all $\underline{x}$. Thus, P is really a no-signalling correlation and so $\omega_{\mathrm{NS}}(G) = 1$. In fact, we can show something stronger, namely the following quantitative relationship.

Lemma 11.1.5. *Consider a game distribution T with full support on $\underline{\mathcal{X}}$. Then there exists $\Gamma = \Gamma(T) \geqslant 0$, which only depends on T, such that for every game G with query distribution T,*

$$\omega_{SNOS}(G) \geqslant 1-\epsilon \;\Rightarrow\; \omega_{NS}(G) \geqslant 1-(\Gamma+1)\epsilon.$$

The definition of Γ can be taken from [40] or [7].

Proof. Take an optimal strategy $P \in \mathrm{SNOS}(\underline{\mathcal{A}}|\underline{\mathcal{X}})$, so that $P(a_I|\underline{x}) \leqslant Q(a_I|x_I)$ for all I, a_I, $\underline{x}$. Then,

$$\sum_{\underline{a},\underline{x}} T(\underline{x})P(\underline{a}|\underline{x}) \geqslant \sum_{\underline{a},\underline{x}} T(\underline{x})V(\underline{a},\underline{x})P(\underline{a}|\underline{x}) = \omega_{\mathrm{SNOS}}(G) \geqslant 1-\epsilon.$$

And so we get, for all I,

$$\left\|T_{\underline{\mathcal{X}}}P_{\mathcal{A}_I|\underline{\mathcal{X}}} - T_{\underline{\mathcal{X}}}Q_{\mathcal{A}_I|\mathcal{X}_I}\right\|_1 = \sum_{a_I,\underline{x}} T(\underline{x})\big(Q(a_I|x_I) - P(a_I|\underline{x})\big) \leqslant \epsilon,$$

because the difference term in the sum is non-negative.

Now simply "bump up" the sub-normalized probability distribution $P_{\underline{\mathscr{A}}|\underline{\mathfrak{X}}}$ to a properly normalized conditional probability distribution $P'_{\underline{\mathscr{A}}|\underline{\mathfrak{X}}}$, adding at most an averaged weight over $T_{\underline{\mathfrak{X}}}$ of ϵ, and hence, for all I,

$$\frac{1}{2}\left\|T_{\underline{\mathfrak{X}}}P'_{\mathscr{A}_I|\mathfrak{X}} - T_{\underline{\mathfrak{X}}}Q_{\mathscr{A}_I|\mathfrak{X}_I}\right\|_1 \leqslant \epsilon.$$

At this point we can invoke the stability of linear programmes, used in [40] and [7] to conclude that there is $\Gamma = \Gamma(T) \geqslant 0$ such that there is a no-signalling correlation $P''_{\underline{\mathscr{A}}|\underline{\mathfrak{X}}} \in \mathrm{NS}(\underline{\mathscr{A}}|\underline{\mathfrak{X}})$ with

$$\frac{1}{2}\left\|T_{\underline{\mathfrak{X}}}P''_{\underline{\mathscr{A}}|\underline{\mathfrak{X}}} - T_{\underline{\mathfrak{X}}}P'_{\underline{\mathscr{A}}|\underline{\mathfrak{X}}}\right\|_1 \leqslant \Gamma\epsilon.$$

This gives

$$\begin{aligned}
\omega_{\mathrm{NS}}(G) &\geqslant \sum_{\underline{a},\underline{x}} T(\underline{x})V(\underline{a},\underline{x})P''(\underline{a}|\underline{x}) \\
&\geqslant \sum_{\underline{a},\underline{x}} T(\underline{x})V(\underline{a},\underline{x})P'(\underline{a}|\underline{x}) - \Gamma\epsilon \\
&\geqslant \sum_{\underline{a},\underline{x}} T(\underline{x})V(\underline{a},\underline{x})P(\underline{a}|\underline{x}) - \Gamma\epsilon \\
&\geqslant 1 - (\Gamma+1)\epsilon,
\end{aligned}$$

where we have used the total variational bound on $P'' - P'$, the fact that P' dominates P and the assumption on the probability of winning G when played P. □

The rest of the chapter is structured as follows: In Section 11.2 we introduce parallel repetition of games, and state our main results, which improve upon, and partly clarify, earlier findings by Holenstein [104], Buhrman *et al.* [40] and Arnon-Friedman *et al.* [7]. In Section 11.3, we present the main technical tool, one of the constrained de Finetti reductions from Chapter 10, adapted to our present needs, followed by the proofs of the main theorems and corollaries in Section 11.4. We conclude in Section 11.5.

11.2 Parallel repetition: definitions and main results

Given an ℓ-player game G, with probability distribution $T(\underline{x})$ on $\underline{\mathfrak{X}}$ and binary predicate $V(\underline{a},\underline{x}) \in \{0,1\}$ on $\underline{\mathscr{A}} \times \underline{\mathfrak{X}}$, we are interested in playing the same game n times independently in parallel, and in looking at the probability of winning all n or a subset of t of them. The reader is referred to Chapter 10 for entirely analogous considerations, but in the setting of support functions of convex sets of states instead of multi-player non-local games.

Formally, the *n-fold parallel repetition of G* is the ℓ-player game G^n having the product probability distribution on $\underline{\mathfrak{X}}^n$

$$T^{\otimes n}(\underline{x}^n) = T\big(\underline{x}^{(1)}\big)\cdots T\big(\underline{x}^{(n)}\big),$$

and the product binary predicate on $\underline{\mathscr{A}}^n \times \underline{\mathfrak{X}}^n$

$$V^{\otimes n}(\underline{a}^n,\underline{x}^n) = V\big(\underline{a}^{(1)},\underline{x}^{(1)}\big)\cdots V\big(\underline{a}^{(n)},\underline{x}^{(n)}\big) \in \{0,1\}.$$

The no-signalling, resp. sub-no-signalling, value of this n-fold parallel repetition game, denoted $\omega_{\mathrm{NS}}(G^n)$, resp. $\omega_{\mathrm{SNOS}}(G^n)$, is thus the maximum of the winning probability

$$\mathbf{P}\,(\mathrm{win}) = \sum_{\underline{a}^n,\underline{x}^n} T^{\otimes n}(\underline{x}^n)V^{\otimes n}(\underline{a}^n,\underline{x}^n)P(\underline{a}^n|\underline{x}^n)$$

over all $P \in \mathrm{NS}(\underline{\mathscr{A}}^n|\underline{\mathfrak{X}}^n)$, resp. $P \in \mathrm{SNOS}(\underline{\mathscr{A}}^n|\underline{\mathfrak{X}}^n)$.

In words, the players win G^n if they win all n instances of G played in parallel. So we obviously always have (for the allowed set of strategies being $X \in \{\mathrm{NS},\mathrm{SNOS}\}$)

$$\big(\omega_X(G)\big)^n \leqslant \omega_X(G^n) \leqslant \omega_X(G). \tag{11.3}$$

However, in the case where $\omega_X(G) < 1$, the gap between the lower and upper bounds in equation (11.3) grows exponentially with n, making equation (11.3) very little informative. The parallel repetition problem is thus the following: If none of the players' allowed strategies can make them win 1 instance of G with probability 1, does it necessarily imply that they have an exponentially decaying probability of winning n of them at the same time? And if so at which rate?

More generally, we can study the game $G^{t/n}$, whose winning predicate is defined as winning any t (or more) out of n repetitions [40], i.e.

$$V^{t/n}(\underline{a}^n, \underline{x}^n) := \left\{ \sum_{i=1}^{n} V\big(\underline{a}^{(i)}, \underline{x}^{(i)}\big) \geqslant t \right\} = \begin{cases} 1 & \text{if } \sum_{i=1}^{n} V\big(\underline{a}^{(i)}, \underline{x}^{(i)}\big) \geqslant t, \\ 0 & \text{otherwise.} \end{cases}$$

Note that, with our notation, $G^n = G^{n/n}$.

The main results of the present chapter are gathered below, where we set $C_\ell := 2^{\ell+1} - 3$.

Theorem 11.2.1 (Parallel repetition for the sub-no-signalling value of ℓ-player games)**.** *Let G be an ℓ-player game such that $\omega_{SNOS}(G) \leqslant 1 - \delta$ for some $0 < \delta < 1$. Then, for any $n \in \mathbf{N}$, and any $t \geqslant (1 - \delta + \alpha)n$ for some $0 < \alpha \leqslant \delta$, we have*

$$\omega_{SNOS}(G^n) \leqslant \left(1 - \frac{\delta^2}{5C_\ell^2}\right)^n,$$
$$\omega_{SNOS}(G^{t/n}) \leqslant \exp\left(-n\frac{\alpha^2}{5C_\ell^2}\right).$$

As immediate consequences or refinements of Theorem 11.2.1, we can get parallel repetition results for the no-signalling value of multiplayer games in some particular instances.

Corollary 11.2.2 (Parallel repetition for the no-signalling value of full support ℓ-player games)**.** *Let G be an ℓ-player game whose distribution T has full support, and such that $\omega_{NS}(G) \leqslant 1 - \delta$ for some $0 < \delta < 1$. Then, for any $n \in \mathbf{N}$, and any $t \geqslant (1 - \delta + \alpha)n$ for some $0 < \alpha \leqslant \delta$, we have*

$$\omega_{NS}(G^n) \leqslant \left(1 - \frac{\delta^2}{5C_\ell^2(\Gamma+1)^2}\right)^n,$$
$$\omega_{NS}(G^{t/n}) \leqslant \exp\left(-n\frac{\alpha^2}{5C_\ell^2(\Gamma+1)^2}\right),$$

where $\Gamma = \Gamma(T) \geqslant 0$ is the constant from Lemma 11.1.5, which only depends on T.

Note that the constant Γ in this corollary depends on the game, and in the worst case carries a heavy dependence on the players' alphabet sizes. This is in contrast to Holenstein's two-player result for no-signalling games, which has no alphabet dependence at all [104]. This is generalized in our Theorem 11.2.1, since for two players we know by Lemma 11.1.3 that NS $\equiv$ SNOS, and we could directly read off bounds with constants already improving on Holenstein's. Looking a little into the proof allows us to optimize the constants even more, which we record as follows.

Theorem 11.2.3 (Parallel repetition for the no-signalling value of 2-player games, cf. Holenstein [104])**.** *Let G be a 2-player game with $\omega_{NS}(G) \leqslant 1 - \delta$ for some $0 < \delta < 1$. Then, for any $n \in \mathbf{N}$, and any $t \geqslant (1 - \delta + \alpha)n$ for some $0 < \alpha \leqslant \delta$, we have*

$$\omega_{NS}(G^n) \leqslant \left(1 - \frac{\delta^2}{27}\right)^n,$$
$$\omega_{NS}(G^{t/n}) \leqslant \exp\left(-n\frac{\alpha^2}{33}\right).$$

11.3 Constrained de Finetti reduction

De Finetti reductions are a useful tool when trying to understand any permutation-invariant information processing task. Indeed, these enable to restrict the analysis to that of i.i.d. scenarios, which are usually trivially understood, as it was exemplified in Chapter 10. In the context of multi-player games played n times in parallel,

one would like to use the fact that the numbering of the n instances of the repeated game is irrelevant to reduce the study of strategies for the latter to the study of so-called *de Finetti strategies* (i.e. convex combinations of n i.i.d. strategies).

The seminal de Finetti reduction (aka post-selection) lemma was stated in [49], later finding applications in many areas of quantum information theory, from quantum cryptography [41] to quantum Shannon theory [29]. Our proofs though, will rely on two more recently established de Finetti reduction results, which are stated below. Just to fix some definitions: we will say that a (sub-)probability distribution $P_{\mathcal{Z}^n}$, resp. a conditional (sub-)probability distribution $P_{\mathcal{B}^n|\mathcal{Y}^n}$, is n-symmetric if for any permutation π of n elements, $\forall\, z^n,\ P(\pi(z^n)) = P(z^n)$, resp. $\forall\, b^n, y^n,\ P(\pi(b^n)|\pi(y^n)) = P(b^n|y^n)$.

Lemma 11.3.1 (de Finetti reduction for conditional p.d.'s, [6])**.** *Let $\mathcal{B}, \mathcal{Y}$ be finite alphabets. There exists a probability measure $dR_{\mathcal{B}|\mathcal{Y}}$ on the set of conditional probability distributions $R_{\mathcal{B}|\mathcal{Y}}$ such that, for any n-symmetric conditional probability distribution $P_{\mathcal{B}^n|\mathcal{Y}^n}$,*

$$P_{\mathcal{B}^n|\mathcal{Y}^n} \leqslant \mathrm{poly}(n) \int_{R_{\mathcal{B}|\mathcal{Y}}} R_{\mathcal{B}|\mathcal{Y}}^{\otimes n}\, \mathrm{d}R_{\mathcal{B}|\mathcal{Y}},$$

where the polynomial pre-factor may be upper bounded as $\mathrm{poly}(n) \leqslant (n+1)^{|\mathcal{B}||\mathcal{Y}|}$.

Lemma 11.3.2 (Constrained de Finetti reduction for (sub-)p.d.'s, [Chapter 10, Section 10.2, Corollary 10.2.6])**.** *Let $\mathcal{X}$ be a finite alphabet. There exists a probability measure $dQ_{\mathcal{X}}$ on the set of probability distributions $Q_{\mathcal{X}}$ on $\mathcal{X}$ such that, for any n-symmetric (sub-)probability distribution $P_{\mathcal{X}^n}$ on $\mathcal{X}^n$,*

$$P_{\mathcal{X}^n} \leqslant \mathrm{poly}(n) \int_{Q_{\mathcal{X}}} F\left(P_{\mathcal{X}^n}, Q_{\mathcal{X}}^{\otimes n}\right)^2 Q_{\mathcal{X}}^{\otimes n}\, \mathrm{d}Q_{\mathcal{X}},$$

where the polynomial pre-factor may be upper bounded as $\mathrm{poly}(n) \leqslant (n+1)^{3|\mathcal{X}|^2}$.

In Lemma 11.3.2 above, as well as in the remainder of this chapter, $F(P,Q)$ stands for the fidelity between probability distributions P and Q, defined as $F(P,Q) = \|\sqrt{P}\sqrt{Q}\|_1$.

We are now ready to present the technical lemma that will allow us in Section 11.4 to reduce the study of strategies for repeated games to the study of so-called *de Finetti strategies*, and hence prove our main results.

Lemma 11.3.3 (de Finetti reduction for sub-no-signalling correlations)**.** *There exists a probability measure dQ on the set of probability distributions Q on $\underline{\mathcal{A}} \times \underline{\mathcal{X}}$ such that for any probability distribution T on $\underline{\mathcal{X}}$ and any $P \in SNOS(\underline{\mathcal{A}}^n|\underline{\mathcal{X}}^n)$ an n-symmetric sub-no-signalling correlation, it holds that*

$$T_{\underline{\mathcal{X}}}^{\otimes n} P_{\underline{\mathcal{A}}^n|\underline{\mathcal{X}}^n} \leqslant \mathrm{poly}(n) \int_{Q_{\underline{\mathcal{A}}\underline{\mathcal{X}}}} \widetilde{F}\left(Q_{\underline{\mathcal{A}}\underline{\mathcal{X}}}\right)^{2n} Q_{\underline{\mathcal{A}}\underline{\mathcal{X}}}^{\otimes n}\, \mathrm{d}Q_{\underline{\mathcal{A}}\underline{\mathcal{X}}}, \tag{11.4}$$

where we defined

$$\widetilde{F}\left(Q_{\underline{\mathcal{A}}\underline{\mathcal{X}}}\right) := \min_{\emptyset \neq I \subsetneq [\ell]} \max_{R_{\mathcal{A}_I|\mathcal{X}_I}} F\left(T_{\underline{\mathcal{X}}} R_{\mathcal{A}_I|\mathcal{X}_I}, Q_{\mathcal{A}_I\underline{\mathcal{X}}}\right).$$

We mention for the sake of completeness that the $\mathrm{poly}(n)$ *pre-factor in equation* (11.4) *may be upper bounded by $(n+1)^{3|\underline{\mathcal{A}}|^2|\underline{\mathcal{X}}|^2+2|\underline{\mathcal{A}}||\underline{\mathcal{X}}|}$.*

Proof. Since $T_{\underline{\mathcal{X}}}^{\otimes n} P_{\underline{\mathcal{A}}^n|\underline{\mathcal{X}}^n}$ is an n-symmetric sub-probability distribution on $(\underline{\mathcal{A}}\underline{\mathcal{X}})^n$, we first of all have by Lemma 11.3.2 that

$$T_{\underline{\mathcal{X}}}^{\otimes n} P_{\underline{\mathcal{A}}^n|\underline{\mathcal{X}}^n} \leqslant \mathrm{poly}(n) \int_{Q_{\underline{\mathcal{A}}\underline{\mathcal{X}}}} F\left(T_{\underline{\mathcal{X}}}^{\otimes n} P_{\underline{\mathcal{A}}^n|\underline{\mathcal{X}}^n}, Q_{\underline{\mathcal{A}}\underline{\mathcal{X}}}^{\otimes n}\right)^2 Q_{\underline{\mathcal{A}}\underline{\mathcal{X}}}^{\otimes n}\, \mathrm{d}Q_{\underline{\mathcal{A}}\underline{\mathcal{X}}}.$$

Notice next that, for any $\emptyset \neq I \subsetneq [\ell]$,

$$F\left(T_{\underline{\mathcal{X}}}^{\otimes n} P_{\underline{\mathcal{A}}^n|\underline{\mathcal{X}}^n}, Q_{\underline{\mathcal{A}}\underline{\mathcal{X}}}^{\otimes n}\right) \leqslant F\left(T_{\underline{\mathcal{X}}}^{\otimes n} P_{\mathcal{A}_I^n|\underline{\mathcal{X}}^n}, Q_{\mathcal{A}_I\underline{\mathcal{X}}}^{\otimes n}\right) \leqslant F\left(T_{\underline{\mathcal{X}}}^{\otimes n} P'_{\mathcal{A}_I^n|\mathcal{X}_I^n}, Q_{\mathcal{A}_I\underline{\mathcal{X}}}^{\otimes n}\right).$$

The first inequality is by monotonicity of the fidelity under stochastic maps (in particular taking marginals). While the second inequality is because $P \in \mathrm{SNOS}(\underline{\mathcal{A}}^n|\underline{\mathcal{X}}^n)$, so that $P_{\mathcal{A}_I^n|\underline{\mathcal{X}}^n} \leqslant P'_{\mathcal{A}_I^n|\mathcal{X}_I^n}$ for some conditional p.d. $P'_{\mathcal{A}_I^n|\mathcal{X}_I^n}$, and because the fidelity is order-preserving.

What is more, for any $\emptyset \neq I \subsetneq [\ell]$, $P'_{\mathscr{A}_I^n|\mathscr{X}_I^n}$ can be chosen to be an n-symmetric conditional probability distribution. Indeed, if it were not, its n-symmetrization would still upper bound $P_{\mathscr{A}_I^n|\underline{\mathscr{X}}^n}$ (since the latter is by assumption n-symmetric). We then have by Lemma 11.3.1 that

$$P'_{\mathscr{A}_I^n|\mathscr{X}_I^n} \leqslant \mathrm{poly}(n) \int_{R_{\mathscr{A}_I|\mathscr{X}_I}} R_{\mathscr{A}_I|\mathscr{X}_I}^{\otimes n} \, \mathrm{d}R_{\mathscr{A}_I|\mathscr{X}_I},$$

and subsequently, using first, once more, that the fidelity is order-preserving, and second that it is multiplicative on tensor products,

$$F\left(T_{\underline{\mathscr{X}}}^{\otimes n} P'_{\mathscr{A}_I^n|\mathscr{X}_I^n}, Q_{\mathscr{A}_I\underline{\mathscr{X}}}^{\otimes n}\right) \leqslant \mathrm{poly}(n) \max_{R_{\mathscr{A}_I|\mathscr{X}_I}} F\left(T_{\underline{\mathscr{X}}}^{\otimes n} R_{\mathscr{A}_I|\mathscr{X}_I}^{\otimes n}, Q_{\mathscr{A}_I\underline{\mathscr{X}}}^{\otimes n}\right) = \mathrm{poly}(n) \max_{R_{\mathscr{A}_I|\mathscr{X}_I}} F\left(T_{\underline{\mathscr{X}}} R_{\mathscr{A}_I|\mathscr{X}_I}, Q_{\mathscr{A}_I\underline{\mathscr{X}}}\right)^n.$$

Recapitulating, we get

$$T_{\underline{\mathscr{X}}}^{\otimes n} P_{\underline{\mathscr{A}}^n|\underline{\mathscr{X}}^n} \leqslant \mathrm{poly}(n) \int_{Q_{\underline{\mathscr{A}}\underline{\mathscr{X}}}} \left(\min_{\emptyset \neq I \subsetneq [\ell]} \max_{R_{\mathscr{A}_I|\mathscr{X}_I}} F\left(T_{\underline{\mathscr{X}}} R_{\mathscr{A}_I|\mathscr{X}_I}, Q_{\mathscr{A}_I\underline{\mathscr{X}}}\right) \right)^{2n} Q_{\underline{\mathscr{A}}\underline{\mathscr{X}}}^{\otimes n} \, \mathrm{d}Q_{\underline{\mathscr{A}}\underline{\mathscr{X}}},$$

as announced. □

11.4 Proofs of the main Theorems

In this section we prove Theorem 11.2.1, Corollary 11.2.2 and Theorem 11.2.3.

We need first of all the following extension of Lemma 9.5 in [104]:

Lemma 11.4.1. *For $\underline{\mathscr{X}} = \bigtimes_{j=1}^m \mathscr{X}_j$ and $\underline{\mathscr{B}} = \bigtimes_{j=1}^m \mathscr{B}_j$, consider probability distributions T on $\underline{\mathscr{X}}$ and P on $\underline{\mathscr{B}} \times \underline{\mathscr{X}}$ satisfying*

$$\frac{1}{2} \left\| P_{\underline{\mathscr{X}}} - T_{\underline{\mathscr{X}}} \right\|_1 \leqslant \epsilon_0.$$

If for each $j \in [m]$ there exists a conditional probability distribution $Q(b_j|z_j)$ such that

$$\left\| P_{\mathscr{B}_j\underline{\mathscr{X}}} - T_{\underline{\mathscr{X}}} Q_{\mathscr{B}_j|\mathscr{X}_j} \right\|_1 \leqslant \epsilon_j,$$

then there exists a conditional probability distribution $P'(\underline{b}|\underline{z})$ such that, for each $j \in [m]$, $P'(b_j|\underline{z}) = P'(b_j|z_j)$ for all $b_j, \underline{z}$, and

$$\frac{1}{2} \left\| T_{\underline{\mathscr{X}}} P'_{\underline{\mathscr{B}}|\underline{\mathscr{X}}} - P_{\underline{\mathscr{B}}\underline{\mathscr{X}}} \right\|_1 \leqslant \epsilon_0 + \sum_{j=1}^m 2\epsilon_j.$$

Proof. This works exactly as the proof of Lemma 9.5 in [104], which is a successive application (m times) of Lemma 9.4 in the same paper. The latter relies on the fact that the statistical distance $\|P_1 - P_2\|_1/2$ between two probability distributions P_1, P_2 can be equivalently characterized as the minimum probability that X_1 differs from X_2 over pairs of random variables (X_1, X_2) sampled from P having (P_1, P_2) as marginals. □

Note that the conditions enforced in Lemma 11.4.1 are not enough to ensure no-signalling of P' for three or more players. They would be sufficient though to guarantee that P' satisfies the relaxed no-signalling constraints considered in [154], namely that any group of $\ell - 1$ players together cannot signal to the remaining player. Nevertheless, we can leverage this result to approximate the given no-signalling correlation by a sub-no-signalling correlation.

Lemma 11.4.2. *Let P be a probability distribution on $\underline{\mathscr{A}} \times \underline{\mathscr{X}}$ and T be a probability distribution on $\underline{\mathscr{X}}$. If the no-signalling conditions (11.1) hold approximately, namely*

$$\forall\ I \subsetneq [\ell],\ \exists\ Q(\cdot|x_I)\ \textit{p.d.'s on}\ \mathscr{A}_I\ \textit{s.t.}\ \frac{1}{2} \left\| P_{\mathscr{A}_I\underline{\mathscr{X}}} - T_{\underline{\mathscr{X}}} Q_{\mathscr{A}_I|\mathscr{X}_I} \right\|_1 \leqslant \epsilon_I,$$

then there exists a sub-no-signalling correlation $P' \in SNOS(\underline{\mathscr{A}}|\underline{\mathscr{X}})$ that approximates P, in the sense that

$$\frac{1}{2} \left\| T_{\underline{\mathscr{X}}} P'_{\underline{\mathscr{A}}|\underline{\mathscr{X}}} - P_{\underline{\mathscr{A}}\underline{\mathscr{X}}} \right\|_1 \leqslant \epsilon_\emptyset + \sum_{\emptyset \neq I \subsetneq [\ell]} 2\epsilon_I.$$

In the two-player case $\ell = 2$, P' can be chosen to be no-signalling itself, $P' \in NS(\underline{\mathscr{A}}|\underline{\mathscr{X}})$.

Proof. We will apply Lemma 11.4.1, with $m = 2^\ell - 2$, the index j identifying a non-empty and non-full set $\emptyset \neq I \subsetneq [\ell]$ (for instance via the expansion of j into ℓ binary digits). The local input and output alphabets are

$$\mathcal{X}_j = \prod_{i \in I} \mathcal{X}_i, \quad \mathcal{B}_j = \prod_{i \in I} \mathcal{A}_i,$$

and the distribution we apply it to is

$$\widehat{P}(\underline{b}\underline{z}) = \begin{cases} P(\underline{a}\underline{x}) & \text{if } \forall j,\ b_j = (a_i : i \in I),\ z_j = (x_i : i \in I), \\ 0 & \text{otherwise.} \end{cases}$$

Likewise, the prior distribution on $\underline{\mathcal{X}}$ is given by

$$\widehat{T}(\underline{z}) = \begin{cases} T(\underline{x}) & \text{if } \forall j,\ z_j = (x_i : i \in I), \\ 0 & \text{otherwise,} \end{cases}$$

and we use the conditional distributions $Q(b_j|z_j) = Q(a_I|x_I)$.

Now, the prerequisites of Lemma 11.4.1 are given, with $\epsilon_j = \epsilon_I$, and thus we get a conditional probability distribution $\widehat{P}'$ with $\widehat{P}'(b_j|\underline{z}) = \widehat{P}'(b_j|z_j)$ for all j, and

$$\frac{1}{2}\big\|\widehat{T}_{\underline{\mathcal{X}}}\widehat{P}'_{\underline{\mathcal{B}}|\underline{\mathcal{X}}} - \widehat{P}_{\underline{\mathcal{B}}\underline{\mathcal{X}}}\big\|_1 \leqslant \epsilon_0 + \sum_{j=1}^{n} 2\epsilon_j =: \epsilon.$$

We would like to conclude here by "pulling back" this conditional distribution to a (it would seem: no-signalling) correlation on $\underline{\mathcal{A}} \times \underline{\mathcal{X}}$, except that P' has support outside the image of the diagonal embedding

$$\begin{aligned} \Delta : \underline{\mathcal{A}} &\longrightarrow \underline{\mathcal{B}} \\ \underline{a} &\longmapsto \underline{b} \text{ s.t. } \forall j,\ b_j = (a_i : i \in I), \end{aligned}$$

and likewise for $\Delta : \underline{\mathcal{X}} \longrightarrow \underline{\mathcal{X}}$.

To resolve this issue, we simply remove this part of the distribution, and define the desired sub-normalized conditional densities by letting

$$P'(\underline{a}|\underline{x}) := \widehat{P}'\big(\Delta(\underline{a})|\Delta(\underline{x})\big).$$

From this we see directly that

$$\frac{1}{2}\big\|T_{\underline{\mathcal{X}}}P'_{\underline{\mathcal{A}}|\underline{\mathcal{X}}} - P_{\underline{\mathcal{A}}\underline{\mathcal{X}}}\big\|_1 \leqslant \epsilon,$$

because $\widehat{P}(\underline{b},\underline{z}) = P(\underline{a},\underline{x})$ for $\underline{b} = \Delta(\underline{a})$ and $\underline{z} = \Delta(\underline{x})$, and it is 0 outside the image of Δ.

It remains to check that P' is sub-no-signalling. Let $\emptyset \neq I \subsetneq [\ell]$ be a subset with corresponding index $1 \leqslant j \leqslant 2^\ell - 2$. Let also $\underline{x} \in \underline{\mathcal{X}}$, $a_I \in \mathcal{A}_I$ be tuples, and set $\underline{z} = \Delta(\underline{x})$, $\underline{b} = \Delta(\underline{a})$ (so that $z_j = x_I \in \mathcal{X}_I = \mathcal{X}_j$, $b_j = a_I \in \mathcal{A}_I = \mathcal{B}_j$). Then,

$$\begin{aligned} P'(a_I|\underline{x}) &= \sum_{a_{I^c} \in \mathcal{A}_{I^c}} P'(\underline{a}|\underline{x}) \\ &= \sum_{a_{I^c} \in \mathcal{A}_{I^c}} \widehat{P}'\big(\Delta(\underline{a})|\Delta(\underline{x})\big) \\ &\leqslant \sum_{b_k \in \mathcal{B}_k,\ k \neq j} \widehat{P}'\big(\underline{b}|\underline{z}\big) \\ &= \widehat{P}'(b_j|\underline{z}\big) \\ &= \widehat{P}'(b_j|z_j) =: Q'(a_I|x_I). \end{aligned}$$

Here, we have used the definition of the marginal and of P'. The inequality in the third line is because we enlarge the domain of the summation, and the equality in the last line is by the marginal property of $\widehat{P}'$.

The last claim, regarding $\ell = 2$ players, is the original Lemma 9.5 in [104]. □

We are now ready to prove our main theorem, namely the parallel repetition and concentration results for the sub-no-signalling value of multi-player games.

Proof of Theorem 11.2.1. Let $P_{\underline{\mathcal{A}}^n|\underline{\mathcal{X}}^n}$ be a sub-no-signalling correlation which is optimal to win the game G^n. The distribution $T_{\underline{\mathcal{X}}}^{\otimes n}$ and the predicate $V_{\underline{\mathcal{A}}\underline{\mathcal{X}}}^{\otimes n}$ of G^n being n-symmetric, we can assume without loss of generality that $P_{\underline{\mathcal{A}}^n|\underline{\mathcal{X}}^n}$ is also n-symmetric. Indeed, since for any permutation π of n elements, $T \circ \pi = T$ and $V \circ \pi = V$, playing G^n with P or with $P \circ \pi$ yields the same winning probability. And therefore, if P is an optimal strategy then so is its symmetrization over all permutations of n elements. Hence, by Lemma 11.3.3,

$$T_{\underline{\mathcal{X}}}^{\otimes n} P_{\underline{\mathcal{A}}^n|\underline{\mathcal{X}}^n} \leqslant \mathrm{poly}(n) \int_{Q_{\underline{\mathcal{A}}\underline{\mathcal{X}}}} \widetilde{F}\left(Q_{\underline{\mathcal{A}}\underline{\mathcal{X}}}\right)^{2n} Q_{\underline{\mathcal{A}}\underline{\mathcal{X}}}^{\otimes n} \, \mathrm{d}Q_{\underline{\mathcal{A}}\underline{\mathcal{X}}}.$$

Now, fix $0 < \epsilon < 1$ and define

$$\mathscr{P}_\epsilon := \left\{ Q_{\underline{\mathcal{A}}\underline{\mathcal{X}}} \ : \ \max_{\emptyset \neq I \subsetneq [\ell]} \min_{R_{\mathcal{A}_I|\mathcal{X}_I}} \frac{1}{2} \| T_{\underline{\mathcal{X}}} R_{\mathcal{A}_I|\mathcal{X}_I} - Q_{\mathcal{A}_I\underline{\mathcal{X}}} \|_1 \leqslant \epsilon \right\}.$$

Observe that, by well-known relations between fidelity and trace-distance (see e.g. [78]), if $Q_{\underline{\mathcal{A}}\underline{\mathcal{X}}} \notin \mathscr{P}_\epsilon$, then automatically $\widetilde{F}\left(Q_{\underline{\mathcal{A}}\underline{\mathcal{X}}}\right)^2 \leqslant 1 - \epsilon^2$. Hence,

$$T_{\underline{\mathcal{X}}}^{\otimes n} P_{\underline{\mathcal{A}}^n|\underline{\mathcal{X}}^n} \leqslant \mathrm{poly}(n) \left(\int_{Q_{\underline{\mathcal{A}}\underline{\mathcal{X}}} \in \mathscr{P}_\epsilon} Q_{\underline{\mathcal{A}}\underline{\mathcal{X}}}^{\otimes n} \, \mathrm{d}Q_{\underline{\mathcal{A}}\underline{\mathcal{X}}} + (1-\epsilon^2)^n \int_{Q_{\underline{\mathcal{A}}\underline{\mathcal{X}}} \notin \mathscr{P}_\epsilon} Q_{\underline{\mathcal{A}}\underline{\mathcal{X}}}^{\otimes n} \, \mathrm{d}Q_{\underline{\mathcal{A}}\underline{\mathcal{X}}} \right).$$

On the other hand, if $Q_{\underline{\mathcal{A}}\underline{\mathcal{X}}} \in \mathscr{P}_\epsilon$, then by definition

$$\forall \ \emptyset \neq I \subsetneq [\ell], \ \exists \ R_{\mathcal{A}_I|\mathcal{X}_I} : \ \frac{1}{2} \| T_{\underline{\mathcal{X}}} R_{\mathcal{A}_I|\mathcal{X}_I} - Q_{\mathcal{A}_I\underline{\mathcal{X}}} \|_1 \leqslant \epsilon.$$

By Lemma 11.4.2, the latter condition implies that there exists a sub-no-signalling correlation $R'_{\underline{\mathcal{A}}|\underline{\mathcal{X}}}$ such that

$$\frac{1}{2} \| T_{\underline{\mathcal{X}}} R'_{\underline{\mathcal{A}}|\underline{\mathcal{X}}} - Q_{\underline{\mathcal{A}}\underline{\mathcal{X}}} \|_1 \leqslant C_\ell \epsilon, \ \text{where } C_\ell = 1 + 2(2^\ell - 2) = 2^{\ell+1} - 3.$$

Yet, the winning probability when playing G with a strategy $R'_{\underline{\mathcal{A}}|\underline{\mathcal{X}}} \in \mathrm{SNOS}(\underline{\mathcal{A}}|\underline{\mathcal{X}})$ is, by assumption on G, at most $1 - \delta$. So the average of the predicate of G over $Q_{\underline{\mathcal{A}}\underline{\mathcal{X}}} \in \mathscr{P}_\epsilon$ is at most $1 - \delta + 2C_\ell\epsilon$. Putting everything together, we eventually get that the winning probability when playing G^n with strategy $P_{\underline{\mathcal{A}}^n|\underline{\mathcal{X}}^n}$ is upper bounded as

$$\mathbf{P}(\mathrm{win}) \leqslant \mathrm{poly}(n) \left((1 - \delta + 2C_\ell\epsilon)^n + (1-\epsilon^2)^n \right). \tag{11.5}$$

Choosing in equation (11.5)

$$\epsilon = C_\ell \left(\left(1 + \frac{\delta}{C_\ell^2} \right)^{1/2} - 1 \right) \geqslant \frac{99\delta}{200 C_\ell}, \ \text{so that } \epsilon^2 \geqslant \frac{\delta^2}{5C_\ell^2},$$

and recalling that $P_{\underline{\mathcal{A}}^n|\underline{\mathcal{X}}^n}$ is, by hypothesis, an optimal sub-no-signalling strategy, we obtain

$$\omega_{\mathrm{SNOS}}(G^n) \leqslant \mathrm{poly}(n) \left(1 - \frac{\delta^2}{5C_\ell^2} \right)^n. \tag{11.6}$$

In order to conclude, we have to remove the polynomial pre-factor. So assume that there exists a constant $C > 0$ such that for some $N \in \mathbf{N}$, $\omega_{\mathrm{SNOS}}(G^N) \geqslant C \left(1 - \delta^2/5C_\ell^2\right)^N$. Then, for any $n \in \mathbf{N}$, we would have

$$\omega_{\mathrm{SNOS}}(G^{Nn}) \geqslant \left(\omega_{\mathrm{SNOS}}(G^N)\right)^n \geqslant C^n \left(1 - \frac{\delta^2}{5C_\ell^2} \right)^{Nn}.$$

On the other hand, however, we still have by equation (11.6)

$$\omega_{\mathrm{SNOS}}(G^{Nn}) \leqslant \mathrm{poly}(Nn) \left(1 - \frac{\delta^2}{5C_\ell^2} \right)^{Nn}.$$

Letting n grow, we see that the only option to make these two conditions compatible is to have $C \leqslant 1$, which is precisely what we wanted to show.

Following the exact same lines as above, we also get the concentration bound. Indeed, we now have in place of equation (11.5) that, for any $0 < \epsilon < 1$,

$$\omega_{\mathrm{SNOS}}(G^{t/n}) \leqslant \mathrm{poly}(n) \left(\exp\left[-2n(\alpha - 2C_\ell\epsilon)^2 \right] + \exp\left[-n\epsilon^2 \right] \right), \tag{11.7}$$

by Hoeffding's inequality (and because $e^{-x} \geqslant 1-x$ for any $x > 0$).

The announced upper bound follows from choosing in equation (11.7)

$$\epsilon = \frac{(4C_\ell - \sqrt{2})\alpha}{8C_\ell^2 - 1} \geqslant \frac{5(20-\sqrt{2})\alpha}{199 C_\ell}, \text{ so that } \epsilon^2 \geqslant \frac{\alpha^2}{5C_\ell^2},$$

and removing the polynomial pre-factor by the same trick as before. □

Proof of Corollary 11.2.2. By Lemma 11.1.5, we know that if G is an ℓ-player game with full support satisfying $\omega_{\mathrm{NS}}(G) \leqslant 1-\delta$, then $\omega_{\mathrm{SNOS}}(G) \leqslant 1 - \delta/(\Gamma+1)$. And thus by Theorem 11.2.1,

$$\omega_{\mathrm{NS}}(G^n) \leqslant \omega_{\mathrm{SNOS}}(G^n) \leqslant \left(1 - \frac{\delta^2}{5C_\ell^2(\Gamma+1)^2}\right)^n.$$

The concentration bound for $\omega_{\mathrm{NS}}(G^{t/n})$ follows analogously. □

Proof of Theorem 11.2.3. We follow the exact same reasoning as in the proof of Theorem 11.2.1, and keep the same notation. In the case $\ell = 2$, we have by Lemma 11.4.2 that, for any $0 < \epsilon < 1$,

$$Q_{\underline{\mathscr{A}\mathscr{X}}} \in \mathscr{P}_\epsilon \;\Rightarrow\; \exists\, R'_{\underline{\mathscr{A}\mathscr{X}}} \in \mathrm{NS}(\underline{\mathscr{A}}|\underline{\mathscr{X}}) : \; \frac{1}{2}\| T_{\underline{\mathscr{X}}} R'_{\underline{\mathscr{A}}|\underline{\mathscr{X}}} - Q_{\underline{\mathscr{A}\mathscr{X}}} \|_1 \leqslant 5\epsilon.$$

Yet, if the winning probability when playing G with a strategy $R'_{\underline{\mathscr{A}}|\underline{\mathscr{X}}} \in \mathrm{NS}(\underline{\mathscr{A}}|\underline{\mathscr{X}})$ is, by assumption on G, at most $1-\delta$, then the average of the predicate of G over $Q_{\underline{\mathscr{A}\mathscr{X}}} \in \mathscr{P}_\epsilon$ is at most $1 - \delta + 5\epsilon$. This is because we are here dealing with normalised probability distributions. Hence, for any $0 < \epsilon < 1$,

$$\begin{aligned}
&\omega_{\mathrm{NS}}(G^n) \leqslant \mathrm{poly}(n)\left((1-\delta+5\epsilon)^n + (1-\epsilon^2)^n\right), \\
&\omega_{\mathrm{SNOS}}(G^{t/n}) \leqslant \mathrm{poly}(n)\left(\exp\left[-2n(\alpha-5\epsilon)^2\right] + \exp\left[-n\epsilon^2\right]\right).
\end{aligned}$$

We can now choose $\epsilon = (\sqrt{29}-5)\delta/2$ in the parallel repetition estimate and $\epsilon = (10-\sqrt{2})\alpha/49$ in the concentration bound one, and argue as in the proof of Theorem 11.2.1 to remove the polynomial pre-factor, which yields the two advertised results. □

11.5 Discussion

Our main contribution in the present chapter is a concentration result for the sub-no-signalling value of multi-player games under parallel repetition. In fact, we believe that our work is the first to recognize the intrinsic interest of the class of sub-no-signalling correlations, which appears naturally as a relaxation of the no-signalling ones.

Specifically, if an ℓ-player game G has SNOS value $1-\delta$, then the probability for SNOS players to win a fraction at least $1-\delta+\alpha$ of n instances of G played in parallel is at most $\exp(-nC_\ell\alpha^2)$, where $C_\ell > 0$ is a constant which only depends on the number ℓ of players. As mentioned in [40], such result, valid for games involving strictly more than 2 players and where not all queries are asked, could potentially find applications in position-based cryptography [39, 73]. In the case $\ell = 2$, this is actually equivalent to the analogous concentration result for the no-signalling value of G, thus with a universal constant $c = C_2$ in the exponential bound. And we know we cannot hope for a better dependence in α than the obtained one, even in the special case $\alpha = \delta$. Indeed, as explained in [121], strong parallel repetition in general does not hold for no-signalling players. In the case $\ell > 2$, our result implies a concentration bound for the no-signalling value of G, but only if its input distribution has full support. Besides, the constant in the exponential bound is this time highly game-dependent (dependence on the sizes of the input and output alphabets, and on the smallest weight occurring in the input distribution). This is fully comparable to previous work in this direction due to Buhrman, Fehr and Schaffner [40], and Arnon-Friedman, Renner and Vidick [7].

Hence, the most immediate open problem at that point is regarding games with non-full support in the case of three or more players (e.g. the anti-correlation game): does a parallel repetition result hold for the no-signalling value of such multi-player games? Answering this question probably requires to understand first whether in Corollary 11.2.2, the presence of the game parameter Γ is really necessary or is just an artifact of the proof technique. In other words, does the rate at which the no-signalling value of a game decays under parallel repetition truly depends on the game distribution?

Another issue that would be worth investigating is whether constrained de Finetti reductions could also be used to establish parallel repetition results for the classical or quantum value of multi-player games. Formally, the sets of classical correlations $\mathrm{C}(\underline{\mathscr{A}}|\underline{\mathscr{X}})$ and quantum correlations $\mathrm{Q}(\underline{\mathscr{A}}|\underline{\mathscr{X}})$ are defined as follows:

$$P \in \mathrm{C}(\underline{\mathscr{A}}|\underline{\mathscr{X}}) :\Leftrightarrow \forall\ \underline{x}, \underline{a},\ P(\underline{a}|\underline{x}) = \sum_{m \in \mathscr{M}} Q(m) P_1(a_1|x_1\, m) \cdots P_\ell(a_\ell|x_\ell\, m),$$

for some p.d. Q on some alphabet $\mathscr{M}$ and some p.d.'s $P_i(\cdot|x_i\, m)$ on $\mathscr{A}_i$.

$$P \in \mathrm{Q}(\underline{\mathscr{A}}|\underline{\mathscr{X}}) :\Leftrightarrow \forall\ \underline{x}, \underline{a},\ P(\underline{a}|\underline{x}) = \langle\psi| M(x_1)_{a_1} \otimes \cdots \otimes M(x_\ell)_{a_\ell} |\psi\rangle,$$

for some pure state $|\psi\rangle\langle\psi|$ on $\mathrm{H}_1 \otimes \cdots \otimes \mathrm{H}_\ell$ and some POVMs $M(x_i)$ on H_i.
And the classical, resp. quantum, value of an ℓ-player game G with distribution T and predicate V, denoted $\omega_{\mathrm{C}}(G)$, resp. $\omega_{\mathrm{Q}}(G)$, is then naturally defined as the maximum, resp. supremum, of the winning probability

$$\mathbf{P}\,(\mathrm{win}) = \sum_{\underline{a},\underline{x}} T(\underline{x}) V(\underline{a}, \underline{x}) P(\underline{a}|\underline{x})$$

over all $P \in \mathrm{C}(\underline{\mathscr{A}}|\underline{\mathscr{X}})$, resp. $P \in \mathrm{Q}(\underline{\mathscr{A}}|\underline{\mathscr{X}})$.

In the classical case, the first parallel repetition result for two-player games was established by Raz [151], and later improved by Holenstein [104], while Rao [150] gave a concentration bound. However, the proof techniques are arguably not as straightforward as via de Finetti reductions, and do not generalise directly to any number ℓ of players. In the quantum case, even less is known. The best parallel repetition result up to now is the one established by Chailloux and Scarpa [43] (subsequently improved by Chung, Wu and Yuen [52]), which applies to two-player (ℓ-player) free games, and from there to games with full support. That is why being able to export ideas from the de Finetti approach to these two cases would be of great interest. Roughly speaking, the problem we are facing is the following: Given an n-symmetric correlation $P_{\underline{\mathscr{A}}^n|\underline{\mathscr{X}}^n}$, we can always write the first step in the proof of Lemma 11.3.3, i.e.

$$T_{\underline{\mathscr{X}}}^{\otimes n} P_{\underline{\mathscr{A}}^n|\underline{\mathscr{X}}^n} \leqslant \mathrm{poly}(n) \int_{Q_{\underline{\mathscr{A}}\underline{\mathscr{X}}}} F\left(T_{\underline{\mathscr{X}}}^{\otimes n} P_{\underline{\mathscr{A}}^n|\underline{\mathscr{X}}^n}, Q_{\underline{\mathscr{A}}\underline{\mathscr{X}}}^{\otimes n}\right)^2 Q_{\underline{\mathscr{A}}\underline{\mathscr{X}}}^{\otimes n}\, \mathrm{d}Q_{\underline{\mathscr{A}}\underline{\mathscr{X}}}. \tag{11.8}$$

Now, we would like to argue that if $P_{\underline{\mathscr{A}}^n|\underline{\mathscr{X}}^n}$ is a classical, resp. quantum, correlation, then the p.d.'s $Q_{\underline{\mathscr{A}}\underline{\mathscr{X}}}$ for which the fidelity weight in the r.h.s. of equation (11.8) is not exponentially small are necessarily close to being of the form $T_{\underline{\mathscr{X}}} R_{\underline{\mathscr{A}}|\underline{\mathscr{X}}}$ for some classical, resp. quantum, correlation $R_{\underline{\mathscr{A}}|\underline{\mathscr{X}}}$. This was precisely our proof philosophy in the no-signalling case. However, the fact that the classical and quantum conditions are not properties that one can read off on the marginals, contrary to the no-signalling one, seems to be a first obstacle to surmount.

One related legitimate question would be the following: is it possible to make an even stronger statement than the one that, as explained above, we either are looking for (in the classical and quantum cases) or already have (in the no-signalling case)? Namely, could we upper bound $T_{\underline{\mathscr{X}}}^{\otimes n} P_{\underline{\mathscr{A}}^n|\underline{\mathscr{X}}^n}$ by a de Finetti distribution analogous to that in the r.h.s. of equation (11.8), but with weight strictly 0 on p.d.'s $Q_{\underline{\mathscr{A}}\underline{\mathscr{X}}}$ which are not of the form $T_{\underline{\mathscr{X}}} R_{\underline{\mathscr{A}}|\underline{\mathscr{X}}}$, for $R_{\underline{\mathscr{A}}|\underline{\mathscr{X}}}$ belonging to the same class as $P_{\underline{\mathscr{A}}^n|\underline{\mathscr{X}}^n}$? The answer to this question is no. Indeed, such improved de Finetti reduction would imply a strong parallel repetition result, which we know does not hold (see [7] for a similar discussion). So the best we can hope for is really to show that the fidelity weight in our upper bounding de Finetti distribution is exponentially small on the p.d.'s which are too far from being of the desired form.

Finally, let us briefly comment on the main spirit difference between the present work and the one by Arnon-Friedman *et al.* [7]. Our approach consists in using a more "flexible" de Finetti reduction, in which the information on the correlation $P_{\underline{\mathscr{A}}^n|\underline{\mathscr{X}}^n}$ and the p.d. $T_{\underline{\mathscr{X}}}^{\otimes n}$ of interest are kept in the upper bounding de Finetti distribution, through the fidelity weight $F(T_{\underline{\mathscr{X}}}^{\otimes n} P_{\underline{\mathscr{A}}^n|\underline{\mathscr{X}}^n}, Q_{\underline{\mathscr{A}}\underline{\mathscr{X}}}^{\otimes n})^2$. Whereas in [7], any initial correlation is first upper bounded by the same universal de Finetti correlation, on which a test (specifically tailored to the considered game distribution) is performed in a second step, that has the property of letting pass, resp. rejecting, with high probability the strategies which are no-signalling, resp. too signalling. So it seems in the end that both approaches are quite closely related: in our case, the "signalling test" which is applied to a given p.d. $Q_{\underline{\mathscr{A}}\underline{\mathscr{X}}}$ is nothing else than the maximal fidelity of $Q_{\underline{\mathscr{A}}\underline{\mathscr{X}}}$ to the set of p.d.'s of the form $T_{\underline{\mathscr{X}}} R_{\underline{\mathscr{A}}|\underline{\mathscr{X}}}$, with $R_{\underline{\mathscr{A}}|\underline{\mathscr{X}}}$ no-signalling, being above or below a certain threshold value. Also, it would be interesting (and potentially fruitful) to investigate whether one could combine in some way the techniques yielding Lemmas 11.3.1 and 11.3.2, to get a de Finetti reduction result that would have the advantages of both: namely, that is designed for conditional p.d.'s while at the same carrying the relevant information on the conditional p.d. it is applied to.

Part V

Outlook and perspectives

Let us wrap up with a (highly subjective) selection of further prospects, bringing together several of the different topics that were evoked over these pages (and therefore expected to be of less specific nature than the ones already put forward at the end of each chapter). When we allude to notions that were broached in a previous chapter, it is tacitely understood that we refer the reader to it for all needed definitions and notation.

Other models of random states

In many places of this manuscript, we were interested in studying whether random states satisfied a given property or not (e.g. being locally almost indistinguishable in Chapter 7, being separable or satisfying a separability criterion in Chapters 8 and 9 etc.). Depending on the context, the random state model that we chose slightly differed, but in the end all of them were qualitatively comparable. Let us be a bit more specific. Let $d, s \in \mathbf{N}$ and consider the following random state models on $\mathbf{C}^d$ (which were all used at some point or another):

- *Random induction*: $\rho = \operatorname{Tr}_{\mathbf{C}^s} |\psi\rangle\langle\psi|$ with ψ a uniformly distributed unit vector in $\mathbf{C}^d \otimes \mathbf{C}^s$, i.e. equivalently $\rho = W/\operatorname{Tr} W$ with W a Wishart matrix of parameter s on $\mathbf{C}^d$.
- *Random mixture*: $\rho = \big(\sum_{i=1}^s |\psi_i\rangle\langle\psi_i|\big)/s$ with $\psi_1, \ldots, \psi_s$ independent and uniformly distributed unit vectors in $\mathbf{C}^d$.
- *Maximally mixed + Gaussian noise*: In the case where $s > d$, $\rho = \big(\mathrm{Id} + G/2\sqrt{s}\big)/d$ with G a traceless GUE matrix on $\mathbf{C}^d$ (which is a state only with high probability).
- *Random renormalized projection*: In the case where $s < d$, $\rho = P_E/s$ with P_E the projector onto E a uniformly distributed s-dimensional subspace of $\mathbf{C}^d$.

When we talk about "random states", without any further specification, we usually refer to the first mentioned model. However, it may happen that one is only interested, for instance, in computing the average of a given convex function $f : \mathscr{H}(\mathbf{C}^d) \to \mathbf{R}$ over $\Delta = \rho - \mathrm{Id}/d$ with ρ a random state distributed in some way. This was more or less our case throughout this whole manuscript, where we always needed to estimate the average of the support function of a convex body in $\mathscr{H}(\mathbf{C}^d)$ (a set of either measurements or states). Then, in this situation, the nice thing is that the values obtained for $\mathbf{E}\, f$ over each of these four random state ensembles will all be of the same order (or even sometimes equivalent) as $d, s \to +\infty$. Let us briefly try to explain why (see Appendices A and B in [18] for more details, and Section 7.5 in Chapter 7 of this manuscript for a similar reasoning put in practice). First of all, define a function $\widetilde{f} : \mathbf{R}^d \to \mathbf{R}$, associated to f, by

$$\widetilde{f}(x) = \int_{\mathscr{U}(\mathbf{C}^d)} f\left(U \operatorname{diag}(x) U^\dagger\right) \mathrm{d}U.$$

Since all four mentioned shifted random state models are obviously unitarily invariant, $\widetilde{f}$ is such that, for Δ distributed according to one of these, $\mathbf{E}\, f(\Delta) = \mathbf{E}\, \widetilde{f}(\operatorname{spec}(\Delta))$. Next, for each of the considered shifted random state models, the limiting spectral distribution is well-known (Marčenko-Pastur, semi-circular, Bernoulli, with certain parameters). And these distributions can be compared to one another, in the sense that: for each two of them μ, μ', there exist constants $c, C > 0$ such that, for some random variables X, X' sampled from μ, μ', $cX \leqslant X' \leqslant CX$. Now, these asymptotic comparisons can be turned into non-asymptotic ones, because each of the empirical spectral distributions converges both weakly and in terms of extreme eigenvalues. Specifically, there exist constants $c_{d,s}, C_{d,s} > 0$ (going to c, C as $d, s \to +\infty$) such that, for some Δ, Δ' sampled from the two ensembles having μ, μ' as limiting spectra,

$$c_{d,s} \operatorname{spec}(\Delta) \preceq \operatorname{spec}(\Delta') \preceq C_{d,s} \operatorname{spec}(\Delta).$$

By elementary properties of majorization, this in turn implies (due to the convexity and the invariance under coordinate permutation of $\widetilde{f}$) that

$$c_{d,s} \widetilde{f}(\operatorname{spec}(\Delta)) \leqslant \widetilde{f}(\operatorname{spec}(\Delta')) \leqslant C_{d,s} \widetilde{f}(\operatorname{spec}(\Delta)).$$

And eventually, simply taking expectations, what we wanted to show, namely

$$c_{d,s}\, \mathbf{E}\, f(\Delta) \leqslant \mathbf{E}\, f(\Delta') \leqslant C_{d,s}\, \mathbf{E}\, f(\Delta).$$

Hence in a nutshell, the conclusion of this digression: depending on the problem we are looking at (and especially on the extra conditions we want to impose), working with one type or the other of these random states may be more convenient, but in the end the picture is usually more or less the same for all of them.

Yet, there are many settings where we would need to understand completely different kinds of random states. One example which might be worth mentioning is that of random tensor product states, i.e. states of the form $\rho_1 \otimes \cdots \otimes \rho_n$ with $\rho_1, \ldots, \rho_n$ random states (either independent or correlated in some way). Indeed, whenever one wants to study the multiplicative/additive behaviour of a certain function, this is precisely the type of states which shows up. And a quite natural wonder that one may have is: given a function which is known to violate multiplicativity/additivity in general, is it nevertheless (weakly) multiplicative/additive in typical scenarios? Such investigations were for instance carried on in [141] or [79] for the output entropy of quantum channels, yielding partial results. But there would be much more to explore. As a sample, closely related to preoccupations in this manuscript:

- For $\mathbf{M}$ being a set of locally restricted POVMs, such as **SEP** or **PPT**, does the distinguishability norm $\|\cdot\|_{\mathbf{M}}$ (cf. Chapter 7) generically exhibit a multiplicative behaviour? That is concretely, we ask if it is true that for random bipartite states $\rho_{\mathrm{AB}}, \sigma_{\mathrm{AB}}$, we have with high probability

$$\left\| \left(\rho_{\mathrm{AB}} - \sigma_{\mathrm{AB}}\right)^{\otimes n} \right\|_{\mathbf{M}(\mathrm{A}^n:\mathrm{B}^n)} \simeq \|\rho_{\mathrm{AB}} - \sigma_{\mathrm{AB}}\|^n_{\mathbf{M}(\mathrm{A}:\mathrm{B})} \, .$$

- For $\mathscr{K}$ being the set of separable states or one of its relaxations, does the maximum fidelity function $F(\cdot, \mathscr{K})$ (cf. Chapter 10) generically exhibit a multiplicative behaviour? That is concretely, we ask if it is true that for a random bipartite state ρ_{AB}, we have with high probability

$$F\big(\rho_{\mathrm{AB}}^{\otimes n}, \mathscr{K}(\mathrm{A}^n{:}\mathrm{B}^n)\big) \simeq F\big(\rho_{\mathrm{AB}}, \mathscr{K}(\mathrm{A}{:}\mathrm{B})\big)^n .$$

At first sight, one of the main difficulties seems to lie in the fact that there is much less invariance in a random tensor product state than in a random state: e.g. the distribution of $\rho^{\otimes n}$, for ρ a random state on $\mathbf{C}^d$, is not invariant under conjugation by any unitary on $(\mathbf{C}^d)^{\otimes n}$, but only under conjugation by those of the form $U^{\otimes n}$ with U a unitary on $\mathbf{C}^d$. Another major obstacle is that, even once the average behaviour is identified, asserting concentration around it is not so easy. Indeed, tiny fluctuations for ρ may have huge repercussions for $\rho^{\otimes n}$.

Note that the related question of studying random separable states has already been considered in [3], whose main result summarizes as follows: If $s \simeq d^n$, then the distribution of $\big(\sum_{i=1}^s |\psi_i^1\rangle\langle\psi_i^1| \otimes \cdots \otimes |\psi_i^n\rangle\langle\psi_i^n|\big)/s$, with the $\psi_i^1, \ldots, \psi_i^n$ independent and uniformly distributed unit vectors in $\mathbf{C}^d$, is roughly the same as the distribution of $\big(\sum_{i=1}^s |\psi_i\rangle\langle\psi_i|\big)/s$, with the ψ_i independent and uniformly distributed unit vectors in $(\mathbf{C}^d)^{\otimes n}$.

Quantifying the generic amount of entanglement in multipartite quantum systems

In Part III we looked at separability vs entanglement properties in generic high-dimensional multipartite systems from various angles. One common trait of the functions we were trying to quantify though was that they were almost always expressible as the support function of a certain convex set (of either measurements or states). Nevertheless, most more operational measures of entanglement are not of that kind, but rather of entropic form. And loosely speaking, the presence of log's in the formulas defining them makes the problem of estimating their expected value much less attackable by our techniques. This said, in the seminal exploration of typical entanglement, launched in [100], bounds were however put on the expected value of several entropic measures of correlations. In short, the main conclusion of this study was that "bound-entangled-like" features, namely both high entanglement of formation and low distillable entanglement, are in fact quite generic in large bipartite quantum systems. But being able to derive more precise estimates would be highly desirable. One possible application would be in disproving that a given entanglement measure obeys a given property (e.g. of monogamy-type, faithfulness-type etc.). Indeed, counter-examples to the long-standing additivity of minimum output entropy conjecture were precisely provided by appropriately chosen random channels [96, 17]. So it is reasonable to hope that, similarly, random states could be good candidates to rule out certain behaviours for such or such measure of entanglement.

There are nonetheless a few entanglement measures for which getting quite tight bounds (or even the exact asymptotics) for their typical value is actually possible (see [1], in collaboration with G. Adesso, S. Di Martino, M. Huber, M. Piani and A. Winter). Examples include the entanglement of formation or the relative entropy of entanglement. The paradigmatic idea could be summarized as follows: whenever a minimization/maximization has to be performed over a certain set of states, the "obvious" candidate states usually already work pretty well in high dimension. On the other hand, archetypical examples of entanglement measures which cannot be easily tackled are all the regularized ones. Indeed, for a non-additive entanglement measure E, even if the

generic behaviour of $E(\rho)$ is well understood, this is generally not the case of its regularized version $E^{\infty}(\rho) = \lim_{n\to+\infty} E(\rho^{\otimes n})/n$. This is for the reason evoked before that properties of tensor power random states remain ill homed in on.

Another category of entanglement measures, which was briefly evoked in the concluding section of Chapter 7, is that of "filtered through measurements" entanglement measures (initially introduced and studied in [147]). For the sake of concreteness, let us focus here on one specific example, based on the relative entropy distance measure $D(\cdot\|\cdot)$ between two quantum states (or two classical probability distributions), as defined in Chapter 2, Section 2.5. So now, given a state ρ on some bipartite system $\mathrm{A}\otimes\mathrm{B}$, its relative entropy of entanglement is defined as

$$D\left(\rho\|\mathcal{S}\right) = \inf_{\sigma\in\mathcal{S}} D(\rho\|\sigma),$$

and for any class of POVMs $\mathbf{M}$ on $\mathrm{A}\otimes\mathrm{B}$, its relative entropy of entanglement filtered through $\mathbf{M}$ is defined as

$$D_{\mathbf{M}}\left(\rho\|\mathcal{S}\right) = \inf_{\sigma\in\mathcal{S}} \sup_{\mathcal{M}\in\mathbf{M}} D\left(\mathcal{M}(\rho)\|\mathcal{M}(\sigma)\right).$$

One may then ask: for $\mathbf{M}$ being a class of locally restricted POVMs on $\mathrm{A}\otimes\mathrm{B}$, such as **SEP** or **PPT**, what is the generic value of $D_{\mathbf{M}}\left(\rho\|\mathcal{S}\right)$ for ρ a uniformly distributed random state on $\mathrm{A}\otimes\mathrm{B}$? Of course, we know by Pinsker's inequality that $D_{\mathbf{M}}\left(\rho\|\mathcal{S}\right) \geqslant \|\rho-\mathcal{S}\|_{\mathbf{M}}^2/(2\ln 2)$. So a lower bound on the typical $\|\rho-\mathcal{S}\|_{\mathbf{M}}$, which is something we establish in Chapter 7, immediately implies a lower bound on the typical $D_{\mathbf{M}}\left(\rho\|\mathcal{S}\right)$. But what about an upper bound? And more generally, is it possible to directly estimate the correct order of magnitude for $D_{\mathbf{M}}\left(\rho\|\mathcal{S}\right)$? In order to do so, it seems that a useful notion could be that of distinguishability ordering rather than distinguishability norm. Let us explicit a bit what we have in mind. Given a class of POVMs $\mathbf{M}$ and two states ρ,σ on $\mathrm{A}\otimes\mathrm{B}$, one can define the minimal dominating constant between ρ and σ, filtered through $\mathbf{M}$, as

$$\alpha_{\mathbf{M}}(\rho,\sigma) = \inf\left\{\alpha \ : \ \forall\, \mathcal{M}\in\mathbf{M},\ \mathcal{M}(\rho)\leqslant\alpha\,\mathcal{M}(\sigma)\right\}.$$

The reason why we claim that this is an interesting figure of merit is simply because, for two probability distributions p,q, knowing that $p\leqslant Cq$, for some constant $C>1$, implies that $D(p\|q)\leqslant\log C$. So an upper bound on $\alpha_{\mathbf{M}}(\rho,\mathcal{S})$ straightforwardly translates into an upper bound on $D_{\mathbf{M}}\left(\rho\|\mathcal{S}\right)$. Understanding the quantity $\alpha_{\mathbf{M}}(\rho,\sigma)$ requires a bit more work than understanding the quantity $\|\rho-\sigma\|_{\mathbf{M}}$, but can nevertheless be done in some cases (see a forthcoming work in collaboration with M. Christandl).

In search of a supercalifragilistic measure of entanglement

One of our main preoccupations in Chapter 10 was to see under which conditions the support function of a set of quantum states exhibits a multiplicative behaviour. As a case study, we scrutinized what happens for the set of separable states. Let us start with commenting a bit more on the relevance of this example, linking it to the topic of Chapter 5. Let P be the projector onto some subspace $\mathrm{C}\subset\mathrm{A}\otimes\mathrm{B}$, and write $P=VV^{\dagger}$, with $V:\mathrm{C}\hookrightarrow\mathrm{A}\otimes\mathrm{B}$ an isometry. Define next the quantum channel $\mathcal{N}:\mathscr{L}(\mathrm{C})\to\mathscr{L}(\mathrm{A})$ by $\mathcal{N}(X)=\mathrm{Tr}_{\mathrm{B}}(VXV^{\dagger})$. It is then easy to see that

$$\sup_{\sigma\in\mathcal{S}(\mathrm{A}:\mathrm{B})} \mathrm{Tr}(P\sigma) =: h_{\mathcal{S}(\mathrm{A}:\mathrm{B})}(P) = \|\mathcal{N}\|_{1\to\infty} := \sup_{\rho\in\mathscr{D}(\mathrm{C})} \|\mathcal{N}(\rho)\|_{\infty}.$$

Hence, asking if $h_{\mathcal{S}(\mathrm{A}^n:\mathrm{B}^n)}(P^{\otimes n}) \simeq \left(h_{\mathcal{S}(\mathrm{A}:\mathrm{B})}(P)\right)^n$ is nothing else than asking if $\|\mathcal{N}^{\otimes n}\|_{1\to\infty} \simeq \left(\|\mathcal{N}\|_{1\to\infty}\right)^n$, i.e. equivalently if $S_{\infty}^{\min}(\mathcal{N}^{\otimes n}) \simeq n\, S_{\infty}^{\min}(\mathcal{N})$. It is now a well-known fact that the so-called additivity conjecture for $S_{\infty}^{\min}$ is false: inequality can be strict in general between $S_{\infty}^{\min}(\mathcal{N}^{\otimes 2})$ and $2\,S_{\infty}^{\min}(\mathcal{N})$, with even drastic examples where $S_{\infty}^{\min}(\mathcal{N}^{\otimes 2}) \sim S_{\infty}^{\min}(\mathcal{N})$. The two archetypical cases for which this happens are when the associated projector P is either the projector onto the anti-symmetric subspace (see [86]) or a random projector (see [101]). However, it has also been proved more recently that, precisely in these two situations, the extreme 2-copy behaviour does not reflect the n-copy behaviour for large n, where weak additivity of $S_{\infty}^{\min}(\mathcal{N})$ actually holds (see [50] and [141]). Hence the natural question of whether this could not in fact be what occurs for any quantum channel $\mathcal{N}$.

Chapter 10 put forward two ways of showing a weakly multiplicative behaviour of $h_{\mathcal{S}(\mathrm{A}:\mathrm{B})}(P)$. Indeed, let $\rho^{(n)}$ be a permutation-invariant state on $(\mathrm{A}\otimes\mathrm{B})^{\otimes n}$ which is separable across the cut $\mathrm{A}^{\otimes n}$:$\mathrm{B}^{\otimes n}$, and assume that you want to upper bound $\mathrm{Tr}(P^{\otimes n}\rho^{(n)})$. Then, you can follow either the "global" strategy of using a de

Finetti reduction for $\rho^{(n)}$ (cf. Section 10.3), or the more "local" strategy of iteratively projecting $\rho^{(n)}$ on $\mathrm{A}\otimes\mathrm{B}$ and working with the conditional state on the remaining subsystems (cf. Section 10.4). The two approaches can be seen through, but with the annoying feature of eventually yielding a dimensional dependence in the exponentially decaying bound. Concretely, setting $d = \max(|\mathrm{A}|, |\mathrm{B}|)$, one can obtain a statement of the form:

$$h_{\mathcal{S}(\mathrm{A}:\mathrm{B})}(P) \leqslant 1-\varepsilon \;\Rightarrow\; h_{\mathcal{S}(\mathrm{A}^n:\mathrm{B}^n)}(P^{\otimes n}) \leqslant (1-\delta(\varepsilon,d))^n, \tag{11.9}$$

with δ a function which is non-decreasing in ε and non-increasing in d, vanishing for $\varepsilon = 0$ or $d \to +\infty$. Now, it could well be that equation (11.9) actually holds with a dimensionally independent function $\delta(\varepsilon)$.

In order to prove something is that direction via the "local" technique, it would be enough to find a function E, quantifying the amount of entanglement in a bipartite quantum state, which would satisfy the two following conditions:

- Monogamy-like property: $I(\mathrm{AA}':\mathrm{BB}')(\rho_{\mathrm{AA}'\mathrm{BB}'}) \geqslant E(\rho_{\mathrm{AB}}) + E(\rho_{\mathrm{A}'\mathrm{B}'})$ (see Chapter 2, Section 2.5, for the definition of the mutual information I).
- Strong faithfulness property: $E(\rho_{\mathrm{AB}}) \leqslant \varepsilon \;\Rightarrow\; \|\rho_{\mathrm{AB}} - \mathcal{S}(\mathrm{A}:\mathrm{B})\|_1 \leqslant \delta(\varepsilon)$, with δ a universal function (i.e. in particular not depending on $|\mathrm{A}|$ and $|\mathrm{B}|$).

The reason why this is not so easy is because monogamy and faithfulness are two properties of entanglement measures which usually exclude one another (for a more precise formulation, see [1], in collaboration with G. Adesso, S. Di Martino, M. Huber, M. Piani and A. Winter). One candidate though for an entanglement measure which would combine the two above mentioned features is the conditional entanglement of mutual information introduced in [109] (cf. Chapter 10, Section 10.4).

We already mentioned the one-to-one correspondence between estimating $h_{\mathcal{S}}$ and estimating $S_\infty^{\min}$. There is another important one, namely with estimating the acceptance probability in a QMA(2) protocol with unentangled provers. Also out of purely quantum information scope, there is a plethora of problems which are equivalent to pulling apart the two options $h_{\mathcal{S}(\mathrm{A}:\mathrm{B})}(P) = 1$ and $h_{\mathcal{S}(\mathrm{A}:\mathrm{B})}(P) < 1$: distinguishing $\|\psi\|_{\mathrm{inj}} = 1$ from $\|\psi\|_{\mathrm{inj}} < 1$ for $\psi \in \mathrm{A}\otimes\mathrm{B}\otimes\mathrm{C}$ such that $\|\psi\| \leqslant 1$, distinguishing $\|M\|_{2\to 4} = 1$ from $\|M\|_{2\to 4} < 1$ for $M : \mathrm{C} \to \mathrm{A}\otimes\mathrm{B}$ such that $\|M\|_\infty \leqslant 1$ etc. The reader is referred to [94] for the precise definitions of a QMA(2) protocol, the injective tensor norm, the $(2\to 4)$-norm, and for a full list of other (roughly) equivalent discrimination problems. Anyway, all this to say that having an efficient way of amplifying an ε-gap in $h_{\mathcal{S}}$ by parallel repetition would have significant implications!

Relaxations of separability or absolute separability and related complexity issues

As was highlighted in many different ways over these pages, any simple necessary condition for separability (meaning that there exists an efficient way of checking it) is doomed to become very rough as the dimensions of the local Hilbert spaces grow. Let us focus here on two such relaxations of separability, which are of very distinct nature, even though deciding membership can be cast as an SDP for both of them: positivity under partial transposition and k-extendibility. The PPT criterion is believed to be, in several senses, the best amongst separability criteria based on positive maps. Nonetheless, it is not even known whether the set of PPT states has a mean width and a volume radius which asymptotically differ from those of the set of all states. On the contrary, concerning the set of k-extendible states, Theorem 9.2.5 in Chapter 9, Section 9.2, tells us that

$$\forall\, k \in \mathbf{N},\ \mathrm{vrad}\left(\mathscr{E}_k(\mathbf{C}^d:\mathbf{C}^d)\right) \leqslant w\left(\mathscr{E}_k(\mathbf{C}^d:\mathbf{C}^d)\right) \underset{d\to+\infty}{\sim} \frac{2}{\sqrt{k}}\frac{1}{d}.$$

Now, we just have to recall that for the set of all states we have

$$w\left(\mathscr{D}(\mathbf{C}^d\otimes\mathbf{C}^d)\right) \underset{d\to+\infty}{\sim} \frac{2}{d} \text{ and } \mathrm{vrad}\left(\mathscr{D}(\mathbf{C}^d\otimes\mathbf{C}^d)\right) \underset{d\to+\infty}{\sim} \frac{e^{-1/4}}{d}.$$

And we thus get the existence of a constant $0 < c_0 < 1$ and a dimension $d_0 \in \mathbf{N}$ such that, for any $d \geqslant d_0$,

$$\forall\, k \geqslant 2,\ w\left(\mathscr{E}_k(\mathbf{C}^d:\mathbf{C}^d)\right) \leqslant c_0\, w\left(\mathscr{D}(\mathbf{C}^d\otimes\mathbf{C}^d)\right) \text{ and } \forall\, k \geqslant 7,\ \mathrm{vrad}\left(\mathscr{E}_k(\mathbf{C}^d:\mathbf{C}^d)\right) \leqslant c_0\, \mathrm{vrad}\left(\mathscr{D}(\mathbf{C}^d\otimes\mathbf{C}^d)\right).$$

Hence to summarize: for a fixed $k \geqslant 2$, the size of $\mathscr{E}_k(\mathbf{C}^d:\mathbf{C}^d)$ differs from that of $\mathscr{D}(\mathbf{C}^d\otimes\mathbf{C}^d)$ only by an absolute factor, but one can at least certify that the latter is strictly smaller than 1 (for all $k \geqslant 2$ if size is

measured in terms of mean width, and for sure as soon as $k \geqslant 7$ if size is measured in terms of volume radius). Whereas oppositely, it could be that the ratio between the size of $\mathscr{P}(\mathbf{C}^d{:}\mathbf{C}^d)$ and that of $\mathscr{D}(\mathbf{C}^d \otimes \mathbf{C}^d)$ approaches 1 as d increases. It may be worth pointing out that, when volume radius is used as size parameter, this problem is equivalent to: does there exist an absolute constant $c > 0$ such that

$$\forall\ d \in \mathbf{N},\ \mathbf{P}_{\rho \sim \mu_{d^2,d^2}} \left(\rho \in \mathscr{P}(\mathbf{C}^d{:}\mathbf{C}^d) \right) \leqslant e^{-cd^4}\ ?$$

As a side technical comment, let us mention that, for the case of k-extendibility, we already have as a corollary of Theorem 9.6.4 in Chapter 9, Section 9.6, that

$$\forall\ k \geqslant 6,\ \exists\ c_k > 0 :\ \forall\ d \in \mathbf{N},\ \mathbf{P}_{\rho \sim \mu_{d^2,d^2}} \left(\rho \in \mathscr{E}_k(\mathbf{C}^d{:}\mathbf{C}^d) \right) \leqslant e^{-c_k d^2}.$$

But by all the remarks above, we see that (at least for $k \geqslant 7$) this probability estimate is far from optimal: the dimensional dependence in the concentration bound can actually be improved from d^2 to d^4.

However complicated to characterize the set of separable states is, it could be that its so-called *absolute* version is much easier to grasp. The latter, initially introduced in [125], is defined as the set of states which remain separable under conjugation by any unitary, i.e. it encompasses a notion of separability for states which only depends on their eigenvalues, and not on their eigenvectors. Similarly, one can define, for each separability criterion, the set of states which absolutely satisfy it. Quite surprisingly, there are strong evidences towards the fact that, for instance, being absolutely PPT could already be equivalent to being absolutely separable. This conjecture is known to hold in the special case of $\mathbf{C}^2 \otimes \mathbf{C}^d$, for any $d \in \mathbf{N}$ [119]. If it were true in general, it would imply an extremely simple description of absolutely separable states, as a finite number of semidefinite conditions [103]. With this kind of motivations in mind, trying to understand the hierarchy of absolutely k-extendible states looks like an interesting project: does it collapse to absolutely separable states for a finite k? and if so a dimension dependent or independent k? etc.

Typical degradability or anti-degradability of quantum channels

Let $\mathcal{N} : \mathscr{L}(\mathrm{A}) \to \mathscr{L}(\mathrm{B})$ be a quantum channel, such that $\mathcal{N}(X) = \mathrm{Tr}_{\mathrm{E}}(VXV^\dagger)$ for some isometry $V : \mathrm{A} \hookrightarrow \mathrm{B} \otimes \mathrm{E}$. Its complementary channel is then the quantum channel $\mathcal{N}^c : \mathscr{L}(\mathrm{A}) \to \mathscr{L}(\mathrm{E})$ defined by $\mathcal{N}^c(X) = \mathrm{Tr}_{\mathrm{B}}(VXV^\dagger)$. $\mathcal{N}$ is called *degradable* if there exists a quantum channel $\mathcal{F} : \mathscr{L}(\mathrm{B}) \to \mathscr{L}(\mathrm{E})$ such that $\mathcal{N}^c = \mathcal{F} \circ \mathcal{N}$, and it is called *anti-degradable* if $\mathcal{N}^c$ is degradable (i.e. if there exists a quantum channel $\mathcal{F} : \mathscr{L}(\mathrm{E}) \to \mathscr{L}(\mathrm{B})$ such that $\mathcal{N} = \mathcal{F} \circ \mathcal{N}^c$). These channels are interesting from many respects. One main feature of theirs is that an anti-degradable channel does not have any quantum capacity, while a degradable channel has an additive coherent information, and therefore a quantum capacity admitting a single-letter formula. In [152], an approximate notion of (anti-)degradability was more generally defined. And it was shown there that, with this definition, the smallest ε such that a channel $\mathcal{N}$ is ε-anti-degradable can be used to give an upper bound on the quantum capacity of $\mathcal{N}$. This is an important result because this minimal ε can be efficiently determined (by solving an SDP), hence providing an easily computable upper bound on the quantum capacity, a quantity for which no method to even just estimate it is known in general.

Besides, another nice fact about (anti-)degradability is that it is a property which can be easily characterized at the level of the Choi states of either $\mathcal{N}$ or $\mathcal{N}^c$. Specifically, $\mathcal{N}$ is anti-degradable if and only if $\tau(\mathcal{N})$ is a 2-extendible state, and thus conversely, $\mathcal{N}$ is degradable if and only if $\tau(\mathcal{N}^c)$ is a 2-extendible state (cf. Chapter 9). So now, assume that $\mathcal{N} : \mathscr{L}(\mathrm{A}) \to \mathscr{L}(\mathrm{B})$ is a random quantum channel, defined by $\mathcal{N}(X) = \mathrm{Tr}_{\mathrm{E}}(VXV^\dagger)$ for $V : \mathrm{A} \hookrightarrow \mathrm{B} \otimes \mathrm{E}$ a uniformly distributed isometry. Then, as $|\mathrm{A}|, |\mathrm{B}|, |\mathrm{E}|$ grow, $\tau(\mathcal{N})$ is more or less distributed as a random state on $\mathrm{A} \otimes \mathrm{B}$ induced by some environment E (this hand-waving claim can be made mathematically precise). Hence, understanding when such random channels are typically (anti-)degradable or not roughly boils down to understanding when random-induced states are typically 2-extendible or not. And this is now something that we know quite well how to do thanks to the techniques developed in Chapter 9. An interesting question at that point would therefore be: can this approach be extended in order to estimate, given a random channel, its minimal ε-anti-degradability (and subsequently an upper bound on its quantum capacity)?

There are several reasons why this does not look like a completely straightforward affair. Characterizing when a random channel is with high probability exactly (anti-)degradable is probably not the hardest part: from the results of Chapter 9, one can reasonably expect that there should be a threshold for anti-degradability vs non-anti-degradability at $|\mathrm{E}| = c|\mathrm{A}||\mathrm{B}|$, for some universal constant $c > 0$ (and correspondingly, one for degradability vs non-degradability at $|\mathrm{B}| = c|\mathrm{A}||\mathrm{E}|$). The main difficulty rather comes from the fact that the

notion of ε-(anti-)degradability introduced in [152] is not simply being ε-close to an exactly (anti-)degradable channel, but instead being ε-close to satisfying the (anti-)degradability condition. It is a problem we already stumbled upon in Chapters 10 and 11 (in a totally different context) that, even though sounding like a sensible hope, it is not always true that the latter (for some ε) implies the former (for some $\delta(\varepsilon)$).

Bibliography

[1] G. Adesso, S. Di Martino, M. Huber, C. Lancien, M. Piani, and A. Winter. Should entanglement measures be monogamous or faithful? arXiv[quant-ph]:1604.02189.

[2] A. Ambainis and J. Emerson. Quantum t-designs: t-wise independence in the quantum world. In *Proc. 22nd IEEE Conference on Computational Complexity*, pages 129–140, Piscataway, NJ, 2007.

[3] A. Ambainis, A.W. Harrow, and M.B. Hastings. Random tensor theory: extending random matrix theory to random product states. *Commun. Math. Phys.*, 310(1):25–74, 2012.

[4] G.W. Anderson, A. Guionnet, and O. Zeitouni. *An Introduction to Random Matrices*. Cambridge Studies in Advanced Mathematics, Vol. 118. Cambridge University Press, Cambridge, 2010.

[5] H. Araki and E.H. Lieb. Entropy inequalities. *Commun. Math. Phys.*, 18:160–170, 1970.

[6] R. Arnon-Friedman and R. Renner. de Finetti reductions for correlations. *J. Math. Phys.*, 56(052203), 2015.

[7] R. Arnon-Friedman, R. Renner, and T. Vidick. Non-signalling parallel repetition using de Finetti reductions. *IEEE Transactions on Information Theory*, 62(3), 2016.

[8] G. Aubrun. An inequality about the largest eigenvalue of a random matrix. *Séminaire de Probabilités XXXVIII Lecture Notes in Math.*, 1857:320–337, 2005.

[9] G. Aubrun. On almost randomizing channels with a short Kraus decomposition. *Commun. Math. Phys.*, 288(3):1103–1116, 2009.

[10] G. Aubrun. Partial transposition of random states and non-centered semicircular distributions. *Random Matrices Theory Appl.*, 1(1250001), 2012.

[11] G. Aubrun and C. Lancien. Locally restricted measurements on a multipartite quantum system: data hiding is generic. *Quant. Inf. Comput.*, 15(5-6):512–540, 2014.

[12] G. Aubrun and C. Lancien. Zonoids and sparsification of quantum measurements. *Positivity*, 20(1):1–23, 2016.

[13] G. Aubrun and I. Nechita. Realigning random states. *J. Math. Phys.*, 53(102210), 2012.

[14] G. Aubrun and S.J. Szarek. *Alice and Bob meet Banach.* Book available at `http://math.univ-lyon1.fr/~aubrun/ABMB/index.html`.

[15] G. Aubrun and S.J. Szarek. Dvoretzky's theorem and the complexity of entanglement detection. arXiv[quant-ph]:1510.00578.

[16] G. Aubrun and S.J. Szarek. Tensor product of convex sets and the volume of separable states on n qudits. *Phys. Rev. A.*, 73(022109), 2006.

[17] G. Aubrun, S.J. Szarek, and E. Werner. Hastings' additivity counterexample via Dvoretzky's theorem. *Commun. Math. Phys.*, 305(1):85–97, 2011.

[18] G. Aubrun, S.J. Szarek, and D. Ye. Entanglement thresholds for random induced states. *Commun. Pure App. Math.*, 67(1):129–171, 2013.

[19] K.R.M. Audenaert. A sharp continuity estimate for the von Neumann entropy. *J. Phys. A: Math. Theor.*, 40:8127–8136, 2007.

[20] K. Ball. *An elementary introduction to modern convex geometry.* Flavors of geometry. Cambridge University Press, Cambridge, 1997.

[21] J. Barrett, N. Linden, S. Massar, S. Pironio, S. Popescu, and D. Roberts. Non-local correlations as an information theoretic resource. *Phys. Rev. A*, 71(022101), 2005.

[22] A. Barvinok. *A course in convexity.* Graduate studies in Mathematics, Vol. 54. American Mathematical Society, Providence, 2002.

[23] J. Batson, D.A. Spielman, and N. Srivatsava. Twice-Ramanujan sparsifiers. *SIAM J. Comput.*, 41(6):1704–1721, 2012.

[24] S. Bäuml, M. Christandl, K. Horodecki, and A. Winter. Limitations on quantum-key repeaters. *Nature Commun.*, 6(6908), 2015.

[25] S. Beigi and P.W. Shor. Approximating the set of separable states using the positive partial transpose test. *J. Math. Phys.*, 51(042202), 2010.

[26] J. Bell. On the Einstein Podolsky Rosen paradox. *Physics*, 1(195), 1964.

[27] I. Bengtsson and K. Życzkowski. *Geometry of Quantum States: An Introduction to Quantum Entanglement.* Cambridge University Press, Cambridge, 2006.

[28] G. Bennett. Schur multipliers. *Duke Math. J.*, 44(3):603–639, 1977.

[29] M. Berta, M. Christandl, and R. Renner. The quantum reverse Shannon theorem based on one-shot information theory. *Commun. Math. Phys.*, 306(3):579–615, 2011.

[30] R. Bhatia. *Matrix analysis.* Graduate Texts in Mathematics, Vol. 169. Springer-Verlag, New-York, 1997.

[31] E.D. Bolker. A class of convex bodies. *Trans. AMS*, 145:323–345, 1969.

[32] J. Bourgain, J. Lindenstrauss, and V.D. Milman. Approximation of zonoids by zonotopes. *Acta Math.*, 162(1), 1989.

[33] J. Bourgain and V.D. Milman. New volume ratio properties for convex symmetric bodies in $\mathbf{R}^n$. *Invent. Math.*, 88(2):319–340, 1987.

[34] F.G.S.L. Brandão, M. Christandl, and J.T. Yard. Faithful squashed entanglement. *Commun. Math. Phys.*, 306(3):805–830, 2011.

[35] F.G.S.L. Brandão and A. Harrow. Quantum de Finetti theorems under local measurements with applications. In *Proc. 45th ACM Symposium on Theory of Computing*, pages 861–870, New-York, 2013.

[36] F.G.S.L. Brandao, A.W. Harrow, and M. Horodecki. Local random quantum circuits are approximate polynomial-designs. arXiv[quant-ph]:1208.0692.

[37] F.G.S.L. Brandao and R.O. Vianna. A robust semidefinite programming approach to the separability problem. *Phys. Rev. A*, 70(062309), 2004.

[38] F.G.S.L. Brandao and R.O. Vianna. Separable multipartite mixed states - Operational asymptotically necessary and sufficient conditions. *Phys. Rev. Lett.*, 93(220503), 2004.

[39] H. Buhrman, N. Chandran, S. Fehr, R. Gelles, V. Goyal, R. Ostrovsky, and C. Schaffner. *Position-Based Quantum Cryptography: Impossibility and Constructions*, volume 6841 of *Lecture Notes in Computer Science*, pages 429–446. Springer-Verlag, Berlin, 2011.

[40] H. Buhrman, S. Fehr, and C. Schaffner. On the parallel repetition of multi-player games: The no-signaling case. In *Proc. 9th Conference on the Theory of Quantum Computation, Communication and Cryptography*, volume 27, pages 24–35, 2014.

[41] N.J. Cerf, R. García-Patrón, A. Leverrier, and R. Renner. Security of continuous-variable quantum key distribution against general attacks. *Phys. Rev. Lett.*, 110(030502), 2013.

[42] D. Chafaï, O. Guédon, G. Lecué, and A. Pajor. *Interactions between compressed sensing, random matrices and high dimensional geometry*. Panoramas et synthèse 37. Société Mathématique de France, Paris, 2012.

[43] A. Chailloux and G. Scarpa. Parallel repetition of free entangled games: simplification and improvements. arXiv[quant-ph]:1410.4397.

[44] K. Chen and L.A. Wu. A matrix realignment method for recognizing entanglement. *Quant. Inf. Comput.*, 3(3):193–202, 2003.

[45] E. Chitambar and M-H. Hsieh. Asymptotic state discrimination and a strict hierarchy in distinguishability norms. *J. Math. Phys.*, 55(112204), 2014.

[46] M. Choi. Completely positive linear maps on complex matrices. *Lin. Alg. and its App.*, 10:285–290, 1975.

[47] M.D. Choi and T.Y. Lam. Extremal positive semidefinite forms. *Math. Ann.*, 231(1), 1977.

[48] M. Christandl, R. König, G. Mitchison, and R. Renner. One-and-a-half quantum de Finetti theorems. *Commun. Math. Phys.*, 273(2):473–498, 2007.

[49] M. Christandl, R. König, and R. Renner. Post-selection technique for quantum channels with applications to quantum cryptography. *Phys. Rev. Lett.*, 102(020504), 2009.

[50] M. Christandl, N. Schuch, and A. Winter. Entanglement of the antisymmetric state. *Commun. Math. Phys.*, 311(2):397–422, 2012.

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

[52] M.K. Chung, X. Wu, and H. Yuen. Parallel repetition for entangled k-player games via fast quantum search. arXiv[quant-ph]:1501.00033.

[53] J.I. Cirac and R. Renner. A de Finetti representation theorem for infinite dimensional quantum systems and applications to quantum cryptography. *Phys. Rev. Lett.*, 102(110504), 2009.

[54] J.F. Clauser, R.A. Holt, M.A. Horne, and A. Shimony. Proposed experiment to test local hidden-variable theories. *Phys. Rev. Lett.*, 23(15):880–884, 1969.

[55] R. Cleve, W. Slofstra, F. Unger, and S. Upadhyay. Strong parallel repetition theorem for quantum XOR proof systems. *Comput. Complexity*, 17(2):282–299, 2008.

[56] B. Collins and I. Nechita. Gaussianization and eigenvalue statistics for random quantum channels (iii). *Ann. Appl. Probab.*, 21(3):1136–1179, 2011.

[57] M. Curty, J. Eisert, O. Gühne, and P. Hyllus. Complete hierarchies of efficient approximations to problems in entanglement theory. *Phys. Rev. A*, 70(062317), 2004.

[58] J. de Pillis. Linear transformations which preserve hermitian and positive semidefinite operators. *Pacific J. Math.*, 23(129), 1967.

[59] A. Defant and C. Michels. Norms of tensor product identities. *Note Mat.*, 25(1):129–166, 2006.

[60] P. Diaconis and D. Freedman. Finite exchangeable sequences. *Ann. Probab.*, 8(4):745–764, 1980.

[61] I. Dinur, D. Steurer, and T. Vidick. A parallel repetition theorem for entangled projection games. *Comput. Complexity*, 24(2):201–254, 2015.

[62] D.P. DiVincenzo, M. Horodecki, D. Leung, J. Smolin, and B.M. Terhal. Locking classical correlation in quantum states. *Phys. Rev. Lett.*, 92(067902), 2004.

[63] D.P. DiVincenzo, D. Leung, and B.M. Terhal. Hiding bits in Bell states. *Phys. Rev. Lett.*, 86(25):5807–5810, 2001.

[64] D.P. DiVincenzo, D. Leung, and B.M. Terhal. Quantum data hiding. *IEEE Trans. Inform. Theory*, 48(3):580–599, 2002.

[65] A.C. Doherty. Entanglement and the shareability of quantum states. *J. Phys. A: Math. Theor.*, 47(42), 2014.

[66] A.C. Doherty, P.A. Parrilo, and F.M. Spedalieri. A complete family of separability criteria. *Phys. Rev. A.*, 69(022308), 2004.

[67] R. Duan, S. Severini, and A. Winter. On zero-error communication via quantum channels in the presence of noiseless feedback. *IEEE Trans. Inform. Theory*, 59(2):1164–1174, 2012.

[68] F. Dupuis, J. Florjanczyk, P. Hayden, and D. Leung. Locking classical information. *Proc. Roy. Soc. Edinburgh Sect. A*, 469(2159), 2013.

[69] T. Eggeling and R.F. Werner. Hiding classical data in multi-partite quantum states. *Phys. Rev. Lett.*, 89(097905), 2002.

[70] C. Eltschka and J. Siewert. Quantifying entanglement resources. *J. Phys. A: Math. Theor.*, 47(424005), 2014.

[71] O. Fawzi, P. Hayden, and P. Sen. From low-distortion norm embeddings to explicit uncertainty relations and efficient information locking. *Journal of the ACM*, 60(6), 2013.

[72] O. Fawzi and R. Renner. Quantum conditional mutual information and approximate Markov chains. *Commun. Math. Phys.*, 340(2):575–611, 2015.

[73] S. Fehr, J. Kaniewski, M. Tomamichel, and S. Wehner. A monogamy-of-entanglement game with applications to device-independent quantum cryptography. *New J. Phys.*, 15(103002), 2013.

[74] U. Feige. On the success probability of the two provers in one-round proof systems. In *Proc. 6th IEEE Symposium on Structure in Complexity Theory*, pages 116–123, 1991.

[75] U. Feige and O. Verbitsky. Error reduction by parallel repetition – a negative result. *Combinatorica*, 2(4):461–478, 2002.

[76] T. Figiel and W.B. Johnson. Large subspaces of ℓ_∞^n and estimates of the Gordon–Lewis constant. *Israel J. Math.*, 37(1-2), 1980.

[77] T. Figiel, J. Lindenstrauss, and V.D. Milman. The dimension of almost spherical sections of convex bodies. *Acta Math.*, 139(1-2), 1977.

[78] C.A. Fuchs and J. van de Graaf. Cryptographic distinguishability measures for quantum-mechanical states. *IEEE Trans. Inform. Theory*, 45(4):1216–1227, 1999.

[79] M. Fukuda and I. Nechita. Asymptotically well-behaved input states do not violate additivity for conjugate pairs of random quantum channels. *Commun. Math. Phys.*, 328(3):995–1021, 2014.

[80] S. Gharibian. Strong NP-hardness of the separability problem. *Quant. Inf. Comput.*, 10(3-4):343–360, 2010.

[81] V. Giovannetti and A.S. Holevo. Quantum channels and their entropic characteristics. *Rep. Prog. Phys.*, 75(046001), 2012.

[82] P. Goodey and W. Weil. Zonoids and generalizations. *Handbook of Convex Geometry*, B:1269–1326, 1993.

[83] Y. Gordon. Some inequalities for Gaussian processes and applications. *Israel J. Math.*, 50(4):265–289, 1985.

[84] I.P. Goulden and D.M. Jackson. Connexion coefficients for the symmetric group, free products in operator algebras and random matrices. *Fields Inst. Commun.*, 12:105–125, 1997.

[85] A. Goupil and G. Schaeffer. Factoring n-cycles and counting maps of given genus. *European. J. Combin.*, 19:819–834, 1998.

[86] A. Grudka, M. Horodecki, and Ł. Pankowski. Constructive counterexamples to additivity of minimum output Rényi entropy of quantum channels for all $p > 2$. *J. Phys. A: Math. Theor.*, 43(425304), 2010.

[87] O. Gühne, M. Huber, C. Lancien, and R. Sengupta. Relaxations of separability in multipartite systems: semidefinite programs, witnesses and volumes. *J. Phys. A: Math. Theor.*, 48(505302), 2015.

[88] O. Gühne, B. Jungnitsch, and T. Moroder. Taming multiparticle entanglement. *Phys. Rev. Lett.*, 106(190502), 2011.

[89] O. Gühne and G. Tóth. Entanglement detection. *Phys. Rep.*, 474(1), 2009.

[90] L. Gurvits. Classical deterministic complexity of edmond's problem and quantum entanglement. In *Proc. 35th ACM Symposium on Theory of Computing*, pages 10–19, New-York, 2003.

[91] U. Haagerup and S. Thorbjørnsen. A new application of random matrices: $\mathrm{Ext}\,(C^*_{\mathrm{red}}(F_2))$ is not a group. *Ann. of Math.*, 162(2):711–775, 2005.

[92] E. Hänggi. *Device-independent quantum key distribution*. PhD thesis, ETH Zürich, 2010.

[93] A.W. Harrow. The church of the symmetric subspace. arXiv[quant-ph]:1308.6595.

[94] A.W. Harrow and A. Montanaro. Testing product states, quantum Merlin-Arthur games and tensor optimisation. *Journal of the ACM*, 60(1), 2013.

[95] A.W. Harrow, A. Montanaro, and A.J. Short. Limitations on quantum dimensionality reduction. *Int. J. Quantum Inf.*, 13(4), 2015.

[96] M.B. Hastings. Superadditivity of communication capacity using entangled inputs. *Nature Physics*, 2009.

[97] M. Hayashi. Optimal sequence of POVMs in the sense of stein's lemma in quantum hypothesis testing. *J. Phys. A: Math. Gen.*, 35(5):10759–10773, 2002.

[98] M. Hayashi and T. Ogawa. On error exponents in quantum hypothesis testing. *IEEE Trans. Inf.*, 50(6):1368–1372, 2004.

[99] P. Hayden, D. Leung, P. Shor, and A. Winter. Randomizing quantum states: constructions and applications. *Commun. Math. Phys.*, 250(2):371–391, 2004.

[100] P. Hayden, D. Leung, and A. Winter. Aspects of generic entanglement. *Commun. Math. Phys.*, 265(1):95–117, 2006.

[101] P. Hayden and A. Winter. Counterexamples to the maximal p-norm multiplicativity conjecture for all $p > 1$. *Commun. Math. Phys.*, 284(1):263–280, 2008.

[102] C.W. Helstrom. Quantum detection and estimation theory. *J. Statist. Phys.*, 1:231–252, 1969.

[103] R. Hildebrand. Positive partial transpose from spectra. *Phys. Rev. A*, 76(052325), 2007.

[104] T. Holenstein. Parallel repetition: simplifications and the no-signaling case. *Theory Comput.*, 5(1):141–172, 2009.

[105] A.S. Holevo. Statistical decision theory for quantum systems. *J. Multivariate Anal.*, 3:337–394, 1973.

[106] K. Horodecki, M. Horodecki, P. Horodecki, and R. Horodecki. Quantum entanglement. *Rev. Mod. Phys.*, 81(865), 2009.

[107] M. Horodecki and P. Horodecki. Reduction criterion for separability and limits for a class of protocols of entanglement didtillation. *Phys. Rev. A.*, 59:4206–4212, 1999.

[108] M. Horodecki, P. Horodecki, and R. Horodecki. Separability of mixed states: necessary and sufficient conditions. *Phys. Lett. A*, 223(1), 1996.

[109] M. Horodecki, Z.D. Wang, and D. Yang. An additive and operational entanglement measure: conditional entanglement of mutual information. *Phys. Rev. Lett.*, 101(140501), 2008.

[110] P. Horodecki, M. Lewenstein, A. Sanpera, and K. Życzkowski. Volume of the set of separable states. *Phys. Rev. A*, 58(883), 1998.

[111] M. Huber and J.I. de Vicente. The structure of multidimensional entanglement in multipartite systems. *Phys. Rev. Lett.*, 110(030501), 2013.

[112] M. Huber and R. Sengupta. Witnessing genuine multipartite entanglement with positive maps. *Phys. Rev. Lett.*, 113(100501), 2014.

[113] P. Indyk. Uncertainty principles, extractors, and explicit embeddings of ℓ_2 into ℓ_1. In *Proc. 39th ACM Symposium on Theory of Computing*, pages 615–620, New-York, 2007.

[114] P. Indyk and S.J. Szarek. Almost-Euclidean subspaces of ℓ_1^n via tensor products: a simple approach to randomness reduction. In *RANDOM*, volume 6302, pages 632–641, Berlin Heidelberg, 2010. Springer-Verlag.

[115] T. Ito. Polynomial-space approximation of no-signaling provers. *Automata, Languages and Programming*, 6198:140–151, 2010.

[116] R. Jain, A. Pereszlenyi, and P. Yao. A parallel repetition theorem for entangled two-player one-round games under product distributions. arXiv[quant-ph]:1311.6309.

[117] A. Jamiołkowski. Linear transformations which preserve trace and positive semidefiniteness of operators. *Rep. Math. Phys.*, 3(275), 1972.

[118] M.A. Jivulescu, N. Lupa, and I. Nechita. Thresholds for reduction-related entanglement criteria in quantum information theory. *Quant. Inf. Comput.*, 15(13-14):1165–1184, 2015.

[119] N. Johnston. Separability from spectrum for qubit-qudit states. *Phys. Rev. A*, 88(062330), 2013.

[120] M. Junge, R. Renner, D. Sutter, M.M. Wilde, and A. Winter. Universal recovery from a decrease of quantum relative entropy. arXiv[quant-ph]:1509.07127.

[121] J. Kempe and O. Regev. No strong parallel repetition with entangled and non-signaling provers. In *Proc. 25th IEEE Conference on Computational Complexity*, pages 7–15, 2010.

[122] J. Kempe, O. Regev, and B. Toner. Unique games with entangled provers are easy. In *Proc. 49th IEEE Symposium on Foundations of Computer Science*, pages 457–466, 2008.

[123] J. Kempe and T. Vidick. Parallel repetition of entangled games. In *Proc. 43rd ACM Symposium on Theory of Computing*, pages 353–362, 2011.

[124] R. König and R. Renner. A de Finetti representation for finite symmetric quantum states. *J. Math. Phys.*, 46(122108), 2005.

[125] M. Kuś and K. Życzkowski. Geometry of entangled states. *Phys. Rev. A*, 63(032307), 2001.

[126] C. Lancien. k-extendibility of high-dimensional bipartite quantum states. arXiv[quant-ph]:1504.06459.

[127] C. Lancien. Quantum channel compression. In preparation.

[128] C. Lancien and A. Winter. Flexible constrained de Finetti reductions and applications. arXiv[quant-ph]:1605.09013.

[129] C. Lancien and A. Winter. Parallel repetition and concentration for (sub-)no-signalling games via a flexible constrained de Finetti reduction. arXiv[quant-ph]:1506.07002.

[130] C. Lancien and A. Winter. Distinguishing multi-partite states by local measurements. *Commun. Math. Phys.*, 323:555–573, 2013.

[131] M. Ledoux and M. Talagrand. *Probability in Banach Spaces: isoperimetry and processes*. Ergebnisse der Mathematik und ihrer Grenzgebiete, Vol. 23. Springer-Verlag, Berlin Heidelberg, 1991.

[132] P. Lévy. *Problèmes concrets d'analyse fonctionnelle*. 2nd ed. Gauthier-Villars, Paris, 1951. (in French).

[133] K. Li and A. Winter. Squashed entanglement, k-extendibility, quantum Markov chains, and recovery maps. arXiv[quant-ph]:1410.4184.

[134] E.H. Lieb and M.B. Ruskai. Proof of the strong subadditivity of quantum-mechanical entropy. *J. Math. Phys.*, 14:1938–1941, 1973.

[135] J. Löfberg. YALMIP: A toolbox for modeling and optimization in Matlab. In *Proc. CACSD Conference*, Taipei, Taiwan, 2004.

[136] S. Lovett and S. Sodin. Almost Euclidean sections of the N-dimensional cross-polytope using $O(N)$ random bits. *Commun. Contemp. Math.*, 10(4):477–489, 2008.

[137] W. Matthews, S. Wehner, and A. Winter. Distinguishability of quantum states under restricted families of measurements with an application to data hiding. *Commun. Math. Phys.*, 291(3):813–843, 2009.

[138] E. Meckes and M. Meckes. Spectral measures of powers of random matrices. *Electron. Commun. Probab.*, 18(78):1–13, 2013.

[139] V.D. Milman. A new proof of the theorem of A. Dvoretzky on sections of convex bodies. *Funct. Annal. Appl.*, 5:28–37, 1971.

[140] V.D. Milman and A. Pajor. Entropy and asymptotic geometry of non-symmetric convex bodies. *Adv. Math.*, 152:314–335, 2000.

[141] A. Montanaro. Weak multiplicativity for random quantum channels. *Commun. Math. Phys.*, 319(2):535–555, 2013.

[142] M. Navascués, M. Owari, and M.B. Plenio. A complete criterion for separability detection. *Phys. Rev. Lett.*, 103(160404), 2009.

[143] M. Navascués, M. Owari, and M.B. Plenio. The power of symmetric extensions for entanglement detection. *Phys. Rev. A*, 80(052306), 2009.

[144] A. Nica and R. Speicher. *Lectures on the Combinatorics of Free Probability*. London Mathematical Society Lecture Note Series, Vol. 335. Cambridge University Press, Cambridge, 2006.

[145] H. Osaka. A class of extremal positive maps in 3×3 matrix algebras. *Publ. Res. Inst. Math. Sci.*, 28(747), 1992.

[146] A. Peres. Separability criterion for density matrices. *Phys. Rev. Lett.*, 77:1413–1415, 1996.

[147] M. Piani. Relative entropy of entanglement and restricted measurements. *Phys. Rev. Lett.*, 103(160504), 2009.

[148] G. Pisier. *The Volume of Convex Bodies and Banach Spaces Geometry*. Cambridge Tracts in Mathematics, Vol. 94. Cambridge University Press, Cambridge, 1989.

[149] S. Popescu and D. Rohrlich. Nonlocality as an axiom. *Foundations of Physics*, 24(3):379–385, 1994.

[150] A. Rao. Parallel repetition in projection games and a concentration bound. *SIAM J. Comput.*, 40(6):1871–1891, 2011.

[151] R. Raz. A parallel repetition theorem. *SIAM J. Comput.*, 27(3):763–803, 1998.

[152] R. Renner, V. Scholz, D. Sutter, and A. Winter. Approximate degradable quantum channels. In *Proc. IEEE Int. Symposium on Inform. Theory*, pages 2767–2771, Hong Kong, 2015.

[153] C.A. Rogers and G.C. Shephard. Convex bodies associated with a given convex body. *J. Lond. Math. Soc.*, 33:270–281, 1958.

[154] R. Rosen. A k-provers parallel repetition theorem for a version of no-signaling model. *Discrete Math. Appl.*, 2(4):457–468, 2010.

[155] H. Rosenthal and S.J. Szarek. *On tensor products of operators from L^p to L^q*, pages 108–132. Lecture Notes in Mathematics, Functional Analysis. Springer-Verlag, Berlin, 1991.

[156] W. Rudin. Trigonometric series with gaps. *J. Math. Mech*, 9(2):203–227, 1960.

[157] W. Rudin. *Functional analysis*. International Series in pure and applied Mathematics. McGraw-Hill, Singapore, 1973.

[158] W. Rudin. *Real and complex analysis*. International Editions, Mathematics Series. McGraw-Hill, Singapore, 1987.

[159] T. Rudolph. Further results on the cross norm criterion for separability. *Quant. Inf. Proc.*, 4(219), 2005.

[160] L. Santaló. An affine invariant for convex bodies of n-dimensional space. *Port. Math.*, 8:155–161, 1949. (in Spanish).

[161] G. Schechtman. More on embedding subspaces of L_p in ℓ_r^n. *Compos. Math.*, 61(2):159–169, 1987.

[162] G. Schechtman. *A remark concerning the dependence on ε in Dvoretzky's theorem*, volume 1376 of *Lecture Notes in Mathematics, Geometric aspects of functional analysis*, pages 274–277. 1987/88.

[163] R. Schneider and W. Weil. *Zonoids and related topics*, pages 296–317. Convexity and its Applications. Birkhäuser, Basel, 1983.

[164] P. Sen. Random measurement bases, quantum state distinction and applications to the hidden subgroup problem. In *Proc. 21st IEEE Conference on Computational Complexity*, Piscataway, NJ, 1960.

[165] B. Simon. *Representations of Finite and Compact Groups*. Graduate Studies in Mathematics, Vol. 10. American Mathematical Society, Providence, 1996.

[166] H-J. Sommers and K. Życzkowski. Induced measures in the space of mixed quantum states. *J. Phys. A.: Gen. Phys.*, 34(35):7111–7124, 2001.

[167] H-J. Sommers and K. Życzkowski. Hilbert-Schmidt volume of the set of mixed quantum states. *J. Phys. A.: Gen. Phys.*, 36(39):10115–10130, 2003.

[168] W.F. Stinespring. Positive functions on C*-algebras. *Proc. Amer. Math. Soc.*, 6:211–216, 1955.

[169] J. Storm. Using SeDuMi 1.02, a Matlab toolbox for optimization over symmetric cones. *Optimization Methods and Software*, 11(1-4):625–653, 1999.

[170] M. Talagrand. Embedding subspaces of L_1 into ℓ_1^n. *Proc. Amer. Math. Soc.*, 108(2):363–369, 1990.

[171] R. Tarrach and G. Vidal. Robustness of entanglement. *Phys. Rev. A.*, 59(141), 1999.

[172] B.S. Tsirelson. Quantum generalizations of Bell's inequality. *Lett. Math. Phys.*, 4(2):93–100, 1980.

[173] R. Vershynin. *Geometric functional analysis*. Lecture notes available at `http://www-personal.umich.edu/~romanv/papers/papers.html`.

[174] J. Watrous. *Theory of quantum information*. Lecture notes available at `https://cs.uwaterloo.ca/~watrous/LectureNotes.html`.

[175] R.F. Werner. Quantum states with Einstein-Rosen-Podolsky correlations admitting a hidden-variable model. *Phys. Rev. A*, 40:4277–4281, 1989.

[176] M.M. Wilde. *Quantum Information Theory*. Cambridge University Press, Cambridge, 2013.

[177] M.M. Wolf. *Quantum channels and operations guided tour*. Lecture notes available at `http://www-m5.ma.tum.de/foswiki/pub/M5/Allgemeines/MichaelWolf/QChannelLecture.pdf`.

[178] A. Zvonkin. Matrix integrals and map enumeration: an accessible introduction. *Math. Comput. Model.*, 26(8-10):281–304, 1997.

Grande dimension et symétries en théorie quantique de l'information

Résumé: S'il fallait résumer le sujet de cette thèse en une expression, cela pourrait être quelque chose comme: phénomènes de grande dimension (mais néanmoins finie) en théorie quantique de l'information. Cela étant dit, essayons toutefois de développer brièvement. La physique quantique a inéluctablement affaire à des objets de grande dimension. Partant de cette observation, il y a, en gros, deux stratégies qui peuvent être adoptées: ou bien essayer de ramener leur étude à celle de situations de plus petite dimension, ou bien essayer de comprendre quels sont les comportements universels précisément susceptibles d'émerger dans ce régime. Nous ne donnons ici notre préférence à aucune de ces deux attitudes, mais au contraire oscillons constamment entre l'une et l'autre.

Notre but dans la première partie de ce manuscrit est de réduire autant que possible la complexité de certains processus quantiques, tout en préservant, évidemment, leurs caractéristiques essentielles. Les deux types de processus auxquels nous nous intéressons sont les canaux quantiques et les mesures quantiques. Dans les deux cas, la complexité d'une transformation est mesurée par le nombre d'opérateurs nécessaires pour décrire son action, tandis que la proximité entre la transformation d'origine et son approximation est définie par le fait que, quel que soit l'état d'entrée, les deux états de sortie doivent être proches l'un de l'autre. Nous proposons des solutions universelles (basées sur des constructions aléatoires) à ces problèmes de compression de canaux quantiques et d'amenuisement de mesures quantiques, et nous prouvons leur optimalité.

La deuxième partie de ce manuscrit est, au contraire, spécifiquement dédiée à l'analyse de systèmes quantiques de grande dimension et certains de leurs traits typiques. L'accent est mis sur les systèmes multi-partites et leurs propriétés ayant un lien avec l'intrication. Les principaux résultats auxquels nous aboutissons peuvent se résumer de la façon suivante: lorsque les dimensions des espaces sous-jacents augmentent, il est générique pour les états quantiques multi-partites d'être à peine distinguables par des observateurs locaux, et il est générique pour les relaxations de la notion de séparabilité d'en être des approximations très grossières. Sur le plan technique, ces assertions sont établies grâce à des estimations moyennes de suprema de processus gaussiens, combinées avec le phénomène de concentration de la mesure.

Dans la troisième partie de ce manuscrit, nous revenons pour finir à notre état d'esprit de réduction de dimensionnalité. Cette fois pourtant, la stratégie est plutôt: pour chaque situation donnée, tenter d'utiliser au maximum les symétries qui lui sont inhérentes afin d'obtenir une simplification qui lui soit propre. En reliant de manière quantitative symétrie par permutation et indépendance, nous nous retrouvons en mesure de montrer le comportement multiplicatif de plusieurs quantités apparaissant en théorie quantique de l'information (fonctions de support d'ensembles d'états, probabilités de succès dans des jeux multi-joueurs non locaux etc.). L'outil principal que nous développons dans cette optique est un résultat de type de Finetti particulièrement malléable.

Mots clés: Théorie quantique de l'information, Analyse géométrique asymptotique, Symétrie par permutation.

Image en couverture: Alice et Bob vivant sur une planète dont la masse se concentre autour de l'équateur, dessin au crayon de couleur et encre par Aurélie Garnier.



UNIVERSITÉ DE LYON