Random States in Quantum Information: from Resource Theories to Quantum Metrology

Dottorando: Giovanni GRAMEGNA

Supervisore: Ch.mo Prof. Paolo FACCHI

Coordinatore: Ch.mo Prof. Giuseppe Iaselli

ESAME FINALE 2021
“So my antagonist said, «Is it impossible that there are flying saucers? Can you prove that it’s impossible?» «No, I can’t prove it’s impossible. It’s just very unlikely». At that he said, «You are very unscientific. If you can’t prove it impossible then how can you say that it’s unlikely?» But that is the way that is scientific. It is scientific only to say what is more likely and what less likely, and not to be proving all the time the possible and impossible.”

Richard Feynman
Acknowledgements

First and foremost, I would like to thank Paolo Facchi for his invaluable guidance, infinite patience and constant dedication, my gratitude to him is enormous. I feel lucky to have worked with such a wonderful person.

I would like to express my gratitude also to all the other people with whom I had the honor to collaborate during these years. I am thankful to Fabio Deelan Cunden for sharing his enlightening intuitions on the topics we investigated together, and Giuseppe Florio for his precious advices. I am grateful to Saverio Pascazio, Daniel Burgarth and Kazuya Yuasa for sharing their knowledge and experience with me, and for conveying a great passion and dedication to their work. I would like also to thank Vincenzo Tamma and Frank Narducci for their positive and encouraging attitude. I am also thankful to all my PhD colleagues, for contributing to create an interactive research environment.

I am immensely grateful for having had the opportunity to share part of this difficult journey with my friend and colleague Danilo Triggiani.

Some special words of gratitude go to my friends Alessandro, Davide and Danilo (again) for lightening the mood when I was a bit discouraged and for providing enjoyable distractions.

I am thankful to my parents, for always being of support for me no matter what, and for always trying to help me as best as they could.

Finally, thanks Rosa, for being there for me and for giving me the extra strength to move forward even when things get difficult.
# Contents

Acknowledgements iii  
Preface ix  

I Random Techniques and Concentration of Measure 1  

1 Concentration of Measure Phenomenon and Typicality 3  
  1.1 Deviation inequalities in probability theory 3  
  1.1.1 Brief recap on Probability Theory 3  
  1.1.2 Law of large numbers and large deviations 5  
  1.1.3 Standard Deviation Inequalities 5  
  1.2 Concentration of measure on the sphere and the isoperimetric problem 9  
  1.2.1 Volume of Euclidean Balls 9  
  1.2.2 Volume of hyperspherical caps 10  
  1.2.3 Isoperimetric Problem and concentration on the sphere 11  
  1.3 Concentration Function 12  
  1.4 General deviation inequalities 13  
  1.5 Typicality 15  

2 Random Quantum States and Haar measure 17  
  2.1 Quantum States 17  
  2.2 Haar measure on the Unitary Group 19  
  2.2.1 Distribution of the eigenvalues and level spacings 20  
  2.2.2 Distribution of the entries 21  
  2.3 Sampling Haar distributed unitary matrices 22  
  2.4 Concentration of Measure on the Unitary Group 26  
  2.5 Typicality of Observables on Random Subspaces 28  
  2.5.1 Averages of Expectation and Variance 29  
  2.5.2 Lipschitz constants 30  

3 Random states on Bipartite Quantum Systems 33  
  3.1 Bipartite quantum systems 33  
  3.1.1 Entanglement and separability 33  
  3.1.2 Measures of Entanglement 34  
  3.2 The unbiased ensemble 35  
  3.2.1 Typicality of Entanglement Entropy and Purity 35  
  3.2.2 Distribution of Schmidts coefficients for the unbiased ensemble 38
3.3 Random pure states and Wishart Matrices
3.3.1 Tridiagonal realization of Wishart Random Matrices
3.4 Spectrum of Wishart Random Matrices
3.4.1 Marchenko-Pastur Distribution
3.4.2 Largest eigenvalue and Tracy-Widom distribution
3.4.3 Smallest eigenvalue of fixed-trace Wishart matrices

II Random states in Quantum Resource Theories

4 Quantum Resource Theories
4.1 Quantum Operations
4.1.1 Stinespring Factorization
4.1.2 Operator-Sum representation
4.2 General structure of Quantum Resource Theories
4.2.1 Consistent QRTs for a given set of free operations
4.2.2 Consistent QRTs for a given set of free states
4.2.3 Resource Quantification
4.3 State conversions in Resource Theories
4.3.1 Single-shot convertibility
4.3.2 Asymptotic convertibility
4.3.3 Catalytic convertibility
4.4 Majorization Theory
4.4.1 Schur-convexity and majorization monotones

5 Random State Conversions in the Resource Theory of Entanglement
5.1 Introduction and some notation
5.2 Deterministic conversions of pure bipartite states
5.2.1 Majorization as the persistence of a random walk above the origin
5.3 Numerical results on deterministic conversion
5.3.1 Confirmation of Nielsen Conjecture
5.3.2 Discussion on the persistence exponents
5.4 Stochastic state conversions
5.4.1 Distribution of the maximal success probability
5.4.2 Exact formulae for $n = 2$
5.4.3 Distribution function of the maximal success probability and scaling limits
5.5 Conclusions and Outlook

6 Resource Theory of Coherence
6.1 Resource theories of (speakable) coherence
6.1.1 Free States
6.1.2 Allowed Operations
6.1.3 Convertibility criterion and majorization relation
6.2 Random pure states
6.3 Volume of the set of IO-convertible states
6.3.1 Asymptotics $n \to \infty$
III Random states for Quantum Metrology

7 Precision Scaling in Quantum Metrology
7.1 Quantum Estimation Theory
7.1.1 Estimators
7.1.2 Cramér-Rao Bound and Fisher Information
7.1.3 Quantum Cramér-Rao Bound
7.1.4 Quantum Fisher Information for parameter independent Hamiltonian evolution
7.2 Standard Quantum Limit and Heisenberg Limit
7.2.1 Separable case
7.2.2 Entangled case
7.3 Typical Sensitivity of Random Quantum States
7.4 Typical Sensitivity of Random Symmetric States
7.4.1 A particular case
7.4.2 Usefulness of random symmetric states for quantum metrology

8 Distributed Quantum Metrology with Gaussian States
8.1 Introduction
8.2 Gaussian States and Operations
8.2.1 Gaussian states
8.2.2 Passive linear operations
8.3 Setup
8.3.1 Metrological scheme
8.4 Heisenberg scaling
8.4.1 Fisher Information of the Setup
8.4.2 Conditions for the Heisenberg Scaling
8.4.3 One-sided adaptivity
8.5 A two-mode example
8.6 Typical sensitivity
8.6.1 Average Behaviour
8.6.2 Typicality of the prefactor
8.6.3 Analytic distribution of the pre-factor for generators with only two distinct eigenvalues
8.7 Conclusions
Conclusions and Outlook 137

8.8 Outlooks ................................................. 138

A Log-Sobolev inequalities and Concentration of measure 141
A.1 Laplace bounds on the concentration function .......... 141
A.2 Concentration from Logarithmic Sobolev inequalities .......... 142

B Miscellaneous results on joint statistics 145
B.1 Eigenvalues of Wishart and fixed-trace Wishart matrices .... 145
B.2 Order statistics of i.i.d. random variables and Markov property 146
B.3 Proof of Proposition 10 ........................................ 149
B.4 Proof of Claim (6.22) in Theorem 22 ..................... 150

C Mathematical characterization of LOCC operations 153
C.1 Finite-round LOCC in bipartite systems ................... 153
C.2 Simplification in the pure bipartite case .................. 155

D Homodyne detection 157
D.1 Probability distribution of the homodyne detection ....... 157
D.1.1 Derivation of (D.3) ..................................... 160

Bibliography 161
Preface

Probabilistic methods have been at the heart of information theory from the very beginning. One of the greatest results of classical information theory, Shannon’s source coding theorem [141], relies on an exponential version of the central limit theorem to prove that the relative volume of a “typical set” approaches unity in a certain limit: hence, for all practical purposes, one can consider only the elements of this set when performing some particular task (in the case of Shannon’s theorem, data compression). The key ingredient of this perspective, \textit{typicality}, can be expressed in its broadest meaning as the property of certain systems in which the overwhelming majority of configurations lead to the same “emergent” behaviour. This idea goes back to the very foundations of classical statistical mechanics: the observable properties of a macroscopic system are determined by what happens on regions of the phase space whose relative volume is close to one. Although a powerful tool, the central limit theorem is suitable only for those situations where one is considering sums (or more generally, linear combinations) of a large number of \textit{independent} random variables. It turns out that these requirements are, in fact, unnecessary, since the feature of the central limit theorem which is relevant for typicality, namely the concentration of measure, can be observed even when the variables involved are not independent, and even if one is not considering just linear combinations of said variables [150, 61, 94]. These more general results of concentration of measure have proven to be useful to show important results in quantum information theory [77, 73, 23, 123] as well as deriving fundamental results of statistical mechanics starting from quantum theory [129, 65, 66].

This thesis collects my research work concerning typical properties of random states in quantum information. This field has grown a lot during the years, becoming immense and covering many topics. Here, we will focus on two of them, namely the resource theoretic approach, which has attracted a lot of attention during the last twenty years [1, 81, 20, 67, 162, 98, 30] and has proven to be useful in capturing the essential properties of quantum resources, and quantum metrology [63, 64, 39, 38, 123, 106], which aims at determining what are the fundamental limits on the precision achievable in a quantum measurement, and how it is possible to achieve them.

The dissertation is tripartite. Part I contains an introduction to the notion of concentration of measure and the use of techniques from Random Matrix Theory for the characterization of typical properties of random states. Chapter 1 contains a general introduction to the topic of concentration of measure and how it determines some typical properties of large systems. Chapter 2 collects some results about the Haar measure on the unitary group, which
are essential for the discussion of the results presented in this thesis. Chapter 3 focuses on the description of pure random states of bipartite systems, and provides techniques from Random Matrix Theory which will be used, in particular, for the discussion of results presented in Chapter 5.

The original material of the thesis is contained in Parts II and III. Part II focuses on the resource theoretic approach to information theory. Chapter 4 contains a general introduction to the subject, and to the mathematical relation of majorization, which plays a fundamental role in characterizing state conversions of the resource theories of Entanglement and Quantum Coherence, analysed respectively in Chapter 5 and Chapter 6. In particular, Chapter 5 contains an extensive investigation of state conversions among pure bipartite states using local operations and classical communication (LOCC). Several numerical results are presented, and an interesting connection between the problem of state conversions and the extreme statistics of an ensemble of random matrices is discussed. In Chapter 6 the analogous problem of conversions among pure states using only incoherent operations is analysed. A conjecture of “generic incomparability” is proved, and the connection between the problem under investigation and the extreme statistics (statistical behaviour of the smallest components of certain random vectors) is confirmed.

Part III is devoted to the subject of Quantum Metrology. In Chapter 7 a brief introduction on the subject is presented, with a discussion of the fundamental limits of precision achievable in the estimation of a parameter performing measurements on random quantum states. Then, in Chapter 8, we focus on analysing a particular set of states of infinite dimensional systems, namely Gaussian states. A metrological scheme for the estimation of a generic parameter embedded into a multimode interferometer is analysed, showing that the scaling of the precision in the number of resources employed (number of photons) is the best achievable (at the Heisenberg limit), and performances for random states (inside the particular subset considered) are analysed.
Part I
Random Techniques and Concentration of Measure
Chapter 1

Concentration of Measure Phenomenon and Typicality

This chapter provides an introduction on the notion of concentration of measure, starting from simple examples and going to more general results. In Section 1.1 we will see how the concentration of measure arises in probability theory as a property of a large number of independent variables, by presenting the traditional large deviation inequalities. Then, in Section 1.2 we will go beyond the scenario of independent variables and consider the concentration of measure from a geometric perspective, focusing on the particular example of the $n$-dimensional sphere. This example will show an interesting and enlightening connection between concentration of measure and isoperimetric inequalities. Using the intuition gathered from this connection, we will introduce the concentration function for general metric probability spaces in Section 1.3, and we will use it to characterize general deviation inequalities in Section 1.4. Finally, in Section 1.5 we will conclude the chapter discussing how the concentration of measure phenomenon is relevant in the characterization of typical properties of large systems.

1.1 Deviation inequalities in probability theory

1.1.1 Brief recap on Probability Theory

In the standard measure-theoretic approach to probability theory (see, for example, [151]), a probability space is defined as a triple $(\Omega, B, \mathcal{P})$: we have some sample space $\Omega$, which is supposed to contain all the possible outcomes $\omega$ of an “experiment”, i.e. all the sources of randomness one is studying. An event is a subset of $E \subset \Omega$, and the set of all events will constitute a $\sigma$-algebra $B$ of subsets of $\Omega$. A probability measure $\mathcal{P} : B \to [0, 1]$, with $\mathcal{P}(\Omega) = 1$, assigns to each event $E \in B$ its probability $\mathcal{P}(E)$ of occurring.

Example 1. One of the paradigmatic examples is the rolling of a six-faced die, which can be characterized by the sample space $\Omega = \{1, 2, \ldots, 6\}$, the $\sigma$-algebra of all the subsets of $\Omega$ and the uniform probability measure $\mathcal{P}(\omega) = 1/6$ for each $\omega \in \Omega$ (for a fair die).

Despite the formal importance of the ambient probability space, when one is not working on the foundational level it is more appealing to work with actual random variables and their distributions.
Chapter 1. Concentration of Measure Phenomenon and Typicality

Definition 1 (Random Variable). Let \((R, \mathcal{R})\) a measurable space (a set \(R\) equipped with a \(\sigma\)-algebra of subsets \(\mathcal{R}\)). A random variable \(X\) is a measurable map from the sample space to \(R\), i.e. a function \(X : \Omega \rightarrow R\) such that \(X^{-1}(S)\) is an event for every \(S \in \mathcal{R}\).

Given a random variable \(X\) taking values in some set \(R\), we define its distribution \(\mu_X\) to be the probability measure on the measurable space \((R, \mathcal{R})\) defined by the formula

\[\mu_X(S) := \mathcal{P}(X \in S).\] (1.1)

Using equation (1.1) it is possible to assign a probability distribution to every random variable \(X\). The converse is also true:

Lemma 1. Let \(\mu\) be a probability measure on a measurable space \((R, \mathcal{R})\). Then (after extending the sample space \(\Omega\) if necessary) there exists an \(R\)-valued random variable with distribution \(\mu\).

Note that the triple \((R, \mathcal{R}, \mu)\) constitutes itself a probability space with ambient sample space \(R\). From a practical point of view, we will always be interested in the properties of some family of random variables, hence we will always be interested on the concentration properties of the associated probability spaces.
1.1.2 Law of large numbers and large deviations

The concentration of measure is a property of a large number of variables, such as in the law of large numbers, which is arguably one of the most significant results in probability theory. Its essence can be expressed by a sentence which is easily understood and agreed upon:

“If a fair coin is tossed many times, it is likely that head will come up nearly half of the times.”

This piece of common knowledge can be expressed more quantitatively by specifying what do we mean exactly with “it is likely” and “nearly half of times”. If we register the result of the \(j\)-th coin toss as \(X_j = 1\) when head comes up and \(X_j = 0\) when tail comes up, the number of heads after \(n\) coin tosses is given by \(S_n = \sum_{j=1}^{n} X_j\). Then, a more quantitative version of the statement above would be:

\[
P\left(\left|\frac{S_n}{n} - \frac{1}{2}\right| \geq \varepsilon\right) \leq \gamma_n(\varepsilon),
\]

where \(\varepsilon\) serves to quantify the “nearly” in the above statement, while \(\gamma_n(\varepsilon)\) quantifies the “likely”, and in particular in order to give quantitative credit to the sentence above we desire the function \(\gamma_n(\varepsilon)\) to decrease sufficiently fast as \(\varepsilon\) grows: in this way, the probability of having large deviations \(\varepsilon\) from the average is bounded by a small quantity \(\gamma_n(\varepsilon)\).

By the central limit theorem we know that fluctuations of \(S_n\) are of order \(\sqrt{n}\), which definitely is not small when many tosses are considered. But a bit of care is needed, since we are considering the fluctuations of a quantity whose average is proportional to \(n\), so this is the right scale to consider. Hence, a more meaningful version of (1.2) is the following:

\[
P\left(\left|\frac{S_n}{n} - \frac{1}{2}\right| \geq \varepsilon\right) \leq \delta_n(\varepsilon).
\]

(Note that if (1.2) holds, then (1.3) holds with \(\delta_n(\varepsilon) = \gamma_n(n\varepsilon)\), reflecting the fact that now we are considering the right scale). An elementary numerical experiment (see Figure 1.1 and Figure 1.2) can be done to show this simple concentration phenomenon in action. Equations in the form of (1.2) or (1.3), where we try to bound the tails of some distribution, are called deviation inequalities. In the next section we will briefly see some standard deviation inequalities of elementary probability theory which can be applied when one has to do with many independent random variables.

1.1.3 Standard Deviation Inequalities

Proposition 1 (Markov’s Inequality). Let \(X\) be a non-negative random variable \(X\). Then, for any \(t > 0\) we can bound its tail as:

\[
P(X > t) \leq \frac{\mathbb{E}[X]}{t}.
\]
Figure 1.2: Distribution of deviations from the average, estimated with a simple numerical experiment on $10^4$ samples for $n = 10, 100, 1000$.

Proof. It follows immediately by:

\[
\mathbb{E}[X] = \int_{0}^{\infty} x \, d\mu_X(x) \geq \int_{t}^{\infty} x \, d\mu_X(x) \geq t \int_{t}^{\infty} d\mu(x) = t \mathbb{P}(X \geq t). \tag{1.5}
\]

\[\square\]

Corollary 1 (Chebyshev’s Inequality). Let $X$ be a random variable with $\mathbb{E}[X] = \mu$ and $\text{Var}[X] = \sigma^2$. Then deviations from the average are bounded as:

\[
\mathbb{P}(|X - \mu| > \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2}. \tag{1.6}
\]

Proof. This is an immediate consequence of Markov’s Inequality since:

\[
\mathbb{P}(|X - \mu| > \varepsilon) = \mathbb{P}(|X - \mu|^2 \geq \varepsilon^2) \leq \frac{\mathbb{E}[(X - \mu)^2]}{\varepsilon^2} = \frac{\sigma^2}{\varepsilon^2}. \tag{1.7}
\]

\[\square\]

Chebyshev inequality is surely useful when one is trying to bound the tail of a random variable with its variance, which is usually easy to compute. Let us apply it to the normalized sum of $n$ i.i.d. random variables $X_1, \ldots, X_n$ with common expectation value $\mathbb{E}[X_j] = \mu$ and variance $\text{Var}[X_j] = \sigma^2$. The normalized sum, defined as:

\[
\overline{X}_n := \frac{1}{n} \sum_{j=1}^{n} X_j \tag{1.8}
\]
1.1. Deviation inequalities in probability theory

has $E[X_n] = \mu$ and variance $\text{Var}(X_n) = \frac{\sigma^2}{n}$, so that the Chebyshev’s inequality yields:

$$\mathcal{P}(|X_n - \mu| > \varepsilon) \leq \frac{\sigma^2}{n\varepsilon^2} \quad (1.9)$$

Applying this result to the example of coin tossing, we see that the inequality (1.3) holds with $\delta_n(\varepsilon) = \frac{\sigma^2}{n\varepsilon^2}$, which decreases both in $n$ and $\varepsilon$ as we desire: for many coin tosses $n$, large deviations $\varepsilon$ are unlikely. However, we can do much better than this: to be more precise, exponentially better.

**Proposition 2 (Chernoff bound).** Let $X$ be a real random variable. Then, for any $\lambda > 0$:

$$\mathcal{P}(X > \varepsilon) \leq e^{-\lambda\varepsilon} E[e^{\lambda X}] \quad (1.10)$$

**Proof.** For any $\lambda \geq 0$ we have that:

$$\mathcal{P}(X > \varepsilon) = \mathcal{P}(\lambda X > \lambda \varepsilon) = \mathcal{P}(e^{\lambda X} > e^{\lambda \varepsilon}) \leq e^{-\lambda \varepsilon} E[e^{\lambda X}], \quad (1.11)$$

where the second equality holds by monotonicity, and Markov’s inequality was applied. \hfill \Box

Chernoff bound can be used to improve Chebyshev’s inequality if one is able to find a suitable bound on $E[e^{\lambda X}]$. The following fundamental result provides such a bound in the case of bounded random variables $X$. It can be proved by using Taylor’s theorem and Jensen’s inequality. See for example [105].

**Lemma 2 (Hoeffding’s lemma).** Let $X$ be a real-valued random variable with $E[X] = \mu$ and $a \leq X \leq b$ almost surely. Then, for all $t > 0$:

$$E[e^{tX}] \leq e^{t\mu + t^2(b-a)^2/8} \quad (1.12)$$

Using this lemma, it is possible to exponentially improve the bound (1.9) on the normalized sum of $n$ independent random variables, further generalizing it to the case in which the random variables $X_1, \ldots, X_n$ are not identically distributed.

**Theorem 1 (Hoeffding’s inequality).** Let $X_1, X_2, \ldots, X_n$ be independent random variables with $\mathcal{P}(a_i \leq X_i \leq b_i) = 1$ and $E[X_i] = \mu_i$. Then, for any $\varepsilon > 0$:

$$\mathcal{P}(|\overline{X}_n - \mu| > \varepsilon) \leq 2e^{-\frac{2n\varepsilon^2}{L^2}} \quad (1.13)$$

where

$$\mu = \frac{1}{n} \sum_{i=1}^{n} \mu_i \quad L^2 = \frac{1}{n} \sum_{i=1}^{n} (a_i - b_i)^2 \quad (1.14)$$

**Proof.** The idea is to apply (1.10) to the random variables $Y_n = \overline{X}_n - \mu$:

$$\mathcal{P}(Y_n > \varepsilon) \leq e^{-\lambda\varepsilon} E[e^{\lambda Y_n}], \quad (1.15)$$
which holds for any $\lambda > 0$. Using the independence of $X_1, \ldots, X_n$ and applying Lemma 2 to each $X_i$, we get:

$$
\mathbb{E}[e^{\lambda Y_n}] = e^{-\lambda \mu} \mathbb{E}[e^{\lambda X_n}] = e^{-\lambda \mu} \mathbb{E}[e^{(\lambda/n) \sum_{i=1}^{n} X_i}] = e^{-\lambda \mu} \prod_{i=1}^{n} \mathbb{E}[e^{(\lambda/n) X_i}]
$$

$$
\leq e^{-\lambda \mu} \prod_{i=1}^{n} e^{(\lambda/n) \mu_i + (\lambda/n)^2 (b_i - a_i)^2 / 8}
$$

$$
= e^{\lambda^2 L^2 / (8n)}.
$$

(1.16)

Substituting in (1.15), we get:

$$
\mathcal{P}(Y_n > \varepsilon) \leq e^{-\lambda \varepsilon} e^{\lambda^2 L^2 / (8n)},
$$

(1.17)

which holds for any $\lambda > 0$. Choosing the optimal $\lambda = 4 \varepsilon n / L^2$, we obtain

$$
\mathcal{P}(Y_n > \varepsilon) \leq e^{-2n \varepsilon^2 / L^2}.
$$

(1.18)

Using the same argument with $-Y_n$, we get the analogous bound on $\mathcal{P}(Y_n < -\varepsilon)$, which together with (1.18) yields (1.13)

Applying this theorem to our initial problem of coin tossing, where $a_i = 0$ and $b_i = 1$, we get that equation (1.3) holds with the exponential deviation bound:

$$
\mathcal{P} \left( \left| \frac{S_n}{n} - \frac{1}{2} \right| \geq \varepsilon \right) \leq 2e^{-2n \varepsilon^2}.
$$

(1.19)

Before proceeding to the next section and going beyond independent random variables, let us consider a slightly different interpretation of measure concentration results that we obtained here for the sum $S_n = X_1 + X_2 + \ldots, X_n$ of many independent random variables. Taking for example the case $\mathcal{P}(X_i = 0) = \mathcal{P}(X_i = 1) = 1/2$ just considered, we can think of $S_n/n$ as a function of the variables $X_i$ which, as a consequence of the concentration result (1.19), is essentially constant (and equal to 1/2). According to this interpretation, due to Talagrand [150], an aspect of measure concentration is that a random variable that depends (in a smooth way) on the influence of many independent variables (but not too much on any of them) is essentially constant.

As we will see in the following, concentration of measure is an interesting feature of some high-dimensional spaces which are more general than the previous cases in two respects: linear functionals (such as sums) can be replaced by arbitrary Lipschitz functions of the samples, and the probability measure does not need to be in a product form (i.e. arising from independent random variables).
1.2 Concentration of measure on the sphere and the isoperimetric problem

Let us look here at the concentration phenomenon with a more geometrical perspective. A nice introduction on the topic can be found in [10]. Due to the geometric structure of quantum states, it is also mandatory to discuss extensively the subject of measure concentration on Euclidean balls.

1.2.1 Volume of Euclidean Balls

The Euclidean unit ball in $\mathbb{R}^n$ will be denoted by $B^n = \{ x \in \mathbb{R}^n : \|x\|_2 \leq 1 \}$, while its boundary, the unit sphere, by $S^{n-1} = \partial B^n = \{ x \in \mathbb{R}^n : \|x\|_2 = 1 \}$. For a (measurable) subset $A \subset \mathbb{R}^n$, we will denote with $|A|$ its Lebesgue measure. The measure of $B^n$ can be computed explicitly with a standard calculation yielding

$$|B^n| = \frac{\pi^{n/2}}{\Gamma(\frac{n}{2} + 1)},$$

(1.20)

where $\Gamma$ denotes the Euler function. The measure of $S^{n-1}$ is related to $|B^n|$ by $|S^{n-1}| = n|B^n|$. For large $n$, we use the Stirling approximation

$$\Gamma\left(\frac{n}{2} + 1\right) \sim \sqrt{\pi n}\left(\frac{n}{2e}\right)^{\frac{n}{2}}$$

(1.21)

to see that

$$|B^n| \sim \frac{1}{\sqrt{\pi n}} \left(\sqrt{\frac{2\pi e}{n}}\right)^n \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$ (1.22)

This shows that the ball with unit radius has vanishing volume as $n \rightarrow \infty$. Equivalently, the ball with unit volume needs to have a large radius as $n \rightarrow \infty$. More precisely, the volume of the ball $B^n(r) = \{ x \in \mathbb{R}^n : \|x\|_2 \leq r \}$ with radius $r > 0$ is given by $|B^n(r)| = |B^n| r^n$, so that by (1.22), the radius of the ball with unit volume is given by

$$r^* = \left(\frac{1}{|B^n|}\right)^{1/n} \sim \sqrt{\frac{n}{2\pi e}}.$$ (1.23)

Let us consider now the volume of slices of the sphere perpendicular to a radius (and parallel between themselves) at different distances $x$ from the center (see Figure 1.3). First, note that a slice of the ball $B^n(r)$ passing through the center is a ball $B^{n-1}(r^*)$ with the same radius. Hence, considering the ball $B^{n-1}(r^*)$ with unit volume, the slice passing through the center has a volume

$$|B^{n-1}(r^*)| = |B^{n-1}|r^{n-1} = |B^{n-1}| \left(\frac{1}{|B^n|}\right)^\frac{n-1}{n} \sim \sqrt{e}.$$ (1.24)

A slice of the ball perpendicular to a radius which is at a distance $x$ from the center will be a ball in $\mathbb{R}^{n-1}$ with radius $\sqrt{r^2} - x^2$ (see Figure 1.3), hence its
Chapter 1. Concentration of Measure Phenomenon and Typicality

volume is given by:

$$|B_n^{-1}(\sqrt{r^2 - x^2})| = |B_n^{-1}(r^2 - x^2)^{\frac{n-1}{2}} = |B_n^{-1}(r_*)| \left(\frac{\sqrt{r^2 - x^2}}{r_*}\right)^{n-1}, \quad (1.25)$$

which, using the approximations (1.23) and (1.24), is given by

$$|B_n^{-1}(\sqrt{r^2 - x^2})| \sim \sqrt{e} \left(1 - \frac{2\pi e x^2}{n}\right)^{\frac{n-1}{2}} \sim \sqrt{e} e^{-\pi e x^2}, \quad (1.26)$$

that is a Gaussian distribution with variance $\sigma^2 = \frac{1}{2\pi e}$. This is a surprising result: despite the fact that the radius of a ball with unit volume in $\mathbb{R}^n$ grows as $\sqrt{n}$ as $n$ goes infinity, the distribution of the mass of parallel slices through the ball has a finite (independent on $n$) variance. In other words, the slices contributing to most of the volume of the normalized ball are those lying near the equator.

1.2.2 Volume of hyperspherical caps

Let us turn now the volume of the sphere $S^{n-1}$. In order to illustrate more formally the concentration of measure of the sphere around the equator, we will first consider the measure of hyperspherical caps (see Figure 1.4). We will denote the spherical cap of height $1 - s$ around the direction $u \in S^{n-1}$ with

$$C(u, s) = \{x \in S^{n-1} : x \cdot u \geq s\}, \quad (1.27)$$

where $\cdot$ denotes the Euclidean product. We will consider the uniform probability measure $\sigma_n$ on $S^{n-1}$, given by the relative volume:

$$\sigma_n(A) = \frac{|A|}{|S^{n-1}|}, \quad (1.28)$$
for every measurable set $A \subset S^{n-1}$.

**Lemma 3 ([10]).** Let $\sigma_n$ be the uniform probability measure on $S^{n-1}$ and let $C(u, \varepsilon)$ defined as in (1.27). Then:

$$\sigma_n(C(u, \varepsilon)) \leq e^{-\frac{n}{2}\varepsilon^2}$$

(1.29)

The bound of (1.29) allow us to state more precisely the concentration of measure (in this case, on the sphere) around the equator (actually, *any* equator!). In fact, denoting with $A_\varepsilon = S^{n-1} \setminus (C(u, \varepsilon) \cup C(-u, \varepsilon))$ the belt of “width” $2\varepsilon$ around the equator, we can bound from below its measure using the result (1.29) as:

$$\sigma_n(A_\varepsilon) \geq 1 - 2e^{-\frac{n}{2}\varepsilon^2},$$

(1.30)

expressing more formally the statement that most of the measure on the $(n-1)$-dimensional sphere is concentrated around the equator. In the following we will show a connection of concentration of measure and the isoperimetric problem which will allow to establish that when concentration of measure occurs, the measure of points which lie near to *any* set of measure $\sigma_n(A) > 1/2$ (in this case, the half-emisphere) is close to one.

### 1.2.3 Isoperimetric Problem and concentration on the sphere

The classical isoperimetric theorem states that among all subsets of $\mathbb{R}^n$ of a given volume, Euclidean balls minimize their surface area. An equivalent formulation can be obtained in terms of the $r$-enlargement of a given region $A \subset \mathbb{R}^n$, defined for $r > 0$ by:

$$A_r = \{ x \in \mathbb{R}^n : d(x, A) \leq r \},$$

(1.31)

where $d(x, A) = \inf_{y \in A} d(x, y)$ and we use the Euclidean distance, i.e. $d(x, y) = \|x - y\|_2$. In terms of enlargements, the isoperimetric theorem states that among all (measurable) subsets $A \in \mathbb{R}^n$ having the same volume, the ones
having minimal enlargements are the Euclidean balls. Hence, if \( A \subset \mathbb{R}^n \) has the same volume as the Euclidean unit ball, \( |A| = |B^n| \), we have that \( |A_r| \geq |B^n_r| \) for any \( r > 0 \). The formulation of the isoperimetric problem in terms of enlargements relates the metric and the measure of a space: in the case of \( \mathbb{R}^n \) considered up to now, these are the Euclidean metric and the Lebesgue measure. Considering the isoperimetric problem on the hypersphere \( S^{n-1} \) equipped with the uniform probability measure \( \sigma_n \) and the Euclidean metric inherited from \( \mathbb{R}^n \), the solution of the isoperimetric problem is given by spherical caps. Note also that spherical caps are nothing but the balls on \( S^{n-1} \), since for example using the Euclidean metric \( \|x - u\|^2 = 2(1 - x \cdot u) \), it is immediate to verify that the ball of radius \( r \) centered on \( u \) is \( C(u, 1 - r^2/2) \).

Using together the upper bound (1.29) on the volume of hyperspherical caps and the fact that they are the solutions of the isoperimetric problem on \( S^{n-1} \) we obtain the following lemma.

**Lemma 4.** Consider \( (S^{n-1}, \|\cdot\|_2) \), where \( S^{n-1} \) is the Euclidean sphere in \( \mathbb{R}^n \) and \( \|\cdot\|_2 \) is the Euclidean metric. Let \( \sigma_n \) be the normalized area measure on \( S^{n-1} \). If \( A \subset S^{n-1} \) is Borel measurable and \( \mu(A) \geq 1/2 \), then, for any \( \varepsilon > 0 \), we have

\[
\sigma_n(x \in S^{n-1} : d(x, A) \geq \varepsilon) \leq 1 - e^{-\frac{n\varepsilon^2}{2}}. \tag{1.32}
\]

The concentration of measure on the sphere has the important consequence that any function on the sphere which is sufficiently “regular” concentrates around its median value. We will state the result here, leaving the proof to the next section, where we will give it in a more general scenario which includes the sphere as a particular case.

**Theorem 2.** Let \( f : S^{n-1} \to \mathbb{R} \) be a Lipschitz function with Lipschitz constant \( L \), i.e.

\[
|f(x) - f(y)| \leq L\|x - y\|_2, \tag{1.33}
\]

and let \( m_f \) be a median of \( f \) with respect to \( \sigma_n \), i.e. a number \( m_f \in \mathbb{R} \) such that:

\[
\sigma_n(x \in S^{n-1} : f(x) \leq m_f) \geq \frac{1}{2}, \quad \sigma_n(x \in S^{n-1} : f(x) \leq m_f) \leq \frac{1}{2}. \tag{1.34}
\]

Then

\[
\sigma_n(|f(x) - m_f| \geq \varepsilon) \leq 2e^{-\frac{n\varepsilon^2}{2}}. \tag{1.35}
\]

### 1.3 Concentration Function

The concentration phenomenon can be generalized and formalized through the concentration function of a probability measure on a metric space [94]. The two basic ingredients of the concentration phenomenon are indeed a (probability) measure and a notion of distance with respect to which the concentration is evaluated.
**Definition 2.** Let \((X, d)\) be a metric space equipped with a probability measure \(\mu\) on the Borel sets of \((X, d)\). The concentration function is then defined as
\[
\alpha(r) = \sup \{ 1 - \mu(A_r) : A \subset X, \mu(A) \geq 1/2 \}, \quad r > 0,
\] (1.36)
where
\[
A_r = \{ x \in X ; d(x, A) < r \}
\] (1.37)
is the (open) \(r\)-neighborhood of \(A\) with respect to \(d\).

Informally, the space \((X, d, \mu)\) exhibits a concentration phenomenon if \(\alpha(r)\) decays very fast as \(r\) grows. More formally, a family of metric measure spaces \((X_n, d_n, \mu_n)\) is called a Lévy family if the corresponding concentration functions \(\alpha_n\) satisfy
\[
\alpha_n(r) \to 0 \quad \text{as} \quad n \to \infty.
\] (1.38)
for each fixed \(r > 0\). In particular, a Lévy family is called a normal Lévy family if it is possible to find some constants \(c, C > 0\) such that
\[
\alpha_n(r) \leq Ce^{-cnr^2}.
\] (1.39)

**Proposition 3** ([94], Proposition 1.2). Let \(F\) be a Lipschitz function between two metric spaces \((X, d)\) and \((Y, \delta)\), i.e. such that
\[
\delta(F(x), F(x')) \leq L d(x, x') \quad \text{for all} \quad x, x' \in X
\] (1.40)
for some \(L > 0\). Let \(\mu\) be a probability measure on the Borel sets of \(X\) and denote by \(\mu_F\) the measure \(\mu\) pushed forward by \(F\) on the Borel sets of \((Y, \delta)\). Then, for every \(r > 0\),
\[
\alpha_{(Y, \delta, \mu_F)}(r) \leq \alpha_{(X, d, \mu)} \left( \frac{r}{L} \right).
\] (1.41)
In particular, \(\alpha_{\mu_F} \leq \alpha_{\mu}\) if \(F : X \to Y\) is \(1\)-Lipschitz, i.e. equation (1.40) holds with \(L = 1\).

This simple contraction property shows that concentration of measure are decreasing under \(1\)-Lipschitz mappings. In the following, we will say that a function \(F\) is Lipschitz with Lipschitz constant \(L\) if the inequality (1.40) holds, with the metric spaces involved usually clear by the context. Note that we do not need to consider the optimal Lipschitz constant, i.e. the minimum \(L\) such that (1.40) holds; however, the optimal one will give better results in terms of concentration of measure. In the following, the optimal Lipschitz constant will be denoted by
\[
\| F \|_{\text{Lip}} = \sup_{x \neq y} \frac{\delta(F(x), F(y))}{d(x, y)}.
\] (1.42)

### 1.4 General deviation inequalities

Let us consider now a probability measure \(\mu\) on the Borel sets of \((X, d)\) and a measurable real-valued function \(F\) on \((X, d)\). We say that \(m_F\) is a median of
Chapter 1. Concentration of Measure Phenomenon and Typicality

$F$ for $\mu$ if

$$\mu(F \leq m_F) \geq \frac{1}{2} \quad \text{and} \quad \mu(F \geq m_F) \geq \frac{1}{2}. \quad (1.43)$$

Observe that in general there can be more than one median. We can bound deviations of a continuous function $F$ on $(X, d)$ from its median by using its modulus of continuity, defined as

$$\omega_F(r) = \sup \{|F(x) - F(y)| : d(x, y) < r\}, \quad r > 0. \quad (1.44)$$

**Proposition 4** ([94]). Let $F$ be a continuous function on $(X, d)$ with modulus of continuity $\omega_F$. Then:

$$\mu(|F - m_F| > \omega_F(r)) \leq 2\alpha_\mu(r) \quad (1.45)$$

**Proof.** Since $A = \{F \leq m_F\}$ has $\mu(A) \geq 1/2$ by definition of $m_F$, we have

$$\mu(A_r) \geq 1 - \alpha_\mu(r) \quad (1.46)$$

for any enlargement $A_r = \{x \in X : d(x, A) < r\}$ of $A$, by the definition of $\alpha_\mu$. Now note that if $x \in A_r$ (but not in $A$), for any $y \in A$ we can write:

$$F(x) \leq F(y) + \omega_F(r) \leq m_F + \omega_F(r), \quad (1.47)$$

which implies that if $F(x) > m_F + \omega_F(r)$, then $x \in X \setminus A_r$. Then, by (1.46):

$$\mu(F > m_F + \omega_F(r)) \leq \alpha_\mu(r). \quad (1.48)$$

Analogously, we have that

$$\mu(F < m_F - \omega_F(r)) \leq \alpha_\mu(r), \quad (1.49)$$

so that putting the two inequalities together we get equation (1.45) \(\square\)

In practice one usually works with functions which are not only continuous, but also Lipschitz continuous and even derivable. In these cases the deviation inequality (1.45) simplifies. In fact, for Lipschitz functions $F$ on $(X, d)$ we clearly have $\omega_F(r) \leq \|F\|_{\text{Lip}} r$, so that equation (1.45) implies:

$$\mu(|F - m_F| > \|F\|_{\text{Lip}} r) \leq \mu(|F - m_F| > \omega_F(r)) \leq 2\alpha_\mu(r). \quad (1.50)$$

We obtain thus the following result.

**Theorem 3.** Let $\mu$ be a Borel probability measure on a metric space $(X, d)$. Let $F$ be a real-valued Lipschitz function on $(X, d)$, and let $m_F$ be one of its medians. Then, for every $r > 0$:

$$\mu(|F - m_F| > r) \leq 2\alpha_\mu \left( \frac{r}{\|F\|_{\text{Lip}}} \right). \quad (1.51)$$

Depending on the relative size of $\alpha_\mu$ and $\|F\|_{\text{Lip}}$, we may state that the Lipschitz function $F$ “concentrates” around one constant value on a portion of space which has a large measure.
1.5. Typicality

The inequality (1.51) describes the concentration of $F$ around one of its median values $m_F$. Intuitively, the concentration of measure means that the function $F$ is essentially constant on sets of measure close to one, which must then coincide with the expected value $\mathbb{E}[F]$. This intuition is formalized by the following proposition.

**Proposition 5** ([94]). Let $F$ be a measurable function on a probability space $(X, \mathcal{B}, \mu)$. Assume that for some $a_F \in \mathbb{R}$ and a non-negative function $\alpha$ on $\mathbb{R}_+$ such that $\alpha(r) \to 0$ as $r \to \infty$,

$$\mu(|F - a_F| \geq r) \leq \alpha(r) \quad (1.52)$$

for all $r > 0$. Then

$$\mu(|F - m_F| \geq r + r_0) \leq \alpha(r), \quad r > 0, \quad (1.53)$$

where $m_F$ is a median of $F$ for $\mu$ and where $r_0 > 0$ is such that $\alpha(r_0) < 1/2$. If moreover $\overline{\alpha} = \int_0^\infty \alpha(r) \, dr < \infty$, then $F$ is integrable, $|a_F - \mathbb{E}_\mu[F]| \leq \overline{\alpha}$ and for every $r > 0$,

$$\mu(|F - \mathbb{E}_\mu[F]| \geq r + \overline{\alpha}) \leq \alpha(r) \quad (1.54)$$

In particular, if $\alpha(r) \leq C e^{-r^p}$, $0 < p < \infty$, $r > 0$, then

$$\mu(|F - M| \geq r) \leq C' e^{-\kappa_p r^p}, \quad (1.55)$$

where $M$ is either the mean or a median of $F$ for $\mu$, and $C' > 0$ only depends on $C$ and $p$, while $\kappa_p > 0$ only depends on $p$.

Informally, what this proposition is telling us is the following: if deviations of some function $F$ from some constant value $a_F$ can be efficiently bounded by some $\alpha(r)$, then we are able to control also deviations of $F$ from any of its medians, (1.53) and from its average (1.54). Moreover, if these deviations can be exponentially bounded, then $F$ concentrates both around its average and its median, in the sense of (1.55), since it is essentially constant.

1.5 Typicality

We say that a property is typical in a system when it is true in the overwhelming majority of cases. In order to assess if a certain property is typical, we need to introduce then a probability measure on the set of states of the physical system under interest, which allows to probe the relative volume of the subset of states where that property holds. This is the approach followed for example in classical statistical mechanics, where randomness is not intrinsic, but it is introduced as a means to investigate emergent properties which do not depend on the particular details of the microstates. As a matter of fact, it could be argued that this view entails the spirit of the modern measure theoretic approach to probability theory.

More precisely, we are interested in the characterization of typical properties of “large” systems, where here “large” is to be intended in terms of the dimensionality of the state space, not in terms of the size of the physical
system, which is indeed a quantum system. Then, one introduces a family of probability metric spaces \( (S_n, \mu_n, d_n) \), where \( n \) represents the dimensionality and \( S_n \) is the set of states with that dimensionality; we say that concentration of measure holds if the corresponding concentration functions \( \alpha_n(r) \) satisfy (1.38) for each \( r > 0 \). In such a case, for a property of the system expressed in terms of a Lipschitz function \( F \) on the set of states \( S_n \), we can use Theorem 3 to write:

\[
\mu_n(|F - m_F| > r) \leq 2\alpha_n \left( \frac{r}{\|F\|_{\text{Lip}}} \right). \tag{1.56}
\]

At this point if \( \|F\|_{\text{Lip}} \) does not depend on \( n \), the right hand-side of (1.56) goes to zero as the dimension of the system grows, implying that \( F \) becomes typical. It can happen, however, that the \( \|F\|_{\text{Lip}} = L_n \) actually depends on the dimensionality of the system \( n \). In such case, one needs to check the behaviour of \( \alpha_n(r/L_n) \) for fixed \( r > 0 \) as \( n \) grows: if it goes to zero, then \( F \) becomes typical for large \( n \). If for example \((S_n, \mu_n, d_n)\) constitutes a Lévy family (which is the case for the Haar measure considered in the next chapter), i.e. the concentration function is in the form (1.39), then functions \( F \) with a Lipschitz constant \( L_n = o(\sqrt{n}) \) become typical for large values of \( n \).
Chapter 2

Random Quantum States and Haar measure

In this chapter we will introduce the invariant probability measure on the set of quantum states and describe its concentration properties. After an introduction on the mathematical description of quantum states in Section 2.1, we will discuss definition and properties of the Haar measure on the unitary group in Section 2.2. In Section 2.3 we will look in detail at a simple and efficient algorithm for the numerical sampling of random Haar unitaries. Then, in Section 2.4 we will focus in particular on the concentration properties characterizing the unitary group, which will be useful in this thesis. Finally, we will use the concentration property to discuss the typicality of quantum observables in Section 2.5

2.1 Quantum States

According to the axioms of quantum mechanics, the observables of a quantum system are described by self-adjoint operators acting on some Hilbert space $\mathcal{H}$. We denote the set of linear operators on $\mathcal{H}$ with $\mathcal{L}(\mathcal{H})$. In this thesis we will deal almost only with finite dimensional Hilbert spaces, $n = \dim \mathcal{H} < \infty$, so that $\mathcal{H} \simeq \mathbb{C}^n$ and the space $\mathcal{L}(\mathcal{H})$ can be identified with the set of $n \times n$ matrices $M_n(\mathbb{C})$. Hence, self-adjoint operators are Hermitian matrices $A = A^\dagger \in M_n(\mathbb{C})$, where $A^\dagger$ denotes the conjugate transpose matrix of $A$.

**Definition 3 (Quantum States).** A quantum state on $\mathcal{H}$ is defined by a positive self-adjoint operator $\rho = \rho^\dagger$ on $\mathcal{H}$ with $\operatorname{tr}\rho = 1$. We denote the set of quantum states on some Hilbert space $\mathcal{H}$ by $S(\mathcal{H})$.

The state of a quantum system determines the expected value of any observable $A$ on $\mathcal{H}$ according to

$$\langle A \rangle = \operatorname{tr}(\rho A). \quad (2.1)$$

The space $S(\mathcal{H})$ is a convex set whose extremal points are the states satisfying $\rho^2 = \rho$ and are called pure states. Pure states are one-dimensional projection operators which can be written as $\rho = |\psi\rangle\langle\psi|$ for some normalized vector $|\psi\rangle \in \mathcal{H}$. As for any convex set, each quantum state $\rho \in S(\mathcal{H})$ can be
written as a convex sum of extremal elements:

$$\rho = \sum_{k=1}^{m} p_k |\varphi_k\rangle\langle \varphi_k|,$$

(2.2)

with $p_k > 0$ for each $k$ and such that $\sum_{k=1}^{m} p_k = 1$. The decomposition (2.2) is not unique, i.e. the convex set $S(\mathcal{H})$ is not a simplex. Among the several decompositions of $\rho$ there is a particular one playing a special role, which is its spectral decomposition:

$$\rho = \sum_{j=1}^{n} \lambda_j |\psi_j\rangle\langle \psi_j|,$$

(2.3)

where $\{\lambda_j\}_{j=1}^{n}$ are the eigenvalues of $\rho$ (we extend the summation also to the vanishing eigenvalues for convenience), $\{|\psi_j\rangle\}_{j=1}^{n}$ are the corresponding eigenvectors and $n = \dim \mathcal{H}$.

![Figure 2.1: The simplex $\Delta_2 \in \mathbb{R}^3$, shaded in green, can be decomposed in $3! = 6$ pieces according to the arrangement of the components $\lambda_1, \lambda_2, \lambda_3$. The $\Delta^{1}_{1}$ component is shown in red in the figure.](image)
2.2. Haar measure on the Unitary Group

Due to the normalization condition $\text{tr}\rho = 1$, the spectrum of $\rho$ identifies a point on the standard $(n - 1)$-dimensional simplex sitting in $\mathbb{R}^n$, defined as
\[ \Delta_{n-1} := \left\{ (\lambda_1, \ldots, \lambda_n) \in \mathbb{R}^n : \lambda_j \geq 0, \sum_{j=1}^n \lambda_j = 1 \right\}. \] (2.4)

It is sometimes more convenient to work with the ordered eigenvalues $\{\lambda_j\}_{j=1}^n$, arranged as $\lambda_1 \geq \cdots \geq \lambda_n \geq 0$. In fact, the set $\Delta_{n-1}$ can be partitioned in $n!$ pieces each being one copy of the other except for the ordering of the components (see Figure 2.1). In particular, we define
\[ \Delta_{n-1}^\downarrow := \left\{ (\lambda_1, \ldots, \lambda_n) \in \mathbb{R}^n : \lambda_1 \geq \cdots \geq \lambda_n \geq 0, \sum_{j=1}^n \lambda_j = 1 \right\}. \] (2.5)

We will denote with $\lambda(\rho)$ the vector of the eigenvalues of $\rho$, arranged in decreasing order, i.e. $\lambda(\rho) = (\lambda_1^\downarrow, \ldots, \lambda_n^\downarrow) \in \Delta_{n-1}^\downarrow$. Note that for any unitary operator $U$ on $\mathcal{H}$, $\lambda(U\rho U^\dagger) = \lambda(\rho)$: the spectrum of a density matrix is invariant under the action of the unitary group $\mathcal{U}(\mathcal{H})$.

Exploiting the identification $\mathcal{L}(\mathcal{H}) \simeq \mathcal{M}_n(\mathbb{C})$ (by choosing a particular basis of $\mathcal{H}$ as a reference frame), we can write the spectral decomposition (2.3) as:
\[ \rho = U\Lambda U^\dagger, \] (2.6)
where $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_n)$ is a diagonal matrix containing the eigenvalues of $\rho$, while $U$ is a unitary matrix whose (orthonormal) columns are the eigenvectors of $\rho$ (in the chosen reference frame).

When considering a probability measure $\sigma$ on the set of quantum states, the decomposition (2.6), together with the fact that the action of $\mathcal{U}(\mathcal{H}) \simeq \mathcal{U}(n)$ leaves the spectrum unchanged, encourages us to consider probability measures in the factorized form:
\[ d\sigma(\rho) = d\nu(\lambda_1, \ldots, \lambda_n) \times d\mu(U), \] (2.7)
where $d\nu$ is a probability measure on the unit simplex $\Delta_{n-1}$, while $d\mu$ is a probability measure on the unitary group $\mathcal{U}(\mathcal{H}) \simeq \mathcal{U}(n)$. In the remaining of this chapter we will focus on the most natural choice for the measure over $\mathcal{U}(n)$ and its properties, while the probability measure of the eigenvalues over the unit simplex will depend on the particular model considered (we will see a physical relevant model in the next chapter). It is however worth to point out before going on that the factorization (2.7) is not only a choice of convenience, but as we will see later, it also arises naturally from a physical point of view.

2.2 Haar measure on the Unitary Group

Identifying minimal knowledge with maximal symmetry, the most natural probability measure over $\mathcal{U}(n)$ is the uniform distribution. To understand
how it is defined, we can take the easiest case, $U(1)$, consisting of complex numbers with unit modulus. These can be geometrically represented as a circle with unit radius in the complex plane, and a natural way to define the uniform distribution over this set is to require it to be invariant under rotations, that is, if we consider a subset $A \subset U(1)$, its measure probability must be the same as any set $e^{i\theta}A := \{e^{i\theta}u : u \in A\}$, for $\theta \in [0, 2\pi]$. The generalization to $n > 1$ goes along the same lines: one defines a uniform measure on $U(n)$ as a rotationally invariant measure, i.e. such that for each fixed $U \in U(n)$:

$$\mu_H(A) = \mu_H(UA) = \mu_H(AU)$$

(2.8)

for each measurable subset $A \subset U(n)$. The existence and uniqueness (up to a multiplicative factor) of such measure was proven by Haar in 1933 for any local compact Hausdorff topological group [75].

### 2.2.1 Distribution of the eigenvalues and level spacings

Since the eigenvalues of a unitary matrix lie on the unit circle of the complex plane $T = \{z \in \mathbb{C} : |z| = 1\}$, the joint probability of the eigenvalues is supported on $T^n$.

**Theorem 4** ([80], Lemma 4.2.1). The measure on $T^n$ induced by the Haar measure $\mu_H$ on $U(n)$ has the density

$$\frac{1}{(2\pi)^n n!} \prod_{i<j} |\lambda_i - \lambda_j|^2 = \frac{1}{(2\pi)^n n!} \prod_{i<j} |e^{i\theta_i} - e^{i\theta_j}|^2$$

(2.9)

with respect to $d\lambda_1 d\lambda_2 \ldots d\lambda_n$, where $d\lambda_j = d\theta_j$ and $\lambda = e^{i\theta_j}$.

Using standard calculations of random matrix theory [108], one can compute statistical correlations among the phases $\theta_j$ of the eigenvalues $\lambda_j$, starting from the joint probability density (2.9). The simplest correlation function to determine is the density of eigenvalues $\rho(\theta)$. As one might intuitively expect, the normalized density is uniform on the interval $[0, 2\pi]$:

$$\rho(\theta) = \frac{1}{2\pi}.$$ 

(2.10)

It is worth emphasizing that this does not imply that the eigenvalues are statistically uncorrelated. On the contrary, there is a repulsion phenomenon between eigenvalues at work here: it can be roughly grasped by looking at the joint eigenvalue density (3.25), which vanishes in configurations such that $\theta_i = \theta_j$. More formally, arranging the arguments of the eigenvalues in ascending order $\theta_1 \leq \theta_2 \leq \cdots \leq \theta_n$, the normalized spacings

$$s_j = \frac{n}{2\pi}(\theta_{j+1} - \theta_j), \quad j = 1, \ldots, n$$

(2.11)
are known to asymptotically follow (as \( n \to \infty \)) the universal distribution \([108]\):
\[
p(s) = \frac{32}{\pi^2} s^2 e^{-\frac{4}{\pi} s^2},
\]
satisfying in particular \( p(s) \to 0 \) as \( s \to 0 \). To sum up, the eigenvalues of a random Haar unitary matrix are uniformly distributed on the unit circle \( \mathbb{T} \) and tend to repel each other.

### 2.2.2 Distribution of the entries

We collect here some results needed for the computation of averages over the unitary group which will be used later. The average of a function \( f : \mathcal{U}(n) \to \mathbb{C} \) is defined as
\[
\mathbb{E}[f(U)] = \int f(U) \, d\mu_H(U)
\]
whenever this integral is defined. In this thesis we will be only interested in the moments of the matrix elements \( U_{ij} \), i.e. the averages of some powers of the matrix elements and their complex conjugates. To begin with, we give the next lemma, which shows that most of the multiple moments of the elements \( U_{ij} \) vanish. Since this is a consequence of the invariance property of the measure \( \mu_H \), it is instructive also to look at the (very simple) proof.

**Lemma 5** ([80], Lemma 4.2.2). Let \( l \in \mathbb{N}, i_1, \ldots, i_l, j_1, \ldots, j_l \in \{1, \ldots, n\} \) and \( k_1, \ldots, k_l, m_1, \ldots, m_l \in \mathbb{Z}^+ \). If either
\[
\sum_{r: \ i_r = i} (k_r - m_r) \neq 0 \quad \text{for some } 1 \leq i \leq n
\]
or
\[
\sum_{r: \ j_r = j} (k_r - m_r) \neq 0 \quad \text{for some } 1 \leq j \leq n,
\]
then:
\[
\mathbb{E} \left[ (U_{i_1 j_1}^{k_1} U_{i_1 j_1}^{-m_1})(U_{i_2 j_2}^{k_2} U_{i_2 j_2}^{-m_2}) \cdots (U_{i_l j_l}^{k_l} U_{i_l j_l}^{-m_l}) \right] = 0.
\]
In particular, if \( \sum_{r=1}^l (k_r - m_r) \neq 0 \) (this is the case when \( \sum_{r=1}^l (k_r + m_r) \) is odd), then (2.16) holds.

**Proof.** From the invariance of the Haar measure, we have that \( \mathbb{E}[f(U)] = \mathbb{E}[f(V U W)] \) for any measurable function \( f : \mathcal{U}(n) \to \mathbb{C} \) and any fixed unitaries \( V \) and \( W \). Considering in particular unitaries in the form \( V = \text{diag}(e^{i\theta_1}, \ldots, e^{i\theta_n}) \) and \( W = \text{diag}(e^{i\psi_1}, \ldots, e^{i\psi_n}) \), we obtain that
\[
\mathbb{E}[f(U)] = \mathbb{E}[f([e^{i\theta_1 + \psi_j} U_{ij}]_{1 \leq i, j \leq n})],
\]
for any \( \theta_i, \psi_j \in \mathbb{R} \). In particular, if for example there is an \( i \) such that \( h := \sum_{r: \ i_r = i} (k_r - m_r) \neq 0 \), we can use (2.17) to write:
\[
\mathbb{E} \left[ (U_{i_1 j_1}^{k_1} U_{i_1 j_1}^{-m_1})(U_{i_2 j_2}^{k_2} U_{i_2 j_2}^{-m_2}) \cdots (U_{i_l j_l}^{k_l} U_{i_l j_l}^{-m_l}) \right] = e^{i\theta_1} \mathbb{E} \left[ (U_{i_1 j_1}^{k_1} U_{i_1 j_1}^{-m_1})(U_{i_2 j_2}^{k_2} U_{i_2 j_2}^{-m_2}) \cdots (U_{i_l j_l}^{k_l} U_{i_l j_l}^{-m_l}) \right]
\]
for every $\theta \in \mathbb{R}$, which immediately gives the statement. \hfill \Box

Having clarified the vanishing moments of the matrix elements $U_{ij}$, we next collect (without proof) the only non-vanishing moments up to the fourth order.

**Lemma 6 ([80], Proposition 4.2.3).** Denoting with $U_{ij}$ the matrix elements of a Haar distributed random unitary matrix $U \in \mathcal{U}(n)$, the only non-vanishing moments up to the fourth order are given by:

\[
\begin{align*}
\mathbb{E}[|U_{ij}|^2] &= \frac{1}{n} & (1 \leq i, j \leq n), \\
\mathbb{E}[|U_{ij}|^4] &= \frac{2}{n(n+1)} & (1 \leq i, j \leq n), \\
\mathbb{E}[|U_{ij}|^2|U_{kj}|^2] &= \frac{1}{n(n+1)} & (i \neq k), \\
\mathbb{E}[|U_{ij}|^2|U_{il}|^2] &= \frac{1}{n(n+1)} & (j \neq l), \\
\mathbb{E}[|U_{ij}|^2|U_{kl}|^2] &= \frac{1}{n^2 - 1} & (i \neq k, j \neq l), \\
\mathbb{E}[U_{ij}U_{kl}U_{il}^*U_{kj}] &= -\frac{1}{n(n^2 - 1)} & (i \neq k, j \neq l).
\end{align*}
\]

The moments gathered in the previous lemma can be expressed with two compact formulas [131] as:

\[
\begin{align*}
\mathbb{E}[U_{ij}U_{kl}^*] &= \frac{\delta_{ik}\delta_{jl}}{n} \quad (2.25) \\
\mathbb{E}[U_{ij}U_{kl}U_{pq}U_{rs}^*] &= \frac{\delta_{ip}\delta_{jq}\delta_{kr}\delta_{ls} + \delta_{ir}\delta_{js}\delta_{kp}\delta_{lq} - \delta_{ip}\delta_{js}\delta_{kr}\delta_{lq} + \delta_{ir}\delta_{jq}\delta_{kp}\delta_{ls}}{n(n^2 - 1)} \quad (2.26)
\end{align*}
\]

We conclude this collection on the statistical properties of the matrix elements of Haar unitaries with the actual distribution of $U_{ij}$, given by the following lemma.

**Lemma 7.** Let $U \in \mathcal{U}(n)$ a Haar random unitary matrix. Then, the distribution of each matrix element $U_{ij} = re^{i\theta}$ is given by:

\[
d\mu(r, \theta) = \frac{n-1}{\pi}(1-r^2)^{n-2}rdrd\theta, \quad 0 \leq r \leq 1, \quad 0 \leq \theta \leq 2\pi. \quad (2.27)
\]

Furthermore, for each $k \in \mathbb{Z}^+$:

\[
\mathbb{E}[|U_{ij}|^{2k}] = \left(\frac{n+k-1}{n-1}\right)^{-1}. \quad (2.28)
\]

### 2.3 Sampling Haar distributed unitary matrices

The invariant properties of the Haar measure can be used for deriving an efficient algorithm to generate random Haar unitary matrices from $\mathcal{U}(n)$ [109].
2.3. Sampling Haar distributed unitary matrices

The starting point of the algorithm is to sample matrices for the Ginibre ensemble, which is supported on the group $GL(n, \mathbb{C})$ of all the invertible $n \times n$ matrices with complex entries. The matrix elements are i.i.d. standard normal complex random variables, i.e. $z_{jk} = x_{jk} + i y_{jk}$ with $x_{jk}, y_{jk} \sim \mathcal{N}(0, 1/2)$, hence having the probability density function

$$p(z_{jk}) = \frac{1}{\pi} e^{-|z_{jk}|^2}, \quad j, k = 1, \ldots, n. \quad (2.29)$$

It can be shown [46] that the columns of a matrix sampled with this distribution are almost surely linearly independent from each other, and as a consequence the resulting matrix is almost surely invertible, i.e. the probability measure is supported on $GL(n, \mathbb{C})$. Since the matrix entries are statistically independent, the joint probability density function for the matrix is

$$P(Z) = \frac{1}{\pi^{n^2}} \prod_{j,k=1}^{n} e^{-|z_{jk}|^2} = \frac{1}{\pi^{n^2}} e^{-\text{tr}(Z^\dagger Z)}. \quad (2.30)$$

The corresponding measure on $GL(n, \mathbb{C})$ describing the Ginibre ensemble is then given by

$$d\mu_G(Z) = P(Z) dZ, \quad (2.31)$$

where $dZ = \prod_{j,k=1}^{n} dx_{jk} dy_{jk}$.

\begin{lemma}
The Ginibre ensemble is invariant under left and right multiplication by arbitrary unitary matrices:

$$d\mu_G(Z) = d\mu_G(UZ) = d\mu_G(ZV), \quad (2.32)$$

for every $U, V \in U(n)$.
\end{lemma}

\begin{proof}
The probability density function $P(Z)$ is manifestly invariant under left and right multiplication with unitary matrices, i.e. $P(UZV) = P(Z)$ since

$$\text{tr}(V^\dagger Z^\dagger U^\dagger UZV) = \text{tr}(Z^\dagger Z).$$

Then, from (2.31), we only need to prove that the Jacobian of the transformations

$$Z \mapsto UZ, \quad Z \mapsto ZV, \quad (2.33)$$

is equal to one. In fact, considering the isomorphism between $n \times n$ matrices and $n^2$ vectors:

$$Z = \begin{pmatrix} z_1 & \cdots & z_n \\ \vdots & \ddots & \vdots \\ z_1 & \cdots & z_n \end{pmatrix} \mapsto z = \begin{pmatrix} z_1 \\ \vdots \\ z_n \end{pmatrix}, \quad (2.34)$$

the transformation $Z \mapsto UZ$ corresponds to a unitary transformation on $n^2$ components vectors given by

$$X = \begin{pmatrix} U & \cdots & U \end{pmatrix} = U \oplus U \oplus \cdots \oplus U, \quad (2.35)$$

hence having $|\det X| = 1$. \hfill \Box

Applying the Gram-Schmidt orthonormalization procedure to columns of a matrix of full rank $Z$, the resulting matrix $Q$ is unitary. In particular, it can be seen that after the Gram-Schmidt procedure is completed, the matrix $Z$ can be written as

$$Z = QR, \quad (2.36)$$

where $R$ is an upper-triangular invertible matrix with positive diagonal entries [46]. In other words, the Gram-Schmidt procedure realizes the QR decomposition of the matrix. It has been proven [46] that if the matrix $Z$ is sampled from the Ginibre ensemble, the corresponding matrix $Q$ arising from the Gram-Schmidt orthonormalization procedure is distributed according to the Haar measure. Basically, this is a consequence of the fact that when the QR decomposition (2.36) arises from the Gram-Schmidt orthonormalization procedure, the map $Z \mapsto (Q,R)$ is such that

$$UZ \mapsto (UQ,R), \quad \forall U \in U(n), \quad (2.37)$$

and this implies that the invariance of the distribution of $Z$ is inherited by the distribution of $Q$.

**Theorem 5.** Let $Z \in \text{GL}(n, \mathbb{C})$ be a random matrix from the Ginibre ensemble, and let $Z \mapsto (Q,R)$ be a QR decomposition satisfying (2.37). Then, the distribution $\nu$ induced on $Q$ by the distribution $\mu_G$ on $Z$ is unitarily invariant.

**Proof.** Let us write $Q \equiv Q(Z)$, to denote explicitly the dependence of $Q$ on $Z$. Then, equation (2.37) can be written as

$$Q(UZ) = UQ(Z), \quad \forall U \in U(n). \quad (2.38)$$

Then, we denote with $\nu$ the distribution on $Q$ induced by the Ginibre distribution $\mu_G$ on $Z$, i.e:

$$\nu(A) = \mu_G(Q(Z) \in A) \quad (2.39)$$

for every measurable set $A \in U(n)$. The unitary invariance of $\nu$ is now a direct consequence of the unitary invariance of $\mu_G$, since for every measurable set $A \subset U(n)$ and any unitary $U \in U(n)$ we have that:

$$\nu(UA) = \mu_G(Q(Z) \in UA) = \mu_G(Q(UZ) \in UA) = \mu_G(UQ(Z) \in UA) = \mu_G(Q(Z) \in A) = \nu(A). \quad (2.40)$$

\hfill \Box
As a consequence of the previous theorem, it would appear that we have found an algorithm to sample a Haar unitary matrix: first we sample a matrix $Z$ with complex Gaussian entries, and then we apply the Gram-Schmidt orthonormalization procedure to obtain a Haar unitary matrix $Q$. However, the problem with this algorithm is that the Gram-Schmidt procedure is numerically unstable. In fact, the algorithm commonly used numerically to implement the QR decomposition (2.36) is based on the Householder reflections, which, on the contrary, are numerically stable; using these algorithms, the property (2.37) is lost. Essentially, the problem here is that the QR decomposition is not unique. In fact, if $Z = QR$ is a QR decomposition of $Z$, then for any unitary diagonal matrix $\Lambda = \text{diag}(e^{i\theta_1}, \ldots, e^{i\theta_n})$, the matrices

$$Q' = Q\Lambda, \quad R' = \Lambda^{-1}R$$

(2.41)

realize another QR decomposition of $Z$, i.e. $Z = Q'R'$. Viceversa, if $Z = QR = Q'R'$ with $Q, Q'$ unitary and $R, R'$ upper triangular and invertible, then there exists a diagonal unitary matrix $\Lambda$ such that (2.41) holds [109]. As shown in [109], this problem can be removed by fixing the diagonal elements of $R$ to be real and positive. This is done by taking the output $(Q, R)$ of the QR decomposition (obtained with any algorithm), and substituting it with $(Q', R')$ defined as in (2.41) with the matrix $\Lambda$ chosen in such a way that the diagonal elements of $R$ are real and positive, i.e:

$$\Lambda = \begin{pmatrix} R_{11} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & R_{nn} \end{pmatrix}.$$  

(2.42)

In this way, the diagonal elements of $R' = \Lambda^{-1}$ are set to be positive, and the QR decomposition $(Q', R')$ is the same which would have been obtained with a Gram-Schmidt orthonormalization procedure.

To sum up, the algorithm consists of:

1) Sample a $n \times n$ matrix $Z$ with complex standard normal entries.

2) Perform any QR decomposition on the matrix $Z$ and let $(Q, R)$ be the output, $Z = QR$.

3) Create the matrix $\Lambda$ as in equation (2.42), starting from the diagonal entries of $R$.

4) Take the matrix $U = QA$, which is distributed according to the Haar measure.

We report in the following the MATLAB implementation of this algorithm.
Algorithm 1 Sampling Haar Unitary Matrices

% Sample a \( n \times n \) matrix from the Ginibre ensemble
Z=(randn(n)+1i*randn(n))/sqrt(2);
% Factorize the matrix with QR decomposition
[Q,R]=qr(Z);
% Remove the arbitrariness in the QR decomposition
L= diag(diag(R)./asb(diag(R)));
U=Q*L;

2.4 Concentration of Measure on the Unitary Group

The \( \mathcal{U}(n) \) can be seen as a submanifold of the Euclidean vector space given by the set \( \mathcal{M}_n(\mathbb{C}) \) endowed with the Hilbert-Schmidt inner product

\[
\langle A, B \rangle_{HS} := \tr(A^\dagger B) \quad \forall A, B \in \mathcal{M}_n(\mathbb{C}).
\]  

(2.43)

The Hilbert-Schmidt inner product induces a norm on \( \mathcal{M}_n(\mathbb{C}) \), by:

\[
\|A\|_{HS} := \sqrt{\langle A, A \rangle} = \sqrt{\tr(A^\dagger A)}, \quad \forall A \in \mathcal{M}_n(\mathbb{C}),
\]  

(2.44)

which is unitarily invariant, i.e.

\[
\|UAU^\dagger\|_{HS} = \|A\|_{HS},
\]  

(2.45)

for any \( A \in \mathcal{M}_n(\mathbb{C}) \) and any \( U, V \in \mathcal{U}(n) \). A distance between matrices can be defined using this norm, by:

\[
d_{HS}(A, B) := \|A - B\|_{HS}.
\]  

(2.46)

Then we can see the space \( \mathcal{U}(n) \) as a metric space with the distance \( d_{HS} \) inherited from \( \mathcal{M}_n(\mathbb{C}) \). We will see now that the Haar probability measure \( \mu_H \) on \( \mathcal{U}(n) \) exhibits a concentration phenomenon with respect to the Euclidean distance \( d_{HS} \). This result can be proven, for example, using the fact that the measure metric space \( (\mathcal{U}(\mathcal{H}), \mu_H, d_{HS}) \) satisfies a logarithmic Sobolev inequality.

Definition 4. The space \( (X, d, \mu) \) is said to satisfy a logarithmic Sobolev inequality (or log-Sobolev inequality) with constant \( C > 0 \) if, for every locally Lipschitz function \( f : X \to \mathbb{R} \),

\[
\text{Ent}(f^2) \leq 2C\mathbb{E}[|\nabla f|^2],
\]  

(2.47)

where

\[
\text{Ent}(f) := \mathbb{E}[f \log(f)] - \mathbb{E}[f] \log(\mathbb{E}[f]).
\]  

(2.48)

and

\[
|\nabla f|(x) := \limsup_{y \to x} \frac{|f(y) - f(x)|}{d(y, x)}.
\]  

(2.49)

It can be proven that log-Sobolev spaces satisfy a concentration inequality. This is expressed by the following theorem, whose derivation is discussed in Appendix A for the sake of completeness.
2.4. Concentration of Measure on the Unitary Group

Figure 2.2: (Left) Distribution of the eigenphases and (Right) distribution of the normalized spacings between the phases, as defined in (2.11), computed numerically in MATLAB with the Algorithm 1.

**Theorem 6.** Let \((X, d, \mu)\) satisfy a log-Sobolev inequality with constant \(C > 0\). Then, every Lipschitz function \(F : X \to \mathbb{R}\) with \(\|F\|_{\text{Lip}} = 1\) has \(\mathbb{E}[F] < \infty\), and:

\[
\mu(|F - \mathbb{E}[F]| \geq r) \leq 2e^{-\frac{r^2}{2C}}.
\]  

(2.50)

In particular, the concentration function satisfies

\[
\alpha(r) \leq e^{-\frac{r^2}{2C}}, \quad r > 0.
\]  

(2.51)

This theorem can be used to derive the concentration of the Haar measure on the unitary group, once we know that \((\mathcal{U}(\mathcal{H}), \mu_H, d_{HS})\) satisfies a log-Sobolev inequality, which follows from the following Lemma.

**Lemma 9 ([107]).** The matrix group \(\mathcal{U}(n)\) endowed with the Haar probability measure \(\mu_H\) and the Hilbert-Schmidt metric satisfies a log-Sobolev inequality (A.9) with \(C = 6/n\).
Chapter 2. Random Quantum States and Haar measure

Then, the concentration of $\mu_H$ on $U(n)$ follows as a simple application of Theorem 6. We state the theorem here explicitly for convenience.

**Theorem 7.** Let $F : U(n) \to \mathbb{R}$ be a Lipschitz function with lipschitz constant $L$, i.e. such that

$$|F(U) - F(V)| \leq L d_{HS}(U, V),$$

(2.52)

for each $U, V \in U(n)$. Then:

$$\mu_H(|F(U) - \mathbb{E}[F]| \geq \varepsilon) \leq 2e^{-\frac{n \varepsilon^2}{12L^2}}.$$  

(2.53)

### 2.5 Typicality of Observables on Random Subspaces

We will now apply the results on the unitary group discussed in this chapter to derive a concentration result on the expectation value of a quantum observable. Let us then consider a quantum system described in some finite dimensional Hilbert space $\mathcal{H} \simeq \mathbb{C}^n$. The observables of the system are Hermitian matrices $A = A^\dagger \in M_n(\mathbb{C})$. The expectation value of some observable $A$ in state $\rho$ is given by (2.1). Let us consider some subspace $\mathcal{S} \subset \mathcal{H}$ and denote with $P = P^\dagger = P^2$ the projection operator on $\mathcal{S}$, $P \mathcal{H} = \mathcal{S}$. If we know that the state of the system lies in $\mathcal{S}$ but we do not know anything else, the most natural assumption is to assume that the system is described by the maximally mixed state on $\mathcal{S}$, given by $\rho = \frac{P}{r}$, where $r = \dim \mathcal{S} = \text{tr} P$. We will call

$$\langle A \rangle = \text{tr}(\rho A) = \frac{\text{tr}(AP)}{\text{tr} P}$$

(2.54)

the expectation value of $A$ on the subspace $\mathcal{S}$. In the particular case of a one-dimensional subspace, i.e. $\mathcal{S} = \text{span}\{\psi\}$ for some state $|\psi\rangle \in \mathcal{H}$, this is simply the expectation value $\langle \psi | A | \psi \rangle$. Let us ask now what happens if the subspace is chosen randomly according to the uniform measure induced by the action of a Haar distributed random unitary, i.e. we want to consider the random subspaces

$$\mathcal{S}_U := U \mathcal{S} = \{U |\psi\rangle : |\psi\rangle \in \mathcal{S}\},$$

(2.55)

with random $U \in U(\mathcal{H}) \simeq U(n)$, distributed according to the Haar measure $\mu_H$. Accordingly, we will have the corresponding projection $P_U = UPU^\dagger$ and the normalized uniform state on $\mathcal{S}_U$ given by $\rho_U = P_U / r$. We will denote by:

$$\langle A \rangle_U = \text{tr}(\rho_U A) = \frac{\text{tr}(AUP_U^\dagger)}{\text{tr} P},$$

(2.56)

$$\langle (\Delta A)^2 \rangle_U = \text{tr}[(A - \langle A \rangle_U)^2 \rho_U] = \langle A^2 \rangle_U - \langle A \rangle_U^2,$$

(2.57)

respectively the expected value of $A$ and its variance on the random subspace $\mathcal{S}_U$ (on the state $\rho_U$).
2.5. Typicality of Observables on Random Subspaces

2.5.1 Averages of Expectation and Variance

**Lemma 10.** Let $A \in \mathcal{M}_n(\mathbb{C})$ be a fixed hermitian matrix, $A = A^\dagger$. Let $U \in \mathcal{U}(n)$ be a random Haar unitary. Then, with the previous notation:

\[
\mathbb{E}[\langle A \rangle_U] = \frac{\text{tr}(A)}{n},
\]  
(2.58)

\[
\mathbb{E}[\langle (\Delta A)^2 \rangle_U] = \frac{nr - 1}{r(n^2 - 1)} \left[ \text{tr}(A^2) - \frac{\text{tr}(A)^2}{n} \right]
\]  
(2.59)

**Proof.** Choosing some orthonormal basis $\{ |e_j \rangle \}_{j=1}^r$ on $\mathcal{H}$ and completing it to an orthonormal basis $\{ |e_j \rangle \}_{j=1}^n$ on $\mathcal{H}$, we can write

\[
P = \sum_{j=1}^r |e_j \rangle \langle e_j|,
\]  
so that, by equation (2.25) we obtain:

\[
\langle A \rangle_U = \frac{1}{r} \sum_{j=1}^r \sum_{k,l=1}^n A_{kl} U_{lj} U_{kj}^*,
\]  
(2.60)

so that, by equation (2.25) we obtain:

\[
\mathbb{E}[\langle A \rangle_U] = \frac{1}{r} \sum_{j=1}^r \sum_{k,l=1}^n A_{kl} \mathbb{E}[U_{lj} U_{kj}^*] = \frac{1}{r} \sum_{j=1}^r \sum_{k,l=1}^n A_{kl} \delta_{kl} = \frac{\text{tr}(A)}{n}.
\]  
(2.61)

To obtain (2.57), we note using the result just obtained for $A$ with $A^2$ we have

\[
\mathbb{E}[\langle (\Delta A)^2 \rangle_U] = \mathbb{E}[\langle A^2 \rangle_U] - \mathbb{E}[\langle A \rangle_U^2]
\]  
(2.62)

Then, we only need to compute the second term:

\[
\mathbb{E}[\langle A^2 \rangle_U] = \frac{1}{r^2} \sum_{j=1}^r \sum_{k,l=1}^n \sum_{j'=1}^r \sum_{k',l'=1}^n A_{kl} A_{k'l'} \mathbb{E}[U_{lj} U_{lj'} U_{kj}^* U_{kj'}^*]
\]  

\[
= \frac{\text{tr}(A)^2}{n^2 - 1} + \frac{\text{tr}(A^2)}{r(n^2 - 1)} - \frac{\text{tr}(A)^2}{nr(n^2 - 1)} - \frac{\text{tr}(A^2)}{n(n^2 - 1)}
\]  

\[
= \frac{n - r}{nr(n^2 - 1)} \text{tr}(A^2) + \frac{nr - 1}{nr(n^2 - 1)} \text{tr}(A)^2,
\]  
(2.63)

where we used (2.26) to compute the average. Inserting this result into (2.62) we finally obtain

\[
\mathbb{E}[\langle (\Delta A)^2 \rangle_U] = \left[ \frac{1}{n} - \frac{n - r}{nr(n^2 - 1)} \right] \text{tr}(A^2) - \frac{nr - 1}{nr(n^2 - 1)} \text{tr}(A)^2
\]  
(2.64)

\[
= \frac{nr - 1}{r(n^2 - 1)} \left[ \text{tr}(A^2) - \frac{\text{tr}(A)^2}{n} \right]
\]  
(2.65)

\[\square\]
2.5.2 Lipschitz constants

Lemma 11. Let $A, B \in \mathcal{M}_n(\mathbb{C})$ be Hermitian, and let $P$ be a $r$-dimensional projection, i.e. $P = P^2 = P^\dagger$, $\text{Tr} \ P = r$. Then, the two functions $f_1, f_2 : U(n) \to \mathbb{R}$ defined as

\begin{align*}
  f_1(U) &= \text{tr}(U^\dagger AU P) \\
  f_2(U) &= \text{tr}(U^\dagger BU P)^2
\end{align*}

are Lipschitz functions with respect to the Hilbert-Schmidt metric on $U(n)$, i.e. satisfy

\begin{equation}
  |f_j(U) - f_j(V)| \leq L_j \|U - V\|_{HS} \quad j = 1, 2,
\end{equation}

for every $U, V \in U(n)$ with $L_1 = 2r^{1/2}\|A\|$ and $L_2 = 4r^{3/2}\|B\|^2$, with $\|\cdot\|$ denoting the operator norm on $\mathcal{M}_n(\mathbb{C})$.

Proof. First of all note that we can extend the definition of the functions $f_1$ and $f_2$ to the whole set $\mathcal{M}_n(\mathbb{C})$, so that they can be evaluated on all the points of the path

\begin{equation}
  Z(t) = tU + (1 - t)V, \quad t \in [0, 1]
\end{equation}

going from $Z(0) = V$ to $Z(1) = U$. Then, we can write

\begin{equation}
  |f_j(U) - f_j(V)| = \left| \int_0^1 \frac{d}{dt} f_j(Z(t)) \ dt \right| \leq \int_0^1 \left| \frac{d}{dt} f_j(Z(t)) \right| \ dt,
\end{equation}

Concerning $f_1$, the derivative can be easily evaluated as

\begin{align*}
  \frac{d}{dt} f_1(Z(t)) &= \frac{d}{dt} \text{tr}[Z(t)^\dagger AZ(t)P] \\
  &= \text{tr}[(U - V)^\dagger AZ(t)P] + \text{tr}[Z(t)^\dagger A(U - V)P].
\end{align*}

Substituting equation (2.71) into equation (2.70) we have:

\begin{align*}
  |f_1(U) - f_1(V)| &\leq \int_0^1 \left| \text{tr}[(U - V)^\dagger AZ(t)P] \right| \ dt + \int_0^1 \left| \text{tr}[P(Z(t)^\dagger A(U - V))P] \right| \ dt \\
  &\leq \int_0^1 (\|AZ(t)P\|_{HS} + \|PZ(t)^\dagger A\|_{HS}) \|U - V\|_{HS} \ dt,
\end{align*}

where we used the Cauchy-Schwarz inequality

\begin{equation}
  |\text{tr}(A^\dagger B)| \leq \text{tr}(A^\dagger A)^{1/2} \text{tr}(B^\dagger B)^{1/2} \equiv \|A\|_{HS} \|B\|_{HS}.
\end{equation}

Then, using the property

\begin{equation}
  \|AB\|_{HS} \leq \|A\| \|B\|_{HS},
\end{equation}

and the fact that \( \| P \|_{HS} = \text{tr}(P^\dagger P)^{1/2} = r^{1/2} \), we finally obtain

\[
\begin{align*}
|f_1(U) - f_1(V)| & \leq 2 \| A \| \left( \int_0^1 \| Z(t) \| \| P \|_{HS} \, dt \right) \| U - V \|_{HS} \\
& \leq 2r^{1/2} \| A \| \left( \int_0^1 (t \| U \| + (1 - t) \| V \|) \, dt \right) \| U - V \|_{HS} \\
& = 2r^{1/2} \| A \| \| U - V \|_{HS},
\end{align*}
\]

(2.74)

where in the last equality we used the fact that \( \| U \| = \| V \| = 1 \) by unitarity. This proves the assertion for \( f_1 \).

An analogous procedure can be carried on with \( f_2 \), starting from the fact that the derivative in such a case is given by:

\[
\frac{d}{dt} f_2(Z(t)) = 2\text{tr}[Z(t)^\dagger BZ(t)P](\text{tr}[(U - V)^\dagger BZ(t)P] + \text{tr}[Z(t)^\dagger B(U - V)P]),
\]

(2.75)

which substituted into equation (2.70) yields:

\[
\begin{align*}
|f_2(U) - f_2(V)| & \leq 2 \int_0^1 |\text{tr}[Z(t)^\dagger BZ(t)P]| |\text{tr}[(U - V)^\dagger BZ(t)P]| \, dt \\
& + 2 \int_0^1 |\text{tr}[Z(t)^\dagger BZ(t)P]| |\text{tr}[Z(t)^\dagger B(U - V)P]| \, dt \\
& \leq 2 \int_0^1 \| Z(t)P \|_{HS} \| BZ(t)P \|_{HS}^2 \| U - V \|_{HS} \, dt, \\
& + 2 \int_0^1 \| Z(t)P \|_{HS} \| BZ(t)P \|_{HS} \| PZ(t)B \|_{HS} \| U - V \|_{HS} \, dt, \\
& \leq 4 \| B \|^2 \left( \int_0^1 \| Z(t) \|^3 \| P \|^3_{HS} \, dt \right) \| U - V \|_{HS} \\
& \leq 4r^{3/2} \| B \|^2 \left( \int_0^1 (t \| U \| + (1 - t) \| V \|)^3 \, dt \right) \| U - V \|_{HS} \\
& = 4r^{3/2} \| B \|^2 \| U - V \|_{HS}.
\end{align*}
\]

(2.76)

The results of the previous lemma allow us to obtain some upper bound for the Lipschitz constants of the functions \( U \mapsto \langle A \rangle_U \) and \( U \mapsto \langle (\Delta A)^2 \rangle_U \) defined by (2.56) and (2.57) respectively. In particular, for the expectation value we obtain:

\[
|\langle A \rangle_U - \langle A \rangle_V| = \frac{1}{r} |\text{tr}(U^\dagger AUP) - \text{tr}(V^\dagger AVP)| \leq \frac{2\| A \|}{r^{1/2}} \| U - V \|_{HS}
\]

(2.77)
while for the variance:
\[
|\langle (\Delta A)^2 \rangle_U - \langle (\Delta A)^2 \rangle_V| \leq |\langle A^2 \rangle_U - \langle A^2 \rangle_V| + |\langle A \rangle_U^2 - \langle A \rangle_V^2|
\]
\[
\leq \frac{2\|A\|}{r^{1/2}} \|U - V\|_{HS} + \frac{1}{r^2} |\text{tr}(A^2 U) - \text{tr}(A^2 V)|
\]
\[
\leq \left( \frac{2\|A\|}{r^{1/2}} + \frac{4\|A\|^2}{r^{1/2}} \right) \|U - V\|_{HS}
\]
\[
\leq 6\|A\|^2 \frac{1}{r^{1/2}} \|U - V\|_{HS}
\]  \hspace{1cm} (2.78)

Using these bounds for the Lipschitz constants of the functions \( U \mapsto \langle A \rangle_U \) and \( U \mapsto \langle (\Delta A)^2 \rangle_U \) and the averages (2.58) and (2.59), we obtain the following \textit{typicality} result for the observable \( A \) as a consequence of the concentration result given by Theorem 7:

\[
\mu_H \left( \left| \frac{\langle A \rangle_U}{n} - \frac{\text{tr}(A)}{n} \right| \geq \varepsilon \right) \leq 2 \exp \left( -\frac{rn}{48\|A\|^2} \varepsilon^2 \right),
\]  \hspace{1cm} (2.80)

\[
\mu_H \left( \left| \frac{\langle (\Delta A)^2 \rangle_U - \langle (\Delta A)^2 \rangle_V}{n} - \frac{\text{tr}(A^2) - \text{tr}(A^2)}{n} \right| \geq \varepsilon \right) \leq 2 \exp \left( -\frac{rn}{432\|A\|^2} \varepsilon^2 \right).
\]  \hspace{1cm} (2.81)
Chapter 3

Random states on Bipartite Quantum Systems

This chapter discusses the peculiarities of random states of bipartite systems and introduces useful techniques from random matrix theory which can be used to investigate their properties. After a brief introduction on the formal description of bipartite quantum systems in Section 3.1, we will introduce the unbiased ensemble in Section 3.2: the typicality of some entanglement measures will be discussed, followed by the characterization of the joint distribution of the Schmidt coefficients associated to the ensemble. Then, in Section 3.3, we will show a connection between the distribution of the Schmidt coefficients of the unbiased ensemble and the distribution of the eigenvalues of Wishart matrices; we will see that this connection can be exploited for an efficient sampling of the Schmidt coefficients from the unbiased ensemble, using theoretical results from random matrix theory. Finally, in Section 3.4 more details on the spectrum of Wishart matrices will be provided, which will be used for the discussion of results presented in Chapter 5.

3.1 Bipartite quantum systems

According to the fundamental postulates of Quantum Mechanics, two subsystems $\mathcal{H}_A$ and $\mathcal{H}_B$ can be studied together by taking their Hilbert tensor product $\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_B$. The term “subsystems” here can refer either to the fact that the systems are physically separated (two spatially separated atoms) or also to different degrees of freedom pertaining to the same physical system (e.g. spin and orbital degrees of freedom).

Given a state $\rho \in S(\mathcal{H}_A \otimes \mathcal{H}_B)$, the states obtained by partial tracing over one subsystem $\rho_A = \text{tr}_B(\rho)$ and $\rho_B = \text{tr}_A(\rho)$ are said to be the reduced state of $A$ and $B$ respectively. In this chapter, we will always denote with $n = \dim \mathcal{H}_A$ and $m = \dim \mathcal{H}_B$.

3.1.1 Entanglement and separability

Definition 5. A state $\rho$ is said to be separable with respect to the bipartition $\mathcal{H}_A \otimes \mathcal{H}_B$ if it can be written as

$$\rho = \sum_{i=1}^{s} p_i \rho_i^A \otimes \rho_i^B.$$  \hspace{1cm} (3.1)
where $\rho^A_i \in S(\mathcal{H}_A)$ and $\rho^B_i \in S(\mathcal{H}_B)$ for each $i = 1, \ldots, s$ and \{\(p_i\)\}_{i=1}^s$ are probability weights ($p_i \geq 0$ and $\sum_{i=1}^s p_i = 1$).

Every non-separable state is said to be entangled. For pure states the definition of separability simplifies, since $|\psi\rangle\langle\psi|$ is separable if and only if $|\psi\rangle = |\psi^A\rangle \otimes |\psi^B\rangle$, with $|\psi^A\rangle \in \mathcal{H}_A$ and $|\psi^B\rangle \in \mathcal{H}_B$.

**Proposition 6** (Schmidt decomposition). Every pure state $|\psi\rangle\langle\psi| \in S(\mathcal{H}_A \otimes \mathcal{H}_B)$ (suppose $n \leq m$) can be written as

$$|\psi\rangle = \sum_{i=1}^n \sqrt{\lambda_i} |e_i\rangle \otimes |f_i\rangle,$$

where $\{ |e_i\rangle \}_{i=1}^n$ and $\{ |f_i\rangle \}_{i=1}^m$ are orthonormal sets of $\mathcal{H}_A$ and $\mathcal{H}_B$ respectively and where $\lambda_i \geq 0$, $\sum_{i=1}^n \lambda_i = 1$. The Schmidt coefficients $\{ \lambda_i \}$ are the eigenvalues of $\rho_A$. The number of non-zero Schmidt coefficients of $|\psi\rangle$ is called Schmidt rank of $|\psi\rangle$.

A maximal entangled state is a state $|\psi\rangle$ with $r = \min(n, m)$ Schmidt coefficients all equal to $1/\sqrt{r}$, hence having the form

$$|\psi\rangle = \frac{1}{\sqrt{r}} \sum_{i=1}^r |e_i\rangle \otimes |f_i\rangle.$$  \tag{3.3}$$

A separable state has Schmidt rank $r = 1$, while a maximally entangled state has maximal Schmidt rank $r = \min(n, m)$. Between these two extremal cases, it is useful to have some quantitative measure of the “amount of entanglement” contained in some state.

### 3.1.2 Measures of Entanglement

It turns out that the “most suitable” measure of entanglement is given by the von-Neumann entropy of entanglement [41], defined as

$$S_{vN}(\psi_A) = -\text{tr}(\rho_A \log \rho_A),$$

where we use the notation

$$\psi = |\psi\rangle\langle\psi|$$  \quad and  \quad $\psi_A = \text{tr}_B(\psi)$ \tag{3.5}$$

for pure states $|\psi\rangle\langle\psi| \in S(\mathcal{H}_A \otimes \mathcal{H}_B)$. Note that for a pure separable state $\psi$ we have $S_{vN}(\psi_A) = 0$, while for a maximally entangled state such as (3.3) we have $S_{vN}(\psi_A) = \log r$, which is the maximum value attainable ($r = \min(n, m)$).

Using the fact that the reduced state of a separable pure state is still pure, while the reduced state is maximally mixed, we can get an intuition for another useful measure of entanglement, through the purity of the reduced state:

$$\pi(\psi_A) = \text{tr}(\rho_A^2),$$

where $\rho_A \in S(\mathcal{H}_A)$ and $\rho_B \in S(\mathcal{H}_B)$ for each $i = 1, \ldots, s$ and \{\(p_i\)\}_{i=1}^s$ are probability weights ($p_i \geq 0$ and $\sum_{i=1}^s p_i = 1$).
3.2 The unbiased ensemble

Let us now discuss the typical properties of random pure states of the bipartite system $\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_B$. The unbiased ensemble is defined by the absence of any prior knowledge on the state of the system, or in other words, by the fact that all the measurements are equivalent. Identifying minimal knowledge with maximal symmetry, the reasonable choice for the sampling pure random states without a priori information on the system is to use the Haar measure induced on the normalized sphere in $\mathcal{H}$ by the action of the unitary group $U(\mathcal{H})$. Then, random pure states of the unbiased ensemble are obtained by considering $|\psi\rangle = U|\psi_0\rangle$, where $|\psi_0\rangle$ is an arbitrarily fixed unit vector on $\mathcal{H}$, while $U$ is drawn from $U(\mathcal{H})$ according to the Haar measure $d\mu_H$. Thanks to the unitary invariance of $d\mu_H$, the initial choice of $|\psi_0\rangle$ is irrelevant.

3.2.1 Typicality of Entanglement Entropy and Purity

The concentration of the Haar measure over the unitary group is reflected also on the induced measure over $\mathcal{H}$.

**Theorem 8.** Let $d = \dim \mathcal{H}$ ($\mathcal{H} \simeq \mathbb{C}^d$), and denote with

$$S(\mathcal{H}) = \{ |\psi\rangle \in \mathcal{H} : \langle \psi | \psi \rangle = 1 \}$$

the unit sphere in $\mathcal{H}$. Let $f : S(\mathcal{H}) \to \mathbb{R}$ be a Lipschitz function with respect to the norm on $\mathcal{H}$, i.e. satisfying:

$$|f(|\psi\rangle) - f(|\varphi\rangle)| \leq L||\psi - |\varphi\rangle||.$$  \hfill (3.8)

Then, denoting with $\mathcal{P}$ the probability measure induced on $S(\mathcal{H})$ by the Haar measure on $U(\mathcal{H})$, we have that:

$$\mathcal{P}(|f(|\psi\rangle) - \mathbb{E}[f]| > \epsilon) \leq 2e^{-\frac{\epsilon^2}{2dL^2}}$$  \hfill (3.9)

**Proof.** This is a consequence of the concentration of $\mu_H$ on $U(\mathcal{H}) \simeq U(d)$. To see this, let us define the function $F : U(d) \to \mathbb{R}$ as

$$F(U) := f(U|\psi_0\rangle),$$  \hfill (3.10)
which is a Lipschitz function with respect to $d_{HS}$ with the same Lipschitz constant $L$ of $f$. In fact, for $U, V \in \mathcal{U}(\mathcal{H})$, we have that

$$|F(U) - F(V)| = |f(U | \psi_0) - f(V | \psi_0)|$$

$$\leq L\|U | \psi_0) - V | \varphi_0\|$$

$$\leq L\|U - V\|\|\psi_0\|$$

$$\leq L\|U - V\|_{HS}.$$

Then, we can apply Theorem 7 and obtain:

$$\mathcal{P}(|f(|\psi\rangle) - \mathbb{E}[f]| \geq \varepsilon) = \mu_H(|F(U) - \mathbb{E}[F]| \geq \varepsilon) \leq 2e^{-\frac{d\varepsilon^2}{12L^2}}$$

Lemma 12. Let $|\psi\rangle \in \mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_B$ be a random pure state distributed according to the uniform Haar measure over $\mathcal{H}$. Then:

$$\mathbb{E}[\pi(\psi_A)] = \frac{n + m}{nm + 1}$$

and

$$\mathbb{E}[S_{vN}(\psi_A)] = \sum_{k=1}^{nm} \frac{1}{k} - \frac{n - 1}{2m}.$$

The average local von Neumann entropy was conjectured by Page in [124] and later proved in [56] and [139]. The average value of the local purity was instead computed by Lubkin [99] using a diagrammatic method. These results show that for high dimensional systems ($n, m \geq 1$) the average entanglement for the unbiased ensemble is pretty large. More precisely, fixing for example the ratio $c = m/n$ (we can assume without loss of generality $c \geq 1$), we have that for $n \gg 1$:

$$\mathbb{E}[\pi(\psi_A)] \simeq \frac{1}{n} \left(1 + \frac{1}{c}\right)$$

which is close to the minimum value $1/n$ corresponding to the maximally entangled case. In particular, it gets more closer as $c$ increases, i.e. when the dimension of the “environment” $m$ grows with respect to that of the subsystem $n$. Analogously for the entropy, we can use the estimate

$$\frac{1}{2(d + 1)} < \sum_{j=1}^{d} \frac{1}{j} - \ln d - \gamma < \frac{1}{2d},$$

where $\gamma \simeq 0.577$ is Euler’s constant, to write the inequality:

$$\mathbb{E}[S_{vN}(\psi_A)] > \ln n - \frac{1}{2c},$$

which is close to its maximum value $\ln n$ for $n \gg 1$, and it gets closer as $c$ grows.
3.2. The unbiased ensemble

We can see that this results are not only valid on average but for typical random states, thanks to the fact that $\pi(\psi_A)$ and $S(\psi_A)$ are Lipschitz functions when considered as functions of normalized vectors $|\psi\rangle \in \mathcal{H}$.

**Lemma 13 ([77])**. The local purity $\pi(\psi_A)$ and the entanglement entropy $S_{vN}(\psi_A)$, satisfy the following inequalities for any $|\psi\rangle, |\varphi\rangle \in \mathcal{H}$:

\[
|\pi(\psi_A) - \pi(\varphi_A)| \leq 4||\psi\rangle - |\varphi\rangle||, \tag{3.21}
\]

and

\[
|S(\psi_A) - S(\varphi_A)| \leq \sqrt{8} \ln n ||\psi\rangle - |\varphi\rangle|| \tag{3.22}
\]

for $n \geq 3$.

Using this lemma and Theorem 8, we get the concentration results:

\[
\mathcal{P}( |S_{vN}(\psi_A) - \mathbb{E}[S_{vN}(\psi_A)] | > \varepsilon ) \leq 2 \exp \left( -\frac{nm}{96(ln n)^2} \varepsilon^2 \right) \tag{3.23}
\]

for $n \geq 3$, and

\[
\mathcal{P}( |\pi(\psi_A) - \mathbb{E}[\pi(\psi_A)] | > \varepsilon ) \leq 2 \exp \left( -\frac{nm}{192} \varepsilon^2 \right), \tag{3.24}
\]

Hence, using $S_{vN}(\psi_A)$ and $1 - \pi(\psi_A)$ as measures of entanglement, we see that high-dimensional bipartite systems are characterized by a large amount
Chapter 3. Random states on Bipartite Quantum Systems

of entanglement across the bipartition. In Figure 3.1, this phenomenon is shown by using numerical results obtained with $n = 2, 8$ and $m = cn$, for several values of $c$. Note that even if the dimension $n$ is fixed to a small value, e.g. $n = 2$, the concentration phenomenon can still be observed for $m \rightarrow \infty$, i.e. $c \rightarrow \infty$.

3.2.2 Distribution of Schmidts coefficients for the unbiased ensemble

Entanglement of pure bipartite states can be characterized entirely in terms of their Schmidt coefficients. It is important then to be able to characterize the distribution of these coefficients for random pure bipartite states (sampled according to the unbiased ensemble). The following theorem, essentially due to Lloyd and Pagels [96], is one of the central results in the theory of random quantum states. For a detailed proof, see [171].

**Theorem 9** (Lloyd and Pagels [96]). Let $|\psi\rangle$ be a state distributed according to the unitarily invariant measure on the unit sphere of $\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_B$, with $\mathcal{H}_A = \mathbb{C}^n$, $\mathcal{H}_B = \mathbb{C}^m$, and $n \leq m$. The local spectrum $\lambda(\psi_A)$ is a random $n$-tuple with probability density

$$g_{n,m}(\lambda) = c_{n,m} \prod_{1 \leq i < j \leq n} (\lambda_i - \lambda_j)^2 \prod_{1 \leq k \leq n} \lambda_k^{m-n} 1_{\Delta_{n-1}}(\lambda)$$

(3.25)

where $1_{\Delta_{n-1}}$ is the indicator function of the simplex, equation (2.4), and

$$c_{n,m} = \frac{\Gamma(nm)}{\prod_{j=1}^{n} \Gamma(j+1) \Gamma(m - n + j)}.$$  

(3.26)

By a first inspection of the probability density function (3.25), we immediately note two remarkable features of the local spectrum distribution (see also Figure 3.2):

- **Eigenvalue repulsion phenomenon**: configurations with equal eigenvalues $\lambda_i = \lambda_j$ are unlikely, due to the vanishing Vandermonde determinant squared $\prod_{i<j} (\lambda_i - \lambda_j)^2$.
- **Configurations with large magnitude of the eigenvalues**: are more and more unlikely as $m - n$ grows, due to the term $\prod_k \lambda_k^{m-n}$.

With the help of techniques from Random Matrix Theory [108, 95], we will see that a lot more can be derived on the local spectrum behaviour $\lambda(\psi_A)$, and we will see in Chapter 5 that this behaviour plays an important role in the characterization of Entanglement conversions.

3.3 Random pure states and Wishart Matrices

The distribution (3.25) closely resembles the joint eigenvalues distribution of Wishart random matrices.
Definition 6. Let $A \in \mathcal{M}_{n,m}(\mathbb{C})$ be a $n \times m$ random matrix with independent and identically distributed standard complex entries, i.e. $A_{jk} = X_{jk} + iY_{jk}$, with $X_{jk}, Y_{jk} \sim \mathcal{N}(0, 1/2)$ for each $j = 1, \ldots, n$ and $k = 1, \ldots, m$. Then $W = AA^\dagger$ is called a complex Wishart random matrix of size $n$ and parameter $m$.

The joint distribution of the eigenvalues of a complex Wishart random matrix is [114, 45, 95]

$$G_{n,m}(\lambda) = C_{n,m} \prod_{1 \leq i < j \leq n} (\lambda_i - \lambda_j)^2 \prod_{k=1}^{n} \lambda_k^{m-n} e^{-\lambda_k} 1_{E_n}(\lambda) \quad (3.27)$$

where

$$C_{n,m} = \frac{1}{\prod_{j=1}^{n} \Gamma(j+1) \Gamma(m-n+j)}.$$  \hspace{1cm} (3.28)

The probability density (3.27) is known in random matrix theory as Laguerre unitary ensemble. One can immediately see the close resemblance between (3.27) and (3.25). It can be actually shown that (see Section B.1) the eigenvalues of the normalised matrix $AA^\dagger / \text{Tr}(AA^\dagger)$ are distributed according to (3.25).

To understand euristically the connection between the Schmidt coefficients of a random bipartite state and the eigenvalues of Wishart matrices, note that if we fix some separable basis $\{|e_j\rangle \otimes |f_k\rangle\}$ of $\mathcal{H}_A \otimes \mathcal{H}_B$, one can
write the pure bipartite state $|\psi\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$ as

$$|\psi\rangle = \sum_{j=1}^{n} \sum_{k=1}^{m} A_{jk} |e_j\rangle \otimes |f_k\rangle,$$

(3.29)

with the components arranged in the rectangular $n \times m$ matrix $A$. The reduced density matrix on the subsystem $\mathcal{H}_A$ is then given by:

$$\psi_A = \text{tr}_B \psi = \sum_{j=1}^{n} \sum_{k=1}^{m} (AA^\dagger)_{jl} |e_j\rangle \langle e_l|.$$

(3.30)

Now note that if the matrix $A$ is chosen with complex i.i.d. standard entries, the resulting distribution of the (unnormalized) vector $|\psi\rangle$ has the probability density function

$$f_{|\psi\rangle}(A) = \frac{1}{(2\pi)^{n/2}} e^{-\frac{1}{2} \text{Tr}(AA^\dagger)}$$

(3.31)

with respect to the coordinates $A_{jk}$ in the fixed basis $\{|e_j\rangle \otimes |f_k\rangle\}$, which is a unitarily invariant distribution (see also Section 2.3). We see then that Wishart random matrices $W = AA^\dagger$ naturally arise in the description of the reduced state of a rotationally invariant random pure bipartite state $|\psi\rangle$.

This simple connection between Wishart matrices and random density matrices has been exploited extensively in the literature [25, 35, 39, 38, 52, 50, 53, 86, 88, 89, 102, 111, 4, 124, 144, 143, 157, 158, 159, 169, 170, 112], and provides a direct route to sample the local spectrum $\lambda(\psi_A)$ of a uniformly distributed pure state $|\psi\rangle \in \mathbb{C}^n \otimes \mathbb{C}^m$. The method proceeds as follows:

1. Sample a $n \times m$ random matrix $A$ with independent complex Gaussian entries;

2. Compute the eigenvalues of the $n \times n$ random matrix $AA^\dagger/ \text{Tr}(AA^\dagger)$.

This method is simple and intuitive, but not the most efficient numerically. We will see in the following section a beautiful result which is interesting both from a theoretical point of view and for its numerical applications [48].

### 3.3.1 Tridiagonal realization of Wishart Random Matrices

It is possible to obtain the spectrum of a Wishart random matrix without sampling an actual Wishart random matrix, but using a tridiagonal realization due to Dumitriu and Edelman [45].

**Theorem 10.** Let $A$ be the $n \times n$ random bidiagonal matrix

$$A = \begin{pmatrix}
\chi_{2m} & \chi_{2m-2} & \cdots & \chi_{2m-2(n-1)} \\
\chi_{2(n-1)} & \chi_{2(n-3)} & \cdots & \\
\cdots & \cdots & \cdots & \\
\chi_2 & \chi_{2m-2(n-1)}
\end{pmatrix},$$

(3.32)
where the $\chi'_\nu$'s are independent chi-distributed (real) random variables with $\nu$ degrees of freedom. Then, the eigenvalues of $AA^T/2$ are distributed according to the Laguerre unitary ensemble (3.27).

**Remark 1.** There is a slight difference with respect to [45] in the theorem above, precisely the normalization factor $1/2$ multiplying $AA^T$, which is absent in [45]. This difference is due to the fact that the authors in [45] actually use Wishart matrices defined with a different variance with respect to Definition 6.

The theorem above suggests an alternative procedure to sample $\lambda(\psi_A)$:

1. Sample a $n \times n$ bidiagonal real random matrix $A$ as in Theorem 10;
2. Compute the eigenvalues of the tridiagonal matrix $AA^T / \text{Tr}(AA^T)$.

In order to understand to which extent this procedure is more efficient than the ordinary one, let us briefly analyze the computational time and storage resources needed with each method.

**Ordinary method.** To sample $A$ we need to generate and store $nm = cn^2$ complex numbers (this is the number of complex coordinates of a vector in $\mathbb{C}^n \otimes \mathbb{C}^m$). The matrix multiplication $AA^\dagger$ using a standard algorithm requires $O(mn^2) = O(cn^3)$ complex operations, and the diagonalization of $AA^\dagger / \text{Tr}(AA^\dagger)$ has a computational cost of order $O(n^3)$ in the field $\mathbb{C}$.

**Tridiagonal method.** We just need to generate a $n \times n$ bidiagonal matrix $A$ with independent (but not identically distributed) entries as in Theorem 10. Note that $A$ is sparse, and only $O(n)$ real numbers need to be stored. Moreover, the diagonalization of the real tridiagonal matrix $AA^T / \text{Tr}(AA^T)$ requires $O(n^2)$ operations in the real field instead of $O(n^3)$ complex operations. Crucially, the value $m = cn$ enters only as a parameter of the distributions of the entries. Therefore, the computational time does not scale with $c$.

The numerical advantage of the tridiagonal construction is twofold. First, the storage and computational costs scale as $O(n)$ and $O(n^2)$ respectively. Second, the dimension $m$ of $\mathcal{H}_B$ enters only as a parameter of the matrix model; this provides a very effective route to sample local spectra of bipartite states from highly unbalanced bipartitions ($m$ much larger than $n$).

### 3.4 Spectrum of Wishart Random Matrices

A comprehensive reference on the subject is the paper by Nechita [114] (that, among other things, contains a collection of exact formulae on moments and precise asymptotic theorems on the largest eigenvalue of random density matrices).
3.4.1 Marchenko-Pastur Distribution

When studying large $n \times m$ random matrices $W$, an important question is to determine their spectrum in the limit $n \to \infty$. In order to answer this question, one formally introduces the empirical spectral measure

$$\mu_W = \frac{1}{n} \sum_{i=1}^{n} \delta_{\lambda_i}, \quad (3.33)$$

where $\delta_x$ denotes the delta measure concentrated on the point $x$, i.e. $\delta_x(A) = 1$ if $x \in A$ and zero otherwise, while $\lambda_i$ are the eigenvalues of the random matrix $W$. Since the eigenvalues $\lambda_i$ are random quantities, the measure (3.33) is a random probability measure on $\mathbb{R}$. One of the striking results of random matrix theory is that for many ensemble of random matrices the associated empirical spectral density converges to a non-random probability measure on the space where the eigenvalues are supported as $n \to \infty$. This is also the case for Wishart matrices, after a suitable rescaling.

**Theorem 11 ([114])**. Let $(W_n)$ be a sequence of Wishart random matrices with parameters $n$ and $m_n$, where $(m_n)$ is a sequence of integers such that $m_n/n \to c > 0$ as $n \to \infty$. Let $\{\lambda_i(W_n)\}$ be the eigenvalues of $W_n$ and let

$$\mu_n = \frac{1}{n} \sum_{i=1}^{n} \delta_{\lambda_i(W_n)} \quad (3.34)$$

denote the empirical spectral measure of its rescaled spectrum. Then $\mu_n \to \mu_c$ almost surely as $n \to \infty$, with

$$\mu_c = \max\{1 - c, 0\} \delta_0 + \frac{\sqrt{(x-a)(b-x)}}{2\pi x} 1_{[a,b]}(x)dx \quad (3.35)$$

where $a = (\sqrt{c} - 1)^2$ and $b = (\sqrt{c} + 1)^2$. The probability distribution $\mu_c$ is called Marchenko-Pastur distribution.

The almost sure convergence in the previous theorem is remarkable, since it practically means that for sufficiently large $n$ the histogram of the rescaled eigenvalues $\lambda_i(W_n)/n$ of a single Wishart matrix $W_n$ will resemble the Marchenko-Pastur distribution (3.35) (see Figure 3.3).

3.4.2 Largest eigenvalue and Tracy-Widom distribution

Another object of interest in random matrix theory is the largest eigenvalue of a large random matrix. For Wishart matrices, the following theorem (see [114, 82, 83]) shows that the largest eigenvalue converges almost surely to the right edge of the support of the Marchenko-Pastur distribution, and also provides some information on the nature of its fluctuations, which are described by the so-called Tracy Widom law [152, 153].
Definition 7. The Tracy-Widom law of order 2 has a distribution defined by

\[
F_2(x) = \exp \left( - \int_x^\infty (y-x)q^2(y)dy \right),
\]

where \( q(x) \) is the solution of a Painlevé equation of type II:

\[
q''(x) = xq(x) + 2q^3(x),
\]

with boundary condition

\[
q(x) \sim \text{Ai}(x) \quad \text{as} \ x \to \infty
\]

and \( \text{Ai}(x) \) is the Airy function.

Theorem 12. Let \( (m_n) \) be a sequence of integers such that \( m_n/n \to c > 0 \) as \( n \to \infty \). Let \( (W_n) \) a sequence of Wishart random matrices with parameters \( n \) and \( m_n \) and let \( \lambda_{\max}(W_n) \) be the largest eigenvalue of \( W_n \). Then, almost surely,

\[
\lim_{n \to \infty} \frac{\lambda_{\max}(W_n)}{n} = b,
\]
FIGURE 3.4: Histogram of the rescaled maximum eigenvalue \( \tilde{\lambda}_{\text{max}} = (\lambda_{\text{max}}(W_n) - nb)/\sigma_n \), with notation of Theorem 12, obtained from \( s = 10000 \) samplings of the largest eigenvalue \( \lambda_{\text{max}}(W_n) \) of Wishart matrices with \( n = 500 \) and \( c = 2 \). The orange curve represents the Tracy-Widom distribution, computed numerically according to Definition 7.

where \( b = (\sqrt{c} + 1)^2 \). Moreover, defining \( \sigma_n = n^{1/3}(1 + \sqrt{c})(1 + \frac{1}{\sqrt{c}})^{1/3} \), the following limit holds:

\[
\lim_{n \to \infty} \mathbb{P} \left( \frac{\lambda_{\text{max}}(W_n) - nb}{\sigma_n} \leq x \right) = F_2(x), \tag{3.40}
\]

where \( F_2(x) \) denotes the Tracy-Widom distribution.

Even if the Tracy Widom distribution is quite complicated, results such as (3.40) are useful to characterize the limit behaviour of the fluctuations of some random quantities in the large dimensional limit \( n \to \infty \). Moreover, the interesting fact about (3.40) is that it describes a universal behaviour, valid for several values of \( c > 0 \), upon rescaling. We will see later that such universal scaling laws can be found for quantities with operational meaning from the resource theoretic point of view.

### 3.4.3 Smallest eigenvalue of fixed-trace Wishart matrices

We collect here some information on the statistics of the smallest eigenvalue of Wishart matrices which will be needed later to analyse some properties of entanglement conversion in bipartite systems. The distribution and moments of the smallest eigenvalue \( \lambda_{\text{min}} \) of complex fixed-trace Wishart matrices can be given in terms of elementary functions in the case \( m = n \) [102]. Let \( (\lambda_1, \ldots, \lambda_n) \) be distributed according to \( g_{n,n} \) in (3.25). Then, \( \lambda_{\text{min}} \) has probability
density
\[ f_{\text{min}}(x) = n(n^2 - 1)(1 - nx)^{n^2 - 2}\theta(1/n - x), \quad (3.41) \]
where \( \theta(y) \) denotes the Heaviside step function, i.e. \( \theta(y) = 1 \) for \( y \geq 0 \) and zero otherwise. The \( k \)-th moment is given by
\[ E[(\lambda_n^k)] = \frac{\Gamma(n^2)\Gamma(k + 1)}{n^k \Gamma(n^2 + k)}. \quad (3.42) \]

In particular,
\[ E[\lambda_n^1] = \frac{1}{n^2}, \quad \text{and} \quad \text{Var}[\lambda_n^1] = \frac{1}{n^6} \left( \frac{n^2 - 1}{n^2 + 1} \right). \quad (3.43) \]

From the explicit formula (3.41) it is easy to see that the rescaled variable \( \lambda_n^1/n^3 \) converges in distribution to an exponential variable with rate 1:
\[ \lim_{n \to \infty} P(\lambda_n^1 \geq xn^3) = e^{-x}. \quad (3.44) \]

This is consistent with earlier results by Edelman [47, Theorem 3.2] on Wishart matrices (without fixed trace). The distribution of \( \lambda_n^1 \) for \( m > n \) is more complicated, and the interested reader might want to check [3, 87]. However, the asymptotic formulae which are useful for our purposes may be also inferred from known results on Wishart matrices without fixed trace. Let \((x_1, \ldots, x_n)\) be distributed according to \( G_{n,m} \) in (3.27), where \( m = cn \) with \( c > 1 \). Set \( a = (1 - \sqrt{c})^2 \) and \( s = |1 - \sqrt{c}|^{4/3}/c^{1/6} \). Then, the rescaled variable \( s^{-1}n^{-1/3}(an - x_n^1) \) converges in distribution to a \( \beta = 2 \) Tracy-Widom random variable:
\[ \lim_{n \to \infty} P \left( \frac{an - x_n^1}{sn^{1/3}} \leq x \right) = F_2(x), \quad (3.45) \]
(See [126] for more details). On the other hand \( \sum_{k=1}^n x_k/cn^2 \) converges in probability to 1 by the law of large numbers. Hence, using the representation
\[ \lambda_j = \frac{x_j}{\sum_{k=1}^n x_k} \quad (3.46) \]
(see (B.1) and the appendix for further details), equation (3.45) can be written as:
\[ \lim_{n \to \infty} P \left( \frac{\lambda_n^1 - a}{2c n^{-5/3}} \leq x \right) = F_2(x), \quad (3.47) \]

hence we see that \( \lambda_n^1 \) has mean \( O(n^{-1}) \) and fluctuation \( O(n^{-5/3}) \) in the asymptotic regime \( n \to \infty \).
Part II

Random states in Quantum Resource Theories
Chapter 4

Quantum Resource Theories

This chapter provides an introduction on the formalism of quantum resource theories, which provide a natural framework to investigate the usefulness of quantum resources such as entanglement or quantum coherence. Section 4.1 gives a brief recap on the formal description of quantum operations. This is preparatory for the description of the general structure of Quantum Resource Theories, which is given in Section 4.2. Then, in Section 4.3 we turn to the central problem of conversion between states within the restrictions imposed by the resource theory. Finally, in Section 4.4 the mathematical relation of Majorization is introduced, which allows to characterize state conversions in several resource theories, and in particular in those which will be investigated here in the following chapters.

4.1 Quantum Operations

The most general transformations of the state of a quantum system are called Quantum Operations [118]. As such, they must include all the operations which make sense from a physical point of view. Formally, a quantum operation is a map

$$\Lambda : \mathcal{S}(\mathcal{H}_1) \to \mathcal{S}(\mathcal{H}_2)$$

(4.1)

which sends quantum states acting on some Hilbert space $\mathcal{H}_1$ into quantum states acting on a possibly different Hilbert space $\mathcal{H}_2$, since we want to take into account the possibility of adding an ancilla system or discarding a subsystem. In order to map quantum states into quantum states, quantum operation must preserve the positivity and the trace, i.e. $\Lambda(\rho) > 0$ whenever $\rho > 0$ and $\text{tr}(\Lambda(\rho)) = \text{tr}(\rho)$. Furthermore, they must satisfy an additional less trivial requirement called complete positivity.

**Definition 8 (Complete positivity).** A map $\Lambda : \mathcal{S}(\mathcal{H}_1) \to \mathcal{S}(\mathcal{H}_2)$ is said to be $k$-positive if the map $\Lambda_k \equiv I_k \otimes \Lambda : \mathcal{S}(\mathbb{C}^k \otimes \mathcal{H}_1) \to \mathcal{S}(\mathbb{C}^k \otimes \mathcal{H}_2)$ is positive, where $I_k$ denotes the identity map on $\mathcal{S}(\mathbb{C}^k)$. The map $\Lambda$ is said to be completely positive if it is $k$-positive for all $k \in \mathbb{N}$.

The physical intuition behind this requirement can be understood recalling that we are including the possibility of considering ancilla systems: this requirement then guarantees the positivity of the output global state when performing the operation $\Lambda$ on the system leaving an arbitrary large ancilla unchanged. Positivity is a trivial consequence of complete positivity, and it
is known that if \( k = \dim(\mathcal{H}_1) \), \( k \)-positivity is enough to guarantee complete positivity. Collecting the requirements together we can give the following formal definition.

**Definition 9 (Quantum Operation).** A quantum operation is a completely positive trace preserving linear map (CPTP).

Quantum operations have three important representations which are frequently used, each one having its own advantage. We will recap them here for the sake of completeness.

### 4.1.1 Stinespring Factorization

The most physically intuitive representation of quantum operations is to regard them as arising from the interaction of the system with an environment, and the generic quantum operation \( \Lambda \) is decomposed as:

\[
\Lambda(\rho) = \text{tr}_{\mathcal{E}}(U(\rho \otimes \sigma)U^\dagger).
\]

(4.2)

In this picture, the state \( \sigma \in S(\mathcal{E}) \) of the environment \( \mathcal{E} \) is included in the description, and the joint product state of the closed total system made up of the system and the environment evolves unitarily. Finally, the environment part of the state is discarded. For physical operations involving initial and final states pertaining to different systems, the partial trace may be taken over an environment \( \mathcal{E}' \) different from the starting one. It is a standard exercise to verify that a map which can be written as (4.2) is indeed a Quantum Operation, according to Definition 9. The converse is also true, thanks to an important result known as Stinespring Dilation Theorem [147].

### 4.1.2 Operator-Sum representation

The operator-sum representation, or Kraus representation, is less intuitive than the Stinespring factorization but is more useful for our purposes. A simple calculation allows to derive the Kraus representation starting from (4.2). Denoting with \( \{ |e_\alpha\rangle \}_\alpha \) an orthonormal basis of the environment \( \mathcal{E} \), we can assume that the initial state of the environment is pure, \( \sigma = |0\rangle\langle 0| \). There is no loss of generality in this assumption since it is always possible to purify a non-pure state by enlarging the environment. Then, equation (4.2) can be written as

\[
\Lambda(\rho) = \sum_\alpha \langle e_\alpha | U(\rho \otimes |e_0\rangle\langle e_0|)U^\dagger | e_\alpha \rangle = \sum_\alpha K_\alpha \rho K_\alpha^\dagger,
\]

(4.3)

where \( K_\alpha = \langle e_\alpha | U | e_0\rangle \) are known as Kraus operators. From the unitarity of \( U \), we have that

\[
\sum_\alpha K_\alpha^\dagger K_\alpha = \langle e_0 | U^\dagger \sum_\alpha |e_\alpha\rangle\langle e_\alpha| U | e_0\rangle = \langle e_0 | \mathbb{1}_H \otimes \mathbb{1}_\mathcal{E} | e_0\rangle = \mathbb{1}_H
\]

(4.4)
which is called completeness relation and implies the trace preserving condition. It is possible to prove that any quantum operation (according to Definition 9) can be expressed in terms of an operator sum representation.

**Theorem 13.** Let $\Lambda : S(\mathcal{H}_1) \to S(\mathcal{H}_2)$ be a CPTP map. Then there exists a (non-unique) set of Kraus operators, $K_\alpha : \mathcal{H}_1 \to \mathcal{H}_2$, such that

$$\Lambda(\rho) = \sum_\alpha K_\alpha \rho K_\alpha^\dagger,$$

with

$$\sum_\alpha K_\alpha^\dagger K_\alpha = I_{\mathcal{H}_1} \quad (4.6)$$

A proof of this theorem can be found in [118]. The authors of [118] actually prove the version of the theorem for maps which are only completely positive and trace non-increasing. In that case equation (4.6) becomes

$$\sum_\alpha K_\alpha^\dagger K_\alpha \leq I_{\mathcal{H}_1}, \quad (4.7)$$

but reduces to (4.6) when the trace preservation is required.

### 4.2 General structure of Quantum Resource Theories

A resource theory is defined by the allowed (or free) operations and by the free states, i.e. states which can be prepared within the restrictions imposed on the allowed operations. In a meaningful quantum resource theory, one needs to define free states and allowed operations consistently: any state which can be prepared with an allowed operation must be considered free, and it should not be possible to use free states alone to perform quantum operations which are not allowed. This considerations motivate the following formal definition [30].

**Definition 10.** Let $A$ be a mapping that assigns any two input/output Hilbert spaces $\mathcal{H}_1$ and $\mathcal{H}_2$ to a unique set of CPTP operations $A(\mathcal{H}_1; \mathcal{H}_2) \subset \mathcal{L}(\mathcal{H}_1, \mathcal{H}_2)$. Let $F$ be the induced mapping $F(\mathcal{H}) := A(C; \mathcal{H})$ for an arbitrary Hilbert space $\mathcal{H}$. Then the pair $(F, A)$ is called a quantum resource theory (QRT) if the following two conditions hold:

1. For any Hilbert space $\mathcal{H}$, the set $A(\mathcal{H}) := A(\mathcal{H}; \mathcal{H})$ contains the identity map $I$.

2. For any three Hilbert spaces $\mathcal{H}_1, \mathcal{H}_2$ and $\mathcal{H}_3$, if $\Lambda \in A(\mathcal{H}_1, \mathcal{H}_2)$ and $\Phi \in A(\mathcal{H}_2; \mathcal{H}_3)$, then $\Phi \circ \Lambda \in A(\mathcal{H}_1, \mathcal{H}_3)$.

In a QRT, $F(\mathcal{H}) \subset S(\mathcal{H})$ defines the free states on $S(\mathcal{H})$, i.e. the ones which can be obtained within the restrictions imposed on the allowed operations, while states in $S(\mathcal{H}) \setminus F(\mathcal{H})$ are called resource states, or static resources.
Likewise, quantum operations in $\mathcal{A}(\mathcal{H}_1, \mathcal{H}_2)$ are called free or allowed operations, while the other Quantum Operations are called dynamical resources. The first condition of Definition 10 is an obvious requirement for any meaningful QRT: doing nothing is an allowed operation. Condition 2 guarantees that a concatenation of free operations is still considered as a free operation. Moreover, condition 2 together with the definition of the free states as $\mathcal{F}(\mathcal{H}) := \mathcal{A}(\mathcal{C}; \mathcal{H})$ implies that allowed operations cannot be used to create resource states starting from free states. This should be considered as the golden rule of Quantum Resource Theories [30].

Definition 10 establishes only the minimal requirements which any meaningful QRT should satisfy. Most physically motivated and previously studied QRTs satisfy also some additional requirements, which can be collected together as a tensor product structure [30].

**Definition 11.** A QRT $(\mathcal{F}, \mathcal{A})$ is said to admit a tensor product structure if the following three conditions hold:

1. Free operations are “completely free”: for any pair of input/output systems $\mathcal{H}_1, \mathcal{H}_2$ and any ancilla system $\mathcal{H}_A$, if $\Lambda \in \mathcal{A}(\mathcal{H}_1, \mathcal{H}_2)$, then $I_A \otimes \Lambda \in \mathcal{A}(\mathcal{H}_A \otimes \mathcal{H}_1, \mathcal{H}_A \otimes \mathcal{H}_2)$, where $I_A$ is the identity map on $\mathcal{S}(\mathcal{H}_A)$.

2. Appending free states is a free operation: for any given free state $\sigma \in \mathcal{F}(\mathcal{H}_2)$, the CPTP map $\Lambda(\rho) := \rho \otimes \sigma$ belongs to $\mathcal{A}(\mathcal{H}_1, \mathcal{H}_1 \otimes \mathcal{H}_2)$.

3. Discarding a system is allowed: for any Hilbert space $\mathcal{H}$, the set $\mathcal{A}(\mathcal{H}, \mathbb{R})$ is not empty (i.e. contains the trace operation).

These additional requirements are very natural in many physical models. All the QRTs treated in this thesis admit a tensor product structure. From a conceptual point of view, the identification of free states as $\mathcal{F}(\mathcal{H}) := \mathcal{A}(\mathcal{C}; \mathcal{H})$ made in Definition 10 places the allowed operations on a slightly more fundamental level than the free states [30], while in standard literature they are usually treated on an equal footing. From a practical point of view, when one wants to model some physical constraints in some QRT, it might be more natural to start either from the specification of free states or from defining the allowed operations, depending on the particular scenario considered. For example, in the entanglement resource theory, where there are some spatially separated parties, it appears more natural to first specify the allowed operations to be local operations and classical communication (LOCC). Conversely, in the resource theory of coherence, the physical constraints are more naturally implemented by first specifying the free states and then turning to the allowed operations. Independently of the starting point, one needs to specify the free states $\mathcal{F}$ and the allowed operations consistently $\mathcal{A}$ consistently, in order to satisfy the golden rule of QRTs [30].

### 4.2.1 Consistent QRTs for a given set of free operations

Suppose the free operations are to be fixed by the physical constraints. For a given Hilbert space $\mathcal{H}$ the allowed operations $\mathcal{A}(\mathcal{H})$ impose a pre-order
(i.e. a reflexive and transitive relation) on the set of quantum states $S(\mathcal{H})$: for any $\rho, \sigma \in S(\mathcal{H})$, we write
\[
\rho \xrightarrow{\Lambda} \sigma
\]
if there exists a free operation $\Lambda \in \mathcal{A}(\mathcal{H})$ such that $\sigma = \Lambda(\rho)$. Reflexivity of the relation $\xrightarrow{\Lambda}$ is guaranteed by the fact that the identity map is always allowed, while transitivity is guaranteed by the fact that the composition of free operations is a free operation. If both $\rho \xrightarrow{\Lambda} \sigma$ and $\sigma \xrightarrow{\Lambda} \rho$, we write $\rho \simeq \sigma$. We can then associate with a given set of allowed operations $\mathcal{A}(\mathcal{H})$, the minimal set of free states, defined as
\[
\mathcal{F}_{\text{min}}(\mathcal{H}) := \{ \rho : \sigma \xrightarrow{\Lambda} \rho \quad \forall \sigma \in S(\mathcal{H}) \} \tag{4.9}
\]
That this is indeed the minimal set of free states is proved in the following proposition.

**Proposition 7.** Let $\mathcal{F}(\mathcal{H})$ be a specification of free states consistent with $\mathcal{A}(\mathcal{H})$. Then $\mathcal{F}_{\text{min}}(\mathcal{H}) \subset \mathcal{F}(\mathcal{H})$. Moreover, if $\rho \simeq \sigma$ for any $\rho, \sigma \in \mathcal{F}(\mathcal{H})$, then $\mathcal{F}(\mathcal{H}) = \mathcal{F}_{\text{min}}(\mathcal{H})$.

**Proof.** Let $\rho \in \mathcal{F}_{\text{min}}$. Then, by definition, for every $\sigma \in S(\mathcal{H})$, and in particular for $\sigma \in \mathcal{F}(\mathcal{H})$, we have $\sigma \xrightarrow{\Lambda} \rho$. But any state which can be obtained from $\mathcal{F}(\mathcal{H}) \subset \mathcal{A}(\mathbb{C}; \mathcal{H})$ by an allowed operation must be itself in $\mathcal{F}(\mathcal{H})$, due to the second requirement of Definition 10, and hence $\rho \in \mathcal{F}(\mathcal{H})$, proving $\mathcal{F}_{\text{min}}(\mathcal{H}) \subset \mathcal{F}(\mathcal{H})$. To prove the second part, we show that if $\mathcal{F}(\mathcal{H})$ is such that $\rho \simeq \sigma$ for any $\rho, \sigma \in \mathcal{F}(\mathcal{H})$, then any $\rho \in \mathcal{F}(\mathcal{H})$ is attainable starting from any $\sigma \in S(\mathcal{H})$ via allowed operations, which implies that $\mathcal{F}_{\text{min}}(\mathcal{H}) = \mathcal{F}(\mathcal{H})$; this follows from the fact that $\sigma \xrightarrow{\Lambda} \omega$ for any $\omega \in \mathcal{F}_{\text{min}}(\mathcal{H}) \subset \mathcal{F}(\mathcal{H})$, which implies that $\omega \simeq \rho$ and in particular $\omega \xrightarrow{\Lambda} \rho$; then, by transitivity, $\sigma \xrightarrow{\Lambda} \rho$. \qed

As a consequence of Proposition 7, if one is interested in QRTs where all free states are interconvertible into each other, the set of free states is uniquely determined by the allowed operations. In all resource theories which admit a tensor product structure free states can be freely converted into each other (more generally, in all resource theories which allow to prepare and discard any free state).

### 4.2.2 Consistent QRTs for a given set of free states

Certain physical situations are more easily modeled as a resource theory by starting from the definition of the free states. This is the case of the resource theory of coherence, which will be studied in Chapter 6. At variance with what happens if one looks at consistent QRTs with a fixed set of allowed operations, fixing only the free states does not identify a unique set of allowed operations. In some cases one might want to choose some set of consistent allowed operations because it is easier to characterize mathematically, in other situations the choice can be dictated by specific physical considerations. A
thorough analysis of the consistent choices of free operations corresponding to a given set of free states can be found in [30]. Here we will mention only some relevant classes.

For a given specification of the free states, it is always possible to find a maximal consistent specification of free operations. This is the set of resource non-generating operations.

**Definition 12 (Resource non-generating (RNG) operations).** Let \( F \) be a mapping which assigns to any Hilbert space \( \mathcal{H} \) a subset \( F(\mathcal{H}) \subset S(\mathcal{H}) \) of the states on \( \mathcal{H} \). For any two Hilbert spaces \( \mathcal{H}_1, \mathcal{H}_2 \), the set of RNG operations \( A_{\text{max}}(\mathcal{H}_1, \mathcal{H}_2) \) consists of all quantum operations \( \Lambda \in \mathcal{L}(\mathcal{H}_1, \mathcal{H}_2) \) having the property that if \( \rho \in F(\mathcal{H}_1) \), then \( \Lambda(\rho) \in F(\mathcal{H}_2) \).

Since in every QRT a resource state cannot be generated from free states, it is obvious that every other set of free operations \( A \) which is consistent with the choice of free states must satisfy \( A \subset A_{\text{max}} \). As one might expect, if one consider the maximal set of free operations \( A_{\text{max}}(\mathcal{H}) \) on some system \( \mathcal{H} \) which are consistent with \( F(\mathcal{H}) \), then \( F(\mathcal{H}) \) will be the minimal set of free states compatible with these free operations. This is true since by definition the operations in \( A_{\text{max}}(\mathcal{H}) \) can transform every state of \( F(\mathcal{H}_1) \) into any other state in \( F(\mathcal{H}) \), and hence by Proposition 7 this set must be \( F_{\text{min}}(\mathcal{H}) \).

With the same intuition which is behind the definition of \( k \)-positive maps and completely positive maps, one can restrict the set of RNG operations to \( k \)-RNG and completely RNG operations.

**Definition 13 (\( k \)-RNG and CRNG operations).** Let \( F \) be a mapping which assigns to any Hilbert space \( \mathcal{H} \) a subset \( F(\mathcal{H}) \subset S(\mathcal{H}) \) of the states on \( \mathcal{H} \). A map \( \Lambda \in A_{\text{max}}(\mathcal{H}_1, \mathcal{H}_2) \) is \( k \)-resource non-generating (\( k \)-RNG) if
\[
I_k \otimes \Lambda \in A_{\text{max}}(C^k \otimes \mathcal{H}_1, C^k \otimes \mathcal{H}_2)
\]
Moreover, \( \Lambda \) is said to be a completely resource non-generating (CRNG) operation if it is \( k \)-RNG for all \( k \in \mathbb{N} \).

**Remark 2.** Notice that if all states are free, \( F(\mathcal{H}) = S(\mathcal{H}) \), then \( k \)-RNG and CRNG correspond respectively to \( k \)-positive and completely positive trace preserving maps respectively.

Similar to \( k \)-positivity, it is known that if a quantum operation is \( k \)-RNG with \( k > \dim \mathcal{H} = d \), then it is CRNG. We then have the following inclusion relations:
\[
\text{RNG} = 1\text{-RNG} \supset 2\text{-RNG} \supset \cdots \supset d\text{-RNG} = \text{CRNG}
\]

### 4.2.3 Resource Quantification

An operationally meaningful quantification of resources is one of the most useful aspects of quantum resource theories. We will briefly introduce here the resource measures, summarizing briefly the main axiomatic requirements necessary to obtain meaningful quantities.
4.2. General structure of Quantum Resource Theories

Formally, a resource measure is some function assigning to any given state \( \rho \in S(\mathcal{H}) \) some positive quantity \( M(\rho) \) quantifying how much resourceful \( \rho \) is. Then, in order to be meaningful, it should satisfy the following requirements:

- **Vanishing for free states.** We should require
  \[
  \rho \in F(\mathcal{H}) \to M(\rho) = 0, \tag{4.12}
  \]
  i.e. free states have zero resource content. One could be tempted to require also the converse of (4.12), i.e. that \( M(\rho) > 0 \) for any resource state \( \rho \in S(\mathcal{H}) \setminus F(\mathcal{H}) \), a property which goes under the name of faithfulness. It turns out that some resource states do not provide operational advantages over free states when performing some particular task, hence a measure quantifying the usefulness of that state for the considered task should rightfully assign to it a zero measure. Thus, faithfulness is not a necessary requirement for a resource measure.

- **Monotonicity.** Since free operations are supposed to generate no resources, a suitable resource measure should be non-increasing under the allowed operations, i.e.
  \[
  M(\Lambda(\rho)) \leq M(\rho), \tag{4.13}
  \]
  for any \( \rho \in S(\mathcal{H}) \) and for any \( \Lambda \in A(\mathcal{H}) \). Note that in resource theories with a tensor product structure, where all free states are interconvertible between each other, this property implies that \( M(\rho) = M(\sigma) \) for any pair of free states \( \rho, \sigma \in F(\mathcal{H}) \). In fact, denoting with \( \Lambda, \Lambda' \) the allowed operations such that \( \Lambda(\rho) = \sigma \) and \( \Lambda'(\sigma) = \rho \), we can write \( M(\sigma) = M(\Lambda(\rho)) \leq M(\sigma) \) and \( M(\rho) = M(\Lambda'(\sigma)) \leq M(\sigma) \), which implies that \( M(\rho) = M(\sigma) \). Then, for these resource theories, once we have found a function satisfying monotonicity it is simple to obtain a function which vanishes for all free states by considering simply a shift so that \( M(\rho) = 0 \) for all free states at once.

- **Convexity.** A resource measure is said to be convex if
  \[
  M\left( \sum_{i=1}^{n} p_i \rho_i \right) \leq \sum_{i=1}^{n} p_i M(\rho_i) \tag{4.14}
  \]
  for any collection of states \( \rho_i \) with corresponding probability distribution \( p_i \). This requirement is usually justified by saying that mixing states should not increase the amount of resource.

Depending on the resource theory one may ask additional requirements to hold for a given resource measure. For a thorough discussion on the topic, see [30].
4.3 State conversions in Resource Theories

Conversion between resource states is the most studied problem in every quantum resource theory. State conversions are important because some resource states are more useful than others for achieving some given task. For example, in the entanglement resource theory a maximally entangled state allows to achieve quantum teleportation with maximal fidelity, while a less entangled state will yield less faithful teleportation protocols.

There are several kinds of state conversions which can be considered, depending on whether one is interested on the conversion of a single copy of a state into another, or if one is interested in what can be done in the asymptotic limit of many copies of the same states. Moreover, one can make a distinction between deterministic (or conclusive) conversions which can always be realized with some protocols, or stochastic conversion, which can be realized with protocols having a certain probability of failure. The presence of a catalyst can also be of some influence in a given state conversion. Here we will review all these possibilities.

4.3.1 Single-shot convertibility

In the single-shot regime one considers a state \( \rho \in S(\mathcal{H}_1) \) on a certain “input” space \( \mathcal{H} \) and a state \( \sigma \in S(\mathcal{H}_2) \) on an “output” space \( \mathcal{H}_2 \), and wants to characterize the possibility of achieving the state conversion \( \rho \rightarrow \sigma \) by using only the allowed operations. There are several possibilities which can be considered here, depending on the necessities of the task considered.

**Definition 14 (Deterministic Conversion).** Given a state \( \rho \in S(\mathcal{H}_1) \) and a state \( \sigma \in S(\mathcal{H}_2) \), we say that the state conversion \( \rho \rightarrow \sigma \) is achievable deterministically (or conclusively) by the allowed operations \( A \) and we will write

\[
\rho \xrightarrow{A} \sigma \quad (4.15)
\]

if there exists an allowed operation \( \Lambda \in A(\mathcal{H}_1, \mathcal{H}_2) \) such that \( \Lambda(\rho) = \sigma \).

If not otherwise specified, when we say that a state conversion is possible we will always mean that it is possible by a deterministic (or conclusive) conversion.

It can happen that a deterministic state conversion \( \rho \rightarrow \sigma \) is not possible, but it is possible to find a protocol involving some measurement which for some outcomes yield the correct state \( \sigma \) starting from \( \rho \). More explicitly, let \( \Lambda \in A(\mathcal{H}_1; \mathcal{H}_2) \) and consider a Kraus representation \( \{ K_\alpha \} \) of \( \Lambda \). When acting on \( \rho \), the output state can be written as a mixture:

\[
\Lambda(\rho) = \sum_\alpha K_\alpha \rho K_\alpha^\dagger = \sum_\alpha p_\alpha \rho_\alpha, \quad (4.16)
\]

where

\[
p_\alpha = \text{tr}(K_\alpha^\dagger K_\alpha \rho), \quad \rho_\alpha = K_\alpha \rho K_\alpha^\dagger / p_\alpha, \quad (4.17)
\]
4.3. State conversions in Resource Theories

represent respectively the probability $p_{\alpha}$ of obtaining the outcome $\alpha$, and the final state $\rho_{\alpha}$ if the outcome $\alpha$ was obtained. We define the probability of success associated to the protocol as:

$$p(\rho, \sigma; \Lambda) = \sup_{\{K_{\alpha}\}} \left\{ \sum_{\alpha: \rho_{\alpha} = \sigma} p_{\alpha} \right\},$$

(4.18)

where $p_{\alpha}$ and $\rho_{\alpha}$ are defined as in equation (4.17) and the supremum is taken over all the possible Kraus decompositions of $\Lambda$. The need to take the supremum stems from the fact that the Kraus decomposition of a quantum operation $\Lambda$ is not unique, and some Kraus decomposition might represent an implementation of $\Lambda$ which can be more useful for realizing the particular state conversion $\rho \rightarrow \sigma$ considered. By making this definition we are implicitly assuming that if the quantum operation $\Lambda(\cdot) = \sum_{\alpha} K_{\alpha}(\cdot)K_{\alpha}^\dagger$ is allowed, then also the “classical flagged” CPTP map $\Lambda_{QC} = \sum_{\alpha} K_{\alpha}(\cdot)K_{\alpha}^\dagger \otimes |\alpha\rangle\langle \alpha|$ and the measurements on the classical register $\{|\alpha\rangle\langle \alpha|\}$ are allowed. This is true for the resource theories of Entanglement and Quantum Coherence for which stochastic conversions will be considered, hence we will take this for granted. We can now give the following definition.

**Definition 15 (Stochastic conversions).** Let $\rho \in S(\mathcal{H}_1)$ and $\sigma \in S(\mathcal{H}_2)$. We say that a stochastic conversion $\rho \rightarrow \sigma$ is achievable by the allowed operations $\mathcal{A}$ if there exists an operation $\Lambda \in \mathcal{A}(\mathcal{H}_1, \mathcal{H}_2)$ such that $p(\rho, \sigma, \Lambda) > 0$. Moreover, we define the maximal success probability of the state conversion as

$$\Pi_{\mathcal{A}}(\rho \rightarrow \sigma) := \sup_{\Lambda \in \mathcal{A}(\mathcal{H}_1, \mathcal{H}_2)} p(\rho, \sigma, \Lambda)$$

(4.19)

We will omit the subscript $\mathcal{A}$ if no possibility of confusion arises.

Both Definitions 14 and 15 concern exact conversions, i.e. conversions in which the state achieved is exactly equal to the target state $\sigma$. It is frequent in physics to be satisfied with a result which is not perfect, but is affected only by a negligible error so that it works well for all practical purposes. It is natural then to define a notion of approximate conversions. In order to do so, however, we first need a way to quantify the distance between two quantum states, or stated otherwise, how well a given quantum state approximates another one.

**Definition 16 (Contractive metric).** Let $\mathcal{D}$ be a metric on the set of quantum states $S(\mathcal{H})$, i.e. a map that assigns to each pair of quantum states $\rho, \sigma \in S(\mathcal{H})$ a positive number $\mathcal{D}(\rho, \sigma)$ satisfying:

- $\mathcal{D}(\rho, \sigma) = \mathcal{D}(\sigma, \rho)$;
- $\mathcal{D}(\rho, \sigma) = 0 \iff \rho = \sigma$;
- $\mathcal{D}(\rho, \sigma) \leq \mathcal{D}(\rho, \omega) + \mathcal{D}(\omega, \sigma)$.

for every $\rho, \sigma, \omega \in S(\mathcal{H})$. Then $\mathcal{D}$ is said to be a contractive metric if for every quantum operation $\Lambda \in \mathcal{L}(\mathcal{H})$ and every pair of states $\rho, \sigma \in S(\mathcal{H})$ it holds that

$$\mathcal{D}(\Lambda(\rho), \Lambda(\sigma)) \leq \mathcal{D}(\rho, \sigma)$$

(4.20)
Now we can give the definition of approximate conversion:

**Definition 17 (Approximate conversion).** Let \( \rho \in S(\mathcal{H}_1) \), \( \sigma \in S(\mathcal{H}_2) \), \( \mathcal{D} \) a contractive metric on \( S(\mathcal{H}_2) \) and \( \varepsilon > 0 \). We say that \( \rho \) can be \( \varepsilon \)-converted in \( \sigma \) and we write

\[
\rho \xrightarrow{\varepsilon} \sigma
\]

if it can be converted into a state \( \Lambda(\rho) \) which is \( \varepsilon \)-close to \( \sigma \), i.e. if there exists a free operation \( \Lambda \in \mathcal{A}(\mathcal{H}_1, \mathcal{H}_2) \) such that \( \mathcal{D}(\Lambda(\rho), \sigma) \leq \varepsilon \).

### 4.3.2 Asymptotic convertibility

In the asymptotic regime we have a large number \( n \) of copies of a system prepared in some input state \( \rho \in S(\mathcal{H}_1) \) and we would like to determine how many copies of some target state \( \sigma \in S(\mathcal{H}_2) \) can be obtained by using the allowed operations. In this regime one does not demand the transformation to be perfect: an approximate result is allowed, as long as the approximation error goes to zero in the “thermodynamic limit” of infinite copies. In other words, one is interested in determining the maximal integer \( m_n \) such that

\[
\rho \otimes n \xrightarrow{\varepsilon_n} \sigma \otimes m_n
\]

with \( \varepsilon_n \to 0 \) as \( n \to \infty \). One of the central task of resource theories is that of determining the optimal value of the ratio \( m_n/n \) as \( n \to \infty \). More precisely, a rate \( R \) is said to be achievable in transforming \( \rho \) to \( \sigma \) if for every \( R' < R \) and every \( \varepsilon > 0 \) there exists a sufficiently large integer \( n \) such that

\[
\rho \otimes n \xrightarrow{\varepsilon} \sigma \otimes \lfloor R'n \rfloor
\]

where \( \lfloor \cdot \rfloor \) denotes the floor of a real number to an integer.

### 4.3.3 Catalytic convertibility

We are not going to address catalytic conversions in this thesis, but we mention them here briefly for the importance they have in quantum resource theories. Borrowing the term from chemistry, catalysis consists in the fact that certain state conversions are not achievable given only the initial state, but they become possible in presence of a system which acts as a catalyst, in the sense that its final state is equal to its initial state. More formally, we talk about a catalytic conversion when the state conversion \( \rho \to \sigma \) is not achievable but there exists some state \( \omega \) (the catalyst) such that

\[
\rho \otimes \omega \xrightarrow{A} \sigma \otimes \omega.
\]

Catalysis is a fascinating phenomenon whose relevance for entanglement conversions was discovered in [84] and turns out to be essential also for other resource theories in order to obtain physically relevant results [20]. For more information on the subject, see the reviews [67, 30] and references therein.
4.4 Majorization Theory

Majorization is a powerful tool which appears in many resource theories to characterize state conversions \[119, 67, 98\]. Good references on majorization theory are the standard books \[16, 103\]. Intuitively, majorization expresses the concept that a given probability distribution is more disordered than another, and for this reason it comes useful in a lot of different areas beyond physics (e.g. economics).

**Definition 18 (Majorization).** For two vectors \( x, y \in \mathbb{R}^n \), we say that \( x \) is majorized by \( y \), and write \( x \prec y \), if

\[
\sum_{j=1}^{k} x_j^k \leq \sum_{j=1}^{k} y_j^k, \quad \text{for} \quad k = 1, \ldots, n - 1, \quad \sum_{j=1}^{n} x_j^k = \sum_{j=1}^{n} y_j^k.
\]

where for a real vector \( x = (x_1, \ldots, x_n) \in \mathbb{R}^n \) we denote by \( x^k \) its permutation with components arranged in decreasing order, i.e. \( x_1^k \geq x_2^k \geq \cdots \geq x_n^k \).

Note that, if \( x, y \) are probability vectors, then the last equality is automatically satisfied by virtue of the normalization condition. The majorization relation is a preorder on the \((n - 1)\)-dimensional simplex of probability vectors

\[
\Delta_{n-1} = \left\{ x \in \mathbb{R}_+^n : \sum_{k=1}^{n} x_k = 1 \right\},
\]

i.e. it is reflexive and transitive. For \( n > 2 \), the ordering is only partial: there are vectors \( x, y \in \Delta_{n-1} \) such that neither \( x \prec y \) nor \( y \prec x \) (sometimes we say that \( x \) and \( y \) are not comparable). Moreover, there are a smallest and a largest element (up to permutations). Indeed, for every probability vector \( x \in \Delta_{n-1} \), we have the following relation:

\[
\left( \frac{1}{n}, \ldots, \frac{1}{n} \right) \prec x \prec (1, 0, \ldots, 0).
\]

The majorization condition can be characterised in terms of the action of bistochastic matrices, i.e. matrices with nonnegative entries and whose rows and columns sum to one. Equivalently, bistochastic matrices are matrices \( B \) mapping probability vectors into probability vectors and leaving the uniform vector \( (1/n, \ldots, 1/n) \) invariant.

**Theorem 14 (Hardy-Littlewood-Pólya).** Let \( x, y \in \mathbb{R}^n \). Then, \( x \prec y \) if and only if there exists a bistochastic matrix \( B \) such that \( x = By \).

This theorem allows us to have a more intuitive grasp on majorization relation in terms of mixing degree of probability vectors. In fact, it is trivial to show that permutations are bistochastic matrices and that mixture of bistochastic matrices are bistochastic as well. More generally, the set of bistochastic matrices can be characterised geometrically thanks to the following theorem.
Theorem 15 (Birkhoff’s Theorem). The set of $n \times n$ bistochastic matrices is a convex set whose extremal points are the permutation matrices.

Combining Theorem 14 and Theorem 15, we have that $x$ is majorized by $y$ if and only if $x$ is a convex combination of permutations of $y$, i.e:

$$x = \sum_{j=1}^{k} p_j P_j y,$$

(4.28)

with $p_j$ being some positive weights summing to one and $P_j$ being some permutations. We see then that the set of vectors majorized by some fixed $y \in \Delta_{n-1}$ is a convex set, and more precisely it is obtained by taking the convex hull of all the $n!$ permutations of $y$ (see Figure 4.1).

![Figure 4.1](image-url)

**Figure 4.1**: The red-shaded region is the set of points $x$ majorized by $y = (\frac{5}{10}, \frac{3}{10}, \frac{2}{10})$. It is a convex subset of $\Delta_2$ whose extremal points are the $3!$ permutations of $y$.

**Definition 19.** A bistochastic matrix $B$ is said to be unistochastic if there exists a unitary matrix $U$ such that $B_{ij} = |U_{ij}|^2$. In the special case that $U$ is an orthogonal matrix (with real elements), the matrix $B$ is said to be orthostochastic

Note that all unistochastic matrices are bistochastic, but the converse is not true, i.e. there are bistochastic matrices which are not unistochastic. Interestingly, it can be shown that bistochastic matrices involved in Theorem 14 can be always taken to be unistochastic. This is a consequence of the following theorem.
Theorem 16 (Schur-Horn Theorem). Let \( d, \lambda \in \mathbb{R}^n \). There is a Hermitian matrix with diagonal entries given by \( d \) and eigenvalues given by \( \lambda \) if and only if \( d \prec \lambda \).

To see that this implies that the matrix \( B \) in Theorem 14 can be always be taken to be unistochastic, simply note that if \( x \prec y \) then there is a Hermitian matrix \( A \) with diagonal entries \( A_{ii} = x_i \) and eigendecomposition \( A = U D U^\dagger \), where the diagonal matrix \( D \) contains the eigenvalues \( D_{ii} = y_i \), which implies that:

\[
x_i = A_{ii} = (UDU^\dagger)_{ii} = \sum_{j=1}^{n} |U_{ij}|^2 y_j,
\]

i.e. \( x = By \) with \( B_{ij} = |U_{ij}|^2 \).

4.4.1 Schur-convexity and majorization monotones

Majorization can be characterized in terms of functions which do not increase under the majorization order.

Definition 20 (Schur-convexity). A function \( f : \mathbb{R}^n \to \mathbb{R} \) is said to be Schur-convex if for any \( x, y \in \mathbb{R}^n \):

\[
x \prec y \implies f(x) \leq f(y).
\]

(4.30)

Schur-convexity is not to be confused with ordinary convexity, even though it is possible to construct Schur-convex functions \( f : \mathbb{R}^n \to \mathbb{R} \) starting from convex functions \( g : \mathbb{R} \to \mathbb{R} \).

Proposition 8 ([103]). Let \( g : \mathbb{R} \to \mathbb{R} \) be convex. Then, the function \( f : \mathbb{R}^n \to \mathbb{R} \) defined as

\[
f(x) = \sum_{i=1}^{n} g(x_i) \quad \forall x \in \mathbb{R}^n,
\]

(4.31)

is Schur-convex.

Using Proposition 8 and the fact that by definition the composition \( h \circ f \) of a Schur convex function \( f : \mathbb{R}^n \to \mathbb{R} \) with a monotone function \( h \) is still a Schur convex function, it is possible to construct several majorization monotones which are commonly used in quantum information theory.

For example considering \( g(t) = t^2 \) in Proposition 8 we obtain the purity:

\[
f(x) = \sum_{i=1}^{n} x_i^2,
\]

(4.32)

while the function \( g(t) = t \log t \), which is convex for \( t > 0 \), yields

\[
-H(x) = \sum_{i=1}^{n} x_i \log x_i,
\]

(4.33)
which is the negative of Shannon entropy. Then, Schur-convexity means that $x < y \Rightarrow H(x) \geq H(y)$, in accordance with the fact that Shannon entropy expresses the amount of “disorder” in a probability distribution. Finally, considering the family

$$g_{\alpha}(t) = \begin{cases} t^\alpha & \alpha < 0 \text{ or } \alpha > 1 \\ -t^\alpha & 0 < \alpha < 1 \end{cases}, \quad (4.34)$$

for all $\alpha \in \mathbb{R} \setminus \{0, 1\}$, which is a family of convex functions for $t > 0$, we get the Schur-convex functions $f_{\alpha}(x) = \sum_{i=1}^{n} g_{\alpha}(x_i)$, and considering the composition of these with the monotone functions

$$h_{\alpha}(t) = \frac{\text{sgn}(\alpha)}{\alpha - 1} \log(|t|) \quad (4.35)$$

we get the so-called $\alpha$-Renyi negentropies:

$$H_{\alpha}(x) := \frac{\text{sgn}(\alpha)}{\alpha - 1} \log \left( \sum_{i=1}^{n} x_i^\alpha \right), \quad (4.36)$$

which are widely used in quantum information theory.
Chapter 5

Random State Conversions in the Resource Theory of Entanglement

When considering systems composed of two parties $A$ and $B$ (or more), the physical constraint corresponds to the paradigm of “distant labs”: $A$ and $B$ can only perform local unitaries and measurements, and can freely communicate classical data, which includes the results of their local measurements. The set of quantum operations describing this scenario is the class of Local Operations assisted by Classical Communication (LOCC). The mathematical characterization of the set of LOCC operations is notoriously difficult [14, 41, 31]. However, when dealing with entanglement manipulations of pure bipartite systems, this description greatly simplifies. Since in this chapter we will consider only transformations among random pure bipartite states, we will omit the technical subtleties involved in the characterization of LOCC. Some of these technical details are collected in Appendix C for the sake of completeness.

The structure of the chapter is as follows. In Section 5.1 the problem under investigation is introduced along with the notation which will be used. In Section 5.2 deterministic conversions among pure states are discussed in detail, based on a criterion due to Nielsen [117] which characterizes these conversions entirely in terms of the majorization relation. A connection between the problem of conversions between random states and the positivity of a peculiar random walk is analyzed. Then, in Section 5.3 some numerical results are presented, based on sampling techniques discussed in Chapter 3. In Section 5.4 we turn to the characterization of maximal success probability in stochastic conversions. Numerical results are presented, along with an explicit computation for a particular case. Finally, in Section 5.5 some conclusions are drawn and possible outlooks are discussed.

5.1 Introduction and some notation

Let us begin by setting the notation which will be used in this chapter. We will be considering stochastic conversions, according to the general definition 15, here applied to LOCC conversions.

**Definition 21.** Let $0 \leq p \leq 1$. We say that $|\psi\rangle$ can be locally converted into $|\varphi\rangle$ with probability of success $p$, and write $|\psi\rangle \xrightarrow{p} |\varphi\rangle$, if there exists a LOCC protocol
which yields the state $\varphi$ with probability $p$ when performed on the starting state $\psi$. For a given pair $|\psi\rangle$ and $|\varphi\rangle$, we denote by

$$\Pi(\psi \rightarrow \varphi)$$

the maximal $p$ such that $|\psi\rangle \xrightarrow{p} |\varphi\rangle$. We will write $|\psi\rangle \rightarrow |\varphi\rangle$ to say that the conversion can be carried out deterministically, i.e. $\Pi(\psi \rightarrow \varphi) = 1$.

Further details on the probability of success associated to a LOCC protocol are provided in Appendix C.

This chapter contains analytical considerations and numerical work, intended to provide insight in questions related to the “volume” of the set of LOCC-convertible quantum states.

**Question 1.** Pick two states $|\psi\rangle$ and $|\varphi\rangle$ at random. What is the probability that $|\psi\rangle \rightarrow |\varphi\rangle$ (i.e. that $\Pi(\psi \rightarrow \varphi) = 1$)?

If the probability measure is unitarily invariant, the question can be rephrased in terms of volumes of the set of convertible states.

Nielsen conjectured that, “in the limit where $\mathcal{H}_A$ and $\mathcal{H}_B$ are of large dimensionality, almost all pairs of pure states picked according to the unitarily invariant measure of $\mathcal{H}_A \otimes \mathcal{H}_B$ will be incomparable” [117]. He informally justified the conjecture with a probabilistic argument which will be analyzed in detail in Section 5.2.1; shortly after, another heuristic explanation based on integral geometry was put forward in [166]. A rigorous result around this question is the proof that for an infinite-dimensional system, the set of pairs of LOCC-convertible pure states is nowhere dense [32] (this statement though does not imply the conjecture).

Questions about the volume of LOCC-convertible states can be relaxed to include local convertibility that succeed with some probability $p$ according to Definition 21.

**Question 2.** Pick two states $|\psi\rangle$ and $|\varphi\rangle$ at random. Given $p \in [0, 1]$, what is the probability that $|\psi\rangle \xrightarrow{p} |\varphi\rangle$ (i.e. that $\Pi(\psi \rightarrow \varphi) \geq p$)?

**Question 3.** Pick two states $|\psi\rangle$ and $|\varphi\rangle$ at random. Find the probability distribution of $\Pi(\psi \rightarrow \varphi)$.

In this chapter we address the above three questions. The investigation of local convertibility of quantum states has led to precise criteria based on the majorization relation [117, 156, 155, 119].

Hereafter we let $|\psi\rangle$ and $|\varphi\rangle$ be independent random pure states distributed according to the unitarily invariant measure on the unit sphere of $\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_B$, where $\mathcal{H}_A = \mathbb{C}^n$ and $\mathcal{H}_B = \mathbb{C}^m$. It is known that, since only local unitary transformations are allowed on $A$ and $B$, and these are locally reversible, any two states with the same Schmidt coefficients are locally equivalent. Thus only the Schmidt coefficients of $|\psi\rangle$ and $|\varphi\rangle$ are relevant as far as local operations are concerned. To be more precise, let $\lambda(\psi_A)$ and
\( \lambda(\varphi_A) \) be the entanglement spectra of \(|\psi\rangle\) and \(|\varphi\rangle\), i.e. the \(n\)-tuples of eigenvalues (squares of the Schmidt coefficients) of the reduced density matrices \( \psi_A = \text{tr}_B(|\psi\rangle\langle\psi|) \) and \( \varphi_A = \text{tr}_B(|\varphi\rangle\langle\varphi|) \) of \(|\psi\rangle\) and \(|\varphi\rangle\), respectively. Vidal [156] discovered that the maximal probability of local conversion depends on the entanglement spectra only and can be written as

\[
\Pi(\psi \rightarrow \varphi) = \Pi(\lambda(\psi_A), \lambda(\varphi_A)),
\]

where \( \Pi(\cdot, \cdot) \) is a known function given in Theorem 19.

The properties of \( \Pi(x, y) \) and its algebraic significance will be discussed later.

We can, therefore, give an equivalent formulation of Question 3 as follows.

**Question 3’.** Pick two states \(|\psi\rangle\) and \(|\varphi\rangle\) at random. Find the probability distribution of \( \Pi(\lambda(\psi_A), \lambda(\varphi_A)) \).

We also remark at this stage that the Schmidt decomposition is symmetric under the interchange of \(A\) and \(B\); hence, without loss of generality we can assume \(n \leq m\).

### 5.2 Deterministic conversions of pure bipartite states

We now turn to the problem of characterizing the deterministic conversions among pure bipartite states using only LOCC transformations. In [117], Nielsen found a criterion which allows to determine if \(|\psi\rangle \xrightarrow{\text{LOCC}} |\varphi\rangle\) for a given pair of states \(|\psi\rangle, |\varphi\rangle\) of a bipartite system \(\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_B\), based on a majorization relation between the Schmidt coefficients of \(|\psi\rangle\) and \(|\varphi\rangle\). Before presenting the criterion, it is useful to get an intuition on how the majorization relation can play a role in the resource theory of entanglement.

Recall that a pure state \(\psi = |\psi\rangle\langle\psi|\) is separable if the reduced state \(\psi_A = \text{tr}_B(|\psi\rangle\langle\psi|)\) on subsystem \(\mathcal{H}_A\) is a rank-one projector; in this case (up to a permutation) \(\lambda(\psi_A) = (1, 0, \ldots, 0)\). The reduced state of a maximally entangled state \(|\psi\rangle\) is \(\psi_A = I/n\); in this case \(\lambda(\psi_A) = (1/n, \ldots, 1/n)\). For \(|\psi\rangle \in \mathcal{H}\), we call the number of nonzero components of \(\lambda(\psi_A)\) the Schmidt number of \(|\psi\rangle\), and entangled states are those characterized by a Schmidt number strictly larger than one. It seems likely that for pure states the “amount of entanglement” is connected in some way to how “uniform” the corresponding vector of Schmidt coefficients is. Furthermore, one expects that under LOCC transformations entanglement cannot increase, so if the intuition on the uniformity of Schmidt coefficients is correct, one expects that it should not be possible to get a more uniform Schmidt vector starting from a more concentrated one. This intuition turns out to be correct, and even more: one finds out that bipartite pure states conversions are entirely governed by how uniform the corresponding Schmidt coefficients are, through the majorization relation.

**Theorem 17** (Nielsen’s criterion [117]). Let \(|\psi\rangle, |\varphi\rangle \in \mathcal{H}\). Then, \(|\psi\rangle \rightarrow |\varphi\rangle\) if and only if \(\lambda(\psi_A) \prec \lambda(\varphi_A)\).
Immediate consequences of Nielsen’s criterion are:

1. A maximally entangled state can be locally converted into any pure state, since the corresponding Schmidt vector is the uniform one \( \lambda(\psi_A) = (1/n, \ldots, 1/n) \), which is majorized by all the other vectors;

2. Any pure state can be locally converted into a separable state, since the Schmidt vector of the latter is the most concentrated one \( \lambda(\psi_A) = (1, 0, \ldots, 0) \), which majorizes all the others.

The relation between LOCC-convertibility and the algebraic notion of majorization evidences the fact that not all the pure quantum states are comparable: it is possible to find pairs of pure bipartite states for which the conversion cannot be carried out in neither direction. For example, if one consider a pair of states \( |\psi\rangle, |\phi\rangle \in H_A \otimes H_B \) with local spectra \( \lambda(\psi_A) = (0.6, 0.2, 0.2) \) and \( \lambda(\varphi_A) = (0.5, 0.4, 0.1) \), it is clear that neither \( \lambda(\psi_A) \prec \lambda(\varphi_A) \) nor \( \lambda(\psi_A) \succ \lambda(\varphi_A) \). Then, as a consequence of Nielsen criterion, a LOCC-conversion between \(|\psi\rangle\) and \(|\phi\rangle\) is not possible in both directions.

In [117] Nielsen conjectured that the probability that two states chosen at random on \( H = \mathbb{C}^n \otimes \mathbb{C}^m \) (according to the natural uniform Haar distribution) are locally convertible goes to zero as \( n, m \to \infty \). Equivalently, the relative volume of LOCC-convertible states of \( H \) vanishes in the limit of large dimension. Nielsen’s criterion translates the problem of LOCC-convertibility in terms of the majorization relation between their local spectra. Furthermore, we know that the local spectra of two Haar-random pure bipartite states on \( H = \mathbb{C}^n \otimes \mathbb{C}^m \) are distributed according to the probability density function (3.25). Therefore, the conjecture can be stated as

**Conjecture 1** (Nielsen [117]). Let \( x \) and \( y \) be independently distributed according to the probability density (3.25). Then \( P(x \prec y) \to 0 \) as \( n, m \to \infty \).

Before analysing extensively this conjecture with numerical results, let us see how one can justify it by interpreting the majorization relation as the positivity of a certain random walk.

### 5.2.1 Majorization as the persistence of a random walk above the origin

For \( x, y \in \Delta_{n-1} \), define \( \delta = y^i - x^i \in \mathbb{R}^n \). Then, the event \( x \prec y \) coincides with the event that the process \( S_k = \sum_{j=1}^{k} \delta_j \), starting at \( S_0 = 0 \), stays positive

\[
P(x \prec y) = P \left( \sum_{j=1}^{k} \delta_j \geq 0, \text{ for } 1 \leq k \leq n \right) = P \left( (S_k) \text{ stays positive} \right). \tag{5.3}
\]

Therefore, the probability that \( x \prec y \) can be recast as the persistence probability (the probability that a stochastic process does not change sign) of the
5.2. Deterministic conversions of pure bipartite states

process

\[ S_k = S_{k-1} + \delta_k, \quad S_0 = 0, \quad (5.4) \]
\[ \delta_k = y^\uparrow_k - x^\uparrow_k \quad (5.5) \]

For "generic" random vectors \( x, y \in \mathbb{R}^n \), it is plausible that \( (S_k) \) is unlikely to stay positive at all times if the number of steps \( n \) is large. This statement would be true if \( S_k \) were a simple random walk (independent and identically distributed steps \( \delta_k \)). However, as already remarked by Nielsen [117], \( (S_k) \) is not a simple random walk:

1. The steps \( \delta_k \) are not independent and identically distributed (the components of \( x^\uparrow_k \) and \( y^\uparrow_k \) are neither independent nor identically distributed);

2. The normalization of \( x \) and \( y \) implies that \( (S_k) \) is a ‘random bridge’ starting at \( S_0 = 0 \) and ending at \( S_n = 0 \).

It is interesting to compare the process defined by (5.4)–(5.5) with the classical examples of random walks of the form

\[ S_k = S_{k-1} + \eta_k. \quad (5.6) \]

In a general investigation of persistence probability (in modern language) of sums of random variables, Sparre Andersen singled out two exactly solvable cases. Denote by \( N_n \) the number of sums \( S_1, \ldots, S_n \), which are nonnegative.

**Theorem 18** (Sparre Andersen [5]).

(i) (Random walks with independent steps) If the steps \( \eta_1, \ldots, \eta_n \) are independent and identically distributed with continuous distribution symmetric around 0, then,

\[ P(N_n = k) = (-1)^n \left( \frac{1}{2} \right) \binom{\frac{1}{2}}{n-k}; \quad (5.7) \]

In particular, the persistence probability is

\[ P((S_k) \text{ stays positive}) = P(N_n = n) = (-1)^n \left( \frac{1}{2} \right) \binom{1/2}{n} \approx \frac{1}{\sqrt{\pi n}}, \quad \text{as } n \to \infty. \quad (5.8) \]

Moreover, the time spent above zero is asymptotically described by the arcsine distribution:

\[ P \left( \frac{N_n}{n} \leq t \right) \to \frac{2}{\pi} \arcsin \sqrt{t}, \quad \text{as } n \to \infty. \quad (5.9) \]

(ii) (Random bridges with exchangeable steps). If the steps \( \eta_1, \ldots, \eta_n \) are exchangeable random variables with continuous distribution and \( S_n = 0 \) almost surely, then \( N_n \) is uniformly distributed over \( \{1, \ldots, n\} \):

\[ P(N_n = k) = \frac{1}{n}; \quad (5.10) \]
In particular, the persistence probability is

\[ P((S_k) \text{ stays positive}) = P(N_n = n) = \frac{1}{n}. \quad (5.11) \]

In various discrete time processes, the persistence probability turns out to decay algebraically at large times, \( P((S_k) \text{ stays positive}) \simeq b/n^\theta \) [49]. The persistence exponent \( \theta \) carries information about the full history of the process. For this reason, \( \theta \) is usually a nontrivial exponent whose calculation is typically very challenging for non-Markovian processes. For an extensive review, see [22]. The two classes covered by the Sparre Andersen theorem, are examples of exactly solvable models where the persistence exponent and the prefactor can be computed explicitly (\( \theta = 1/2, b = 1/\sqrt{\pi} \) for random walks, and \( \theta = 1, b = 1 \) for random bridges).

It is interesting to note that in our original problem, if we drop the ordering on the local spectra, then the persistence probability problem changes dramatically and it becomes exactly solvable. Specifically, if we replace (5.5) with \( \delta_k = y_k - x_k \), then the process \( (S_k) \) becomes a random bridge with exchangeable steps \( x \) and \( y \) follow the distribution \( g_{n,m} \) of Equation (3.25), which is invariant under permutation) and Sparre Andersen’s theorem applies. Hence, in such a simplified problem we can show that the persistence probability goes to zero for large \( n \) (and how fast it goes to zero). Needless to say, majorization is expressed through the ordered vectors \( x^{\downarrow} \) and \( y^{\downarrow} \), and thus we are interested in the process \( (S_k) \) with steps \( \delta_k \) rather than the exchangeable steps \( \tilde{\delta}_k \).

5.3 Numerical results on deterministic conversion

Using the tridiagonal realization presented in 3.3.1 we sampled, for several values of \( n \) and \( m, M = 5 \cdot 10^5 \) pairs of spectra \( \lambda(\psi_A) \) and \( \lambda(\varphi_A) \) distributed according to (3.25). These pairs correspond to the local spectra of states \( |\psi\rangle, |\varphi\rangle \) from bipartite Hilbert spaces \( \mathcal{H} \) of dimension ranging from 4 (two qubits) up to \( 2^{20} \cdot 10^3 \simeq 10^9 \). Using this sample we computed numerically the distribution (and other statistics) of \( \Pi(\psi \rightarrow \varphi) = \Pi(\lambda(\psi_A), \lambda(\varphi_A)) \).

5.3.1 Confirmation of Nielsen Conjecture

Our findings are the following.

1. For fixed \( n \), numerical computations show that the sequence \( P(|\psi\rangle \rightarrow |\varphi\rangle) \) is non-decreasing in \( m \). Therefore, as \( m \rightarrow \infty \), the probability of local conversion converges to a positive constant:

\[ \lim_{m \rightarrow \infty} P(|\psi\rangle \rightarrow |\varphi\rangle) = \kappa(n) > 0. \quad (5.12) \]

The sequence of constants \( \kappa(n) \) is decreasing in \( n \). See Figure 5.1a. The result is of course symmetric under exchange of \( n \) and \( m \).
2. Suppose that both $n, m \to \infty$. We verified numerically that the relative volume of pairs of LOCC-convertible states goes to zero

$$
\lim_{n,m \to \infty} P (|\psi\rangle \to |\varphi\rangle) = 0.
$$

(5.13)

(This is a precise statement of Nielsen’s conjecture.) See Figure 5.1b.

3. We can analyse the asymptotic rate at which the probability of local conversion decays to zero. More precisely, we consider the limit $n, m \to \infty$ with the ratio $\frac{m}{n} = c$ fixed. We can assume $c \geq 1$. In fact, the results are symmetric with respect to the exchange of $n$ and $m$, i.e. the transformation $c \leftrightarrow c^{-1}$ (the fixed point $c = 1$ of the transformation is indeed quite special as discussed later). The analysis of the problem reveals a mapping to the calculation of the persistence probability for a non-Markovian random walk (see Section 5.2.1). This connection, combined with numerical evidence (see Figure 5.1b), suggests that the probability of local conversion decays as a power law at large $n$. Indeed, we find numerically that

$$
P (|\psi\rangle \to |\varphi\rangle) = P (\Pi (\psi \to \varphi) = 1) \simeq \frac{b}{n^\theta}, \quad \text{as } n \to \infty \text{ with } m = cn.
$$

(5.14)

A few values of the so-called ‘persistence exponent’ $\theta$ and prefactor $b$ are shown in Table 5.1.

Hence, we have that the persistence exponent is $\theta \simeq 4/5$ if $c = 1$, and $\theta \simeq 2/5$ if $c > 1$. 

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure5.1.png}
\caption{The probability $P (|\psi\rangle \to |\varphi\rangle)$ for random pure states $\psi$ and $\varphi$ versus the parameter $c$ for several values of $n$ (a), and versus the dimension $n$ for several values of the parameter $c = m/n$ (b).}
\end{figure}
Chapter 5. Random State Conversions in the Resource Theory of Entanglement

<table>
<thead>
<tr>
<th>$c$</th>
<th>$\theta$</th>
<th>$b$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.795 ± 0.013</td>
<td>0.579 ± 0.041</td>
</tr>
<tr>
<td>3/2</td>
<td>0.420 ± 0.006</td>
<td>0.302 ± 0.010</td>
</tr>
<tr>
<td>2</td>
<td>0.418 ± 0.006</td>
<td>0.348 ± 0.010</td>
</tr>
<tr>
<td>5</td>
<td>0.409 ± 0.005</td>
<td>0.394 ± 0.011</td>
</tr>
<tr>
<td>10</td>
<td>0.400 ± 0.005</td>
<td>0.391 ± 0.010</td>
</tr>
<tr>
<td>100</td>
<td>0.408 ± 0.005</td>
<td>0.423 ± 0.011</td>
</tr>
<tr>
<td>1000</td>
<td>0.405 ± 0.005</td>
<td>0.422 ± 0.011</td>
</tr>
</tbody>
</table>

Table 5.1: Parameters $b$ and $\theta$ of the best fitting curve for the probability of local conversion, $P(|\psi\rangle \rightarrow |\phi\rangle) \simeq b/n^{\theta}$. The fit is taken only over the last 4 points of each curve in Figure 5.1b.

5.3.2 Discussion on the persistence exponents

To get some insight, we studied numerically for the process (5.4)–(5.5) the number $N_n$ of sums $S_1, \ldots, S_n$, which are nonnegative, i.e. the total time spent by the process $(S_k)$ above 0. The distribution of $N_n$ is summarised in Figure 5.2, and compared with the uniform distribution (of classical random bridges) and the arcsine distribution (of classical random walks). We see that in the case $c = 1$ (balanced partitions $m = n$), the distribution of $N_n$ is close to uniform, but $(S_k)$ is less likely to cross 0 compared to a random bridge with exchangeable steps. Indeed, we found numerically that the persistence probability decays with exponent $\theta \simeq 4/5$ (smaller than the exponent $\theta = 1$ of classical random bridges). As $c$ increases, the distribution of $N_n$ looks more and more similar to the arcsine distribution. Again however, $(S_k)$ is less likely to cross 0 compared to a random walk with independent steps; indeed, the persistence exponent for $c > 1$ is $\theta \simeq 2/5$, again smaller than the value $\theta = 1/2$ of classical random walks.

In summary, we conclude that the ordering of $x^k$ and $y^k$ defining the steps (5.5) of the process $(S_k)$, makes the persistence exponent $\theta$ nontrivial (and hard to compute). Nevertheless, there is a sharp distinction in the persistence probability corresponding to balanced ($c = 1$) and unbalanced ($c > 1$) partitions. The change of behaviour at $c = 1$ is consistent with the aforementioned symmetry of the problem under exchange of $m$ and $n$, i.e. the symmetry $c \leftrightarrow c^{-1}$. In the following section, we will see that this different behaviour characterizes also the maximal success probability of stochastic conversions, and we will make the point that the existence of two distinct persistence exponents is related to the completely different asymptotic behaviour of the smallest eigenvalue $\lambda(|\psi_A\rangle)^i_n$ when $c$ is strictly larger, or equal to 1.
5.4 Stochastic state conversions

Nielsen’s criterion is a result about deterministic LOCC-conversions, i.e. about the possibility to find an operation in the class LOCC such that, given pure bipartite states $|\psi\rangle, |\varphi\rangle$, the probability of converting $|\psi\rangle$ into $|\varphi\rangle$ is $\Pi(|\psi\rangle \rightarrow |\varphi\rangle) = 1$. Suppose that a deterministic local conversion of $|\psi\rangle$ into $|\varphi\rangle$ cannot be achieved; is there any procedure realizing the conversion with positive probability of success? And if this is the case, is it possible to determine the maximum probability of success $\Pi(|\psi\rangle \rightarrow |\varphi\rangle)$ in the conversion? The precise answer to these questions was found by Vidal [156] and is expressed in the following theorem.

**Theorem 19** (Vidal’s criterion [156]). Let $|\psi\rangle, |\varphi\rangle \in \mathcal{H}$. Then,

$$\Pi (|\psi\rangle \rightarrow |\varphi\rangle) = \Pi(\lambda(\psi_A), \lambda(\varphi_A))$$ (5.15)

where

$$\Pi(x, y) = \min_{1 \leq k \leq n} \frac{\sum_{j=k}^{n} x_j^+}{\sum_{j=k}^{n} y_j^+}.$$ (5.16)

Note that Vidal’s theorem is an extension of Nielsen’s result. Indeed,

$$x \prec y \iff \Pi(x, y) = 1.$$ (5.17)

We can also write (as originally done by Vidal)

$$\Pi(x, y) = \min_{1 \leq k \leq n} \frac{E_k(x)}{E_k(y)},$$ (5.18)

where the functions $E_k(x) = \sum_{j=k}^{n} x_j^+$ are entanglement monotones (Schur-concave functions).
We now give a geometric characterization of the function $\Pi(\cdot, \cdot)$.

**Proposition 9.** Let $e = (1, 0, \ldots, 0) \in \Delta_{n-1}$. Then,

$$\Pi(x, y) = \max \left\{ 0 \leq p \leq 1 : x^k \prec y^k + (1 - p)e \right\}.$$  \hfill (5.19)

**Proof.** To prove (5.19), first note that:

$$\Pi(x, y) = \max \left\{ 0 \leq p \leq 1 : \sum_{j=k}^{n} p y_k^j \leq \sum_{j=k}^{n} x_k^j \text{ for } 1 \leq k \leq n \right\},$$  \hfill (5.20)

and then observe that

$$\sum_{j=k}^{n} p y_k^j \leq \sum_{j=k}^{n} x_k^j \text{ for } 1 \leq k \leq n \iff x^k \prec y^k + (1 - p)e.$$  \hfill (5.21)

Indeed, the term $(1 - p)e$ does not affect the inequalities $\sum_{j=k}^{n} p y_k^j \leq \sum_{j=k}^{n} x_k^j$ for $k = 2, \ldots, n$, while saturates the last inequality for $k = 1$, ensuring the normalization condition.

This remark gives a geometric picture of the maximal probability of success: $|\psi\rangle$ cannot be deterministically converted into $|\varphi\rangle$ (i.e., $\Pi(\psi \rightarrow \varphi) < 1$), if and only if $\lambda(\psi_A)$ is not majorized by $\lambda(\varphi_A)$ or, in other words, if $\lambda(\psi_A)$ is not contained in the polytope given by the convex hull of the permutations of $\lambda(\varphi_A)$ (see again Figure 4.1). The analogue polytope associated to a convex combination of $\lambda(\varphi_A)^k$ with the largest element $e$ is always larger than the polytope associated to $\lambda(\varphi_A)^k$. We can therefore consider an enlarged polytope associated to $p\lambda(\varphi_A)^k + (1 - p)e$ that contains $\lambda(\psi_A)$ (such a polytope always exists). $\Pi(\lambda(\psi_A), \lambda(\varphi_A))$ represents the maximum weight we can assign to $\lambda(\varphi_A)$ in the convex combination such that this is possible. In summary, we have

$$\Pi(\psi \rightarrow \varphi) = \max \left\{ 0 \leq p \leq 1 : \lambda(\psi_A)^k \prec p\lambda(\varphi_A)^k + (1 - p)e \right\}.$$  \hfill (5.22)

### 5.4.1 Distribution of the maximal success probability

In the previous section we considered the probability

$$P(|\psi\rangle \rightarrow |\varphi\rangle) = P(\Pi(\psi \rightarrow \varphi) = 1)$$

for pairs of random states independent and uniformly distributed on the unit sphere of $\mathcal{H} = \mathbb{C}^n \otimes \mathbb{C}^m$. By Nielsen’s criterion (Theorem 17), this is the probability $P(x \prec y)$ that two independent probability vectors distributed according to $g_{n,m}$ (see equation (3.25)) satisfy the majorization relation; recall that $P(x \prec y) = P(\Pi(x, y) = 1)$. In this section we discuss the full distribution of $\Pi(x, y)$, i.e:

$$F(p) = P(\Pi(\psi \rightarrow \varphi) \leq p).$$  \hfill (5.23)
5.4. Stochastic state conversions

Having established that

\[ P(\Pi(\psi \rightarrow \varphi) = 1) = P(\Pi(x, y) = 1) = F(1) - F(1^-) \rightarrow 0 \quad \text{as } n \rightarrow \infty, \]

(5.24)

we studied numerically the asymptotic behaviour of the full distribution of \( \Pi(x, y) \). A first natural step in the analysis is the calculation of the expectation value

\[ E[\Pi(\psi \rightarrow \varphi)] = E[\Pi(\lambda(\psi_A), \lambda(\varphi_A))]. \]

(5.25)

Eq. (5.25) is the average maximal probability of successful LOCC conversions between two states picked at random. The behaviour of this expectation value as a function of \( n \) and \( m \) is summarised in Figure 5.3.

We note an interesting and surprising asymptotic behaviour for unbalanced bipartitions \((c > 1)\). As \( n \) and \( m \) increase with \( c > 1 \) fixed, the average maximal probability of conversion eventually approaches the limit value 1. In formulas:

\[ E[\Pi(\psi \rightarrow \varphi)] = \int_0^1 pf(p)dp \rightarrow 1 \quad \text{as } n \rightarrow \infty \]

(5.26)

This is at first surprising given that the probability of successful conversions goes to zero (5.24). Equations (5.26) and (5.24) disclose an interesting structure of the density \( f(p) \) of the maximal probability of conversion. Since for any finite \( n \) there is a nonvanishing probability to have an exact conversion, i.e. \( P(|\psi \rangle \rightarrow |\varphi \rangle) = F(1) - F(1^-) \neq 0 \), there is a singular part of the distribution at \( p = 1 \) that goes to zero as \( n \) grows. At the same time, equation (5.26) tells that there is a continuous part \( f_{\text{cont}}(p) \) which concentrates around (on the left of) \( p = 1 \) as \( n \rightarrow \infty \). In the next subsection we will compute exactly this distribution for \( n = 2 \) and \( m \) generic to explicitly show this structure. The calculation for generic \( n \) is more complicated, but a numerical computation of this distribution confirms a similar structure.
We conclude this section by stressing the physical significance of equations (5.26) and (5.24): the relative volume of pairs of pure states that can be successfully converted one into another using local operations vanishes in the asymptotic limit of large dimension; on the other hand, if the bipartition is unbalanced \( (c > 1) \), then the overwhelming majority of states are LOCC-convertible if we allow an arbitrarily small probability of failure in the local conversion.

5.4.2 **Exact formulae for** \( n = 2 \)

For \( n = 2 \), \( x \prec y \iff x_1^+ \leq y_1^+ \). Since \( x \) and \( y \) are independent and identically distributed with continuous density \( g_{2,m} \), we have

\[
P(x \prec y) = \frac{1}{2}, \quad \text{for } n = 2. \tag{5.27}
\]

We now compute the probability density \( f(p) = F'(p) \) of \( \Pi(x,y) \) for \( n = 2 \) and \( m \geq n \). For \( n = 2 \),

\[
\Pi(x,y) = \min \left\{ 1, \frac{x_2^+}{y_2^+} \right\}, \tag{5.28}
\]

the function \( \Pi(x,y) \) only depends on the ratio of the smallest eigenvalues \( x_2^+ \) and \( y_2^+ \), and we have

\[
f(p) = \int_{\Delta_1} dx \int_{\Delta_1} dy \, g_{2,m}(x)g_{2,m}(y) \delta \left( p - \min \left\{ 1, \frac{x_2^+}{y_2^+} \right\} \right)
\]

\[
= \int_0^{1/2} ds \int_0^{1/2} dt \, q_{2,m}(s)q_{2,m}(t) \delta \left( p - \min \left\{ 1, \frac{s}{t} \right\} \right),
\]

where \( q_{2,m}(s) = 2g_{2,m}(1 - s, s) \) reads

\[
q_{2,m}(s) = \frac{\Gamma(2m)}{\Gamma(m)\Gamma(m-1)}(s - s^2)^{m-2}(1 - 2s)^2. \tag{5.29}
\]

Then, by splitting the integral we get

\[
f(p) = \int_0^{1/2} dt \int_t^{1/2} ds \, q_{2,m}(s)q_{2,m}(t) \delta(p - 1) + \int_0^{1/2} dt \int_0^t ds \, q_{2,m}(s)q_{2,m}(t) \delta \left( p - \frac{s}{t} \right)
\]

\[
= P(x \succ y) \delta(p - 1) + 1_{[0,1]}(p) \int_0^{1/2} dt \, tq_{2,m}(pt)q_{2,m}(t). \tag{5.30}
\]

We see then that \( f(p) \) is the sum of a singular part and a continuous density

\[
f(p) = f_{\text{cont}}(p) + \frac{1}{2} \delta(p - 1), \tag{5.31}
\]
where we used (5.27) and we have defined

\[ f_{\text{cont}}(p) = 1_{[0,1]}(p) \int_0^{1/2} t \, q_{2,m}(t) q_{2,m}(pt) \, dt. \]  

(5.32)

Of course, \( f_{\text{cont}}(p) \geq 0 \) and \( \int_0^1 f_{\text{cont}}(p) \, dp = 1/2 \). In fact, for \( 0 \leq p \leq 1 \), \( f_{\text{cont}}(p) \) is a polynomial in \( p \) of degree \( 2m - 2 \). Explicit formulae for the first values of \( m \) are

\[ f_{\text{cont}}(p) = \frac{3}{20} \left(p^2 - 4p + 5\right), \quad m = 2, \]  

(5.33)

\[ f_{\text{cont}}(p) = -\frac{5}{224} p \left(13p^3 - 80p^2 + 165p - 120\right), \quad m = 3, \]  

(5.34)

\[ f_{\text{cont}}(p) = \frac{105}{36608} p^2 \left(139p^4 - 1148p^3 + 3549p^2 - 4888p + 2574\right), \quad m = 4. \]  

(5.35)

In Figure 5.4 we show a comparison between numerical results and the analytical expressions (5.33)–(5.35). It is not hard to show that for large \( m \) the continuous density \( f_{\text{cont}}(p) \) concentrates at \( p = 1 \), \( \lim_{m \to \infty} f_{\text{cont}}(p) = \frac{1}{2} \delta(1 - p) \), so that

\[ \lim_{m \to \infty} F(p) = \theta(p - 1), \quad \text{for } n = 2. \]  

(5.36)

### 5.4.3 Distribution function of the maximal success probability and scaling limits

A numerical computation of the distribution of \( \Pi(\lambda(\psi_A), \lambda(\varphi_A)) \) for \( n > 2 \) has been performed, whose results are shown in Figures 5.5a and 5.5b for \( c = 1 \) and \( c > 1 \) respectively. These numerical results show that the structure of the probability density function (5.31) is preserved for \( n > 2 \), i.e. we have

\[ f(p) = f_{\text{cont}}(p) + \kappa \delta(p - 1) \]  

(5.37)

for every \( n \). Clearly the weight of the singular part is associated with the probability of having a successful conversion strategy, i.e. \( P(\Pi(x, y) = 1) = \kappa > 0 \), and we retrieve in Figures 5.5a and 5.5b that \( \kappa \to 0 \) as \( n \to \infty \). A remarkable difference can be noted between the balanced \( (c = 1) \) and unbalanced \( (c > 1) \) bipartition. In the large \( n \) limit, while for \( c = 1 \) the density \( f_{\text{cont}} \) is supported on the whole interval \([0, 1]\), for \( c > 1 \) we see (as announced before) that the support of \( f_{\text{cont}} \) concentrates around \( p = 1 \). This difference explains the different behaviour observed in Figure 5.3 between the balanced \( (c = 1) \) and unbalanced \( (c > 1) \) cases. In formulae, we found evidence that for \( m = cn \ (c > 1 \text{ fixed}) \),

\[ \lim_{n \to \infty} F(p) = \theta(p - 1), \]  

(5.38)
where $\theta(x)$ is the Heaviside step function, defined as $\theta(x) = 1$ for $x \geq 0$ and zero otherwise. This result is symmetric under exchange of $n$ and $m$. This convergence in distribution would imply convergence in probability

$$P(\Pi(\psi \rightarrow \varphi) \leq 1 - \varepsilon) \rightarrow 0, \quad \text{for every } \varepsilon > 0,$$

as $m \rightarrow \infty$ or $n \rightarrow \infty$ (or both with $c > 1$).

We conjecture that the distribution of $\Pi(\lambda(\psi_A), \lambda(\varphi_A))$ is driven by the smallest eigenvalues of $\psi_A$ and $\varphi_A$, i.e. $\lambda(\psi_A)_1^n, \lambda(\varphi_A)_1^n$. Recall that

$$\Pi(\lambda(\psi_A), \lambda(\varphi_A)) =$$

$$= \min \left\{ \frac{\lambda(\psi_A)_1^n}{\lambda(\varphi_A)_n^1}, \frac{\lambda(\psi_A)_2^n + \lambda(\psi_A)_1^n}{\lambda(\varphi_A)_n^1 + \lambda(\varphi_A)_1^n}, \ldots, \frac{\lambda(\psi_A)_n^n + \cdots + \lambda(\psi_A)_2^n}{\lambda(\varphi_A)_n^n + \cdots + \lambda(\varphi_A)_2^n}, 1 \right\}.$$

If each component of $\lambda(\psi_A)_k^1, \lambda(\varphi_A)_k^1$ does not vary too much with respect to its mean, then all the sums in the numerators are very close to the sums in the denominators of (5.40), implying that the minimum of their ratios is unlikely to be much less than one. In order to test this idea, we computed numerically the variance $\text{Var}[\lambda(\psi_A)_k^1]$ and the mean $\mathbb{E}[\lambda(\psi_A)_k^1]$ of the $k$-th largest eigenvalue of the density matrix $\psi_A$, and its relative fluctuation

$$\sqrt{\text{Var}[\lambda(\psi_A)_k^1] / \mathbb{E}[\lambda(\psi_A)_k^1]}.$$

(5.41)
5.4. Stochastic state conversions

Figure 5.5: Distribution function $F(p)$ of $\Pi(\psi \to \varphi)$. The decomposition (5.37) of the corresponding probability density function $f(p)$ into a singular and a continuous component can be seen from the presence of a jump at $p = 1$. In both the balanced and unbalanced cases the height of the jump vanishes as $n$ grows.

In Figure 5.6 we plotted the relative fluctuations (5.41) for large $n$ and several values of $c \geq 1$. The plots show that the main difference between the balanced and the unbalanced cases are in the fluctuations of the smallest eigenvalues, in particular the minimum eigenvalue $\lambda(\psi_A)^n_{\downarrow}$. This is somehow expected. Indeed, for large $n$, the rescaled spectrum $n \lambda(\psi_A)$ concentrates on a bounded interval $[a, b] \subset \mathbb{R}_+$, and has a continuous limit density given by the Marčenko-Pastur distribution (see Theorem 11)

$$
\rho_c(x) = \frac{c}{2\pi \sqrt{(x-a)(b-x)}} 1_{[a,b]}(x),
$$

where $a = (1 - 1/\sqrt{c})^2$ and $b = (1 + 1/\sqrt{c})^2$. When $c > 1$, the support of this density is bounded away from zero ($a > 0$) and vanishes at the lower edge as $(x-a)^{1/2}$ (soft edge); for $c = 1$, the support of the Marčenko-Pastur distribution contains the origin and diverges at $a = 0$ as $x^{-1/2}$ (hard edge). See Figure 5.7.

Using the results presented in section 3.4.3, we see that for large $n$, the fluctuations of the minimum eigenvalue $\lambda(\psi_A)^n_{\downarrow}$ at the soft or hard edges are described respectively by

$$
\frac{\sqrt{\text{Var}[\lambda(\psi_A)^n_{\downarrow}]} }{\mathbb{E}[\lambda(\psi_A)^n_{\downarrow}]} \sim \begin{cases} 1 & \text{if } c = 1 \text{ (hard edge)}, \\ \frac{1}{c^{1/6} |1 - \sqrt{c}|^{2/3}} n^{-2/3} & \text{if } c > 1 \text{ (soft edge)}. \end{cases}
$$

(5.43)

The fluctuations of the smallest eigenvalue are comparable with its mean in the balanced case, suggesting that the ratio

$$
\frac{\sum_{j=k}^{n} \lambda_j(\psi_A)^j}{\sum_{j=k}^{n} \lambda_j(\varphi_A)^j},
$$

(5.44)
Chapter 5. Random State Conversions in the Resource Theory of Entanglement

Figure 5.6: Relative fluctuations \( \sqrt{\text{Var}[\lambda(\psi_A)_{k}]/\text{E}[\lambda(\psi_A)_{k}]} \) of the \( k \)-th largest eigenvalue for \( k = 1, \ldots, n \). Here \( n = 1024 \).

Figure 5.7: Marčenko-Pastur distribution (5.42) with hard-edge for \( c = 1 \) (left) and soft edge for \( c > 1 \) (right).
5.4. Stochastic state conversions

takes values significantly smaller than one with a non-negligible probability for \( k = n \) (the ratio between the smallest eigenvalues); the same must happen to \( \Pi(\lambda(\psi_A), \lambda(\varphi_A)) \), since it is the minimum of (5.44) over \( 1 \leq k \leq n \). On the contrary, for unbalanced bipartitions the relative deviation of the smallest eigenvalue vanishes for large \( n \), and as shown in Figure 5.6 the same happens for all the other eigenvalues. This implies that the ratio (5.44) is typically nearly equal to 1 for every \( 1 \leq k \leq n \).

As a final test of our conjectural relation between \( \Pi(\lambda(\psi_A), \lambda(\varphi_A)) \) and the fluctuations of the minimum eigenvalues \( \lambda(\psi_A)_n^{1/n}, \lambda(\varphi_A)_n^{1/n} \), we ask whether the distribution function \( F(p) \) has a scaling limit driven by the fluctuations (5.43). Specifically, we established that

\[
\lim_{n,m \to \infty} \left( P \left( \Pi(\lambda(\psi_A), \lambda(\varphi_A)) \leq p \right) = \theta(1-p) \right).
\]

Is it possible to rescale the random variable \( \Pi(\lambda(\psi_A), \lambda(\varphi_A)) \) in order to obtain a nontrivial scaling limit? In particular, we would like to centre the variable around its limit value 1 and scale by its typical fluctuations. The previous discussion suggests the rescaling

\[
\tilde{\Pi}(\lambda(\psi_A), \lambda(\varphi_A)) = \frac{\mathbb{E}[\lambda(\psi_A)_n^{1/n}]}{\sqrt{\text{Var}[\lambda(\psi_A)_n^{1/n}]}} (1 - \Pi(\lambda(\psi_A), \lambda(\varphi_A)))
\]

\[
\sim \begin{cases} 
(1 - \Pi(\lambda(\psi_A), \lambda(\varphi_A))) & \text{if } c = 1 \text{ (hard edge)} \\
\left(1 - \frac{1}{\sqrt{c}}\right)^{2/3} n^{2/3} (1 - \Pi(\lambda(\psi_A), \lambda(\varphi_A))) & \text{if } c > 1 \text{ (soft edge)}
\end{cases}.
\]

(5.46)

It turns out that the rescaled variable \( \tilde{\Pi} \) has a nontrivial limit in distribution. The results of the numerical computation are rather convincing and are summarised in Figure 5.8. Formally, numerical results show that for \( m = cn \) with \( c > 1 \) fixed,

\[
\lim_{n \to \infty} P \left( \Pi(\psi \rightarrow \varphi) > 1 - \frac{x}{c^{1/6} \sqrt{1 - \sqrt{c}}^{2/3} n^{2/3}} \right) = H_{\text{unb}}(x),
\]

(5.47)

for some nontrivial scaling functions \( H_{\text{unb}}(x) \). We can equivalently write for the probability distribution

\[
\lim_{n \to \infty} F \left( 1 - \frac{x}{c^{1/6} \sqrt{1 - \sqrt{c}}^{2/3} n^{2/3}} \right) = 1 - H_{\text{unb}}(x).
\]

(5.48)

For \( c = 1 \) (the fixed point of the symmetry \( c \leftrightarrow c^{-1} \)) we observe instead a simpler asymptotics

\[
\lim_{n \to \infty} F(1 - x) = 1 - H_{\text{bal}}(x).
\]

(5.49)

A precise description of the limit distributions \( H_{\text{bal}} \) and \( H_{\text{unb}} \) of \( \tilde{\Pi} \) seems quite challenging and remains beyond reach.
Chapter 5. Random State Conversions in the Resource Theory of Entanglement

5.5 Conclusions and Outlook

In this chapter we have studied the probability of LOCC-convertibility between pairs of random pure states. We have found strong numerical evidence that the proportion of LOCC-convertible states of a bipartite system $AB$ vanishes in the asymptotic limit of large dimensional local Hilbert spaces, thus supporting Nielsen’s conjecture. By mapping the problem to the persistence probability of a discrete-time continuous-space random walk, we argued that the probability $P\left( |\psi\rangle \rightarrow |\varphi\rangle \right)$ of successful LOCC-conversions decays algebraically $\simeq 1/n^\theta$, and we gave precise estimates for the persistence exponents $\theta$. It turns out that the value of $\theta$ depends on whether the bipartition is balanced $m \sim n$ ($\theta \simeq 4/5$) or unbalanced $m \sim cn$ with $c > 1$ ($\theta \simeq 2/5$). Recall that the problem of LOCC-convertibility is symmetric under exchange of parties $A$ and $B$, and the special value $c = 1$ is the fixed point of the symmetry $c \leftrightarrow c^{-1}$. The sharp transition from $c = 1$ to $c > 1$ can also be observed in the full distribution $F(p)$ of $\Pi (|\psi\rangle \rightarrow |\varphi\rangle)$ (see Figure 5.5). Surprisingly enough, for unbalanced bipartitions ($c > 1$), although the probability of successful LOCC-conversion between random states vanishes asymptotically, the overwhelming majority of pairs can be LOCC-converted if any fixed arbitrary small probability of error is allowed. Random matrix theory turns out to be a useful tool to explain this phenomenon and elaborate on it. The main difference between the two regimes is that the density of eigenvalues $\rho_c(x)$ in (5.42) accumulates at zero for $c = 1$, while it has support bounded away from zero for $c > 1$. The fluctuations of the minimum eigenvalues have different scales in the two cases, and we showed that those fluctuations drive the scaling limit of $F(p)$. See Figs. 5.8a–5.8b.

Mapping the majorization relation between random vectors to persistence probabilities of random walks suggests that Nielsen’s conjecture is an instance of a more general phenomenon. It can be conjectured that if $x, y$ are
5.5. Conclusions and Outlook

independently distributed from a continuous and permutation invariant distribution on $\Delta_{n-1}$, then

$$P(x < y) \to 0 \quad \text{as } n \to \infty.$$  \hspace{1cm} (5.50)

In the next chapter we will see that this conjecture can be analytically proved, the proof involving the behaviour of the extreme values for uniformly distributed points on the simplex. One could also extend the problem to discrete measures on the simplex; in fact, an equivalent problem for the dominance order of random integer partitions is stated in Macdonald’s book [101, Chap. 1, Ex. 18] (see also next chapter).

The key numerical ingredient that made possible the analysis presented in this chapter is the tridiagonal realisation of the Laguerre unitary ensemble of section 3.3.1. It is worth emphasising that the general version of the tridiagonal method of Dumitriu and Edelman provides a way to sample $n$-tuples for the Laguerre $\beta$-ensemble for generic $\beta > 0$. In particular, one can easily adapt the method explained in Section 3.3.1 to investigate the set of random real pure states (corresponding to $\beta = 1$), a set that at first seems artificial, but which is in fact quite realistic for many situations (see [157, 164, 172, 170, 12]). In that case a similar behaviour as for complex random states can be expected, but the persistence exponents might change.

Let us conclude with a final remark on the ordered sums of the eigenvalues of random matrices. Natural ‘observables’ in study of spectral properties of random matrices are symmetric under permutations of the eigenvalues (these include moments, correlation functions, averages of characteristic polynomials, etc.). Therefore, the ordered sums $\lambda_1 + \cdots + \lambda_k$ might seem rather exotic for a random matrix theorist. Note however that, for the Laguerre unitary ensemble, the ordered sums $\lambda_1 + \cdots + \lambda_k$ are the outputs of the Robinson-Schensted-Knuth algorithm applied to an array of i.i.d. exponential waiting times [Johansson00, 19, 120]. Therefore, as a consequence of Greene’s theorem [72], the ordered sums have a very explicit and direct relation to last passage percolation models. (Related quantities, dubbed ‘truncated linear statistics’, were recently studied in [68].) There is no space here to elaborate on the representation theory behind this correspondence; the mere purpose of this remark is to stress the relevance of ordered sums of eigenvalues of random matrices.
Chapter 6

Resource Theory of Coherence

In this chapter, we introduce the Resource Theory of Quantum Coherence and present the results about convertibility among pure states in terms of a majorization relation between their coherence vectors. Using the distribution induced on the simplex of the coherence vectors by the uniform Haar measure on the set of pure states, we prove that if two $n$-dimensional random pure states are independently chosen according to the natural uniform distribution, then the probability that they are comparable vanishes as $n$ increases. We also study the maximal success probability of incoherent conversions and find an explicit formula for its large-$n$ asymptotic distribution. This analysis is based on the observation that the extreme values (largest and smallest components) of a random point uniformly sampled from the unit simplex are distributed asymptotically as certain explicit homogeneous Markov chains.

The structure of the chapter is as follows. In Section 6.1 we recall the definitions of incoherent, strictly incoherent and dephasing covariant quantum operations, and the connection between incoherent convertibility and the majorization relation. In Section 6.2 we present the distributions of the smallest and largest “components” of random pure quantum states; these are the main probabilistic properties relevant to our analysis. Section 6.3 contains the main results: the set of comparable states in the resource theory of coherence has volume zero in the limit of large dimension $n \to \infty$ (this is the analogue of Nielsen’s conjecture in the theory of entanglement discussed in the previous chapter); this problem is related to the persistence probability of a non-Markovian random walk and we give numerical estimates on the rate of decay to zero; in the limit $n \to \infty$ the maximal success probability of incoherent conversion between two random independent pure states has a nontrivial asymptotic distribution that is characterized completely in terms of a Markovian random walk. Finally, in Section 6.4, some concluding remarks are given.

6.1 Resource theories of (speakable) coherence

Here, we will consider the resource theories of coherence introduced and studied by Åberg [1], Baumgratz, Cramer, and Plenio [11], Winter and Yang [162], and Chitambar and Gour [29]. These are also referred to as resource theories of “speakable coherence” in the literature, as opposed to the resource theories of “unspeakable coherence” [104, 30]. Here we will focus only on the
former ones. As we will see briefly, several proposals have been considered concerning the definition of the allowed operations, all compatible with the the same choice of the free states, but each leading to a different resource theory, which is the reason why we are talking about multiple resource theories of speakable coherence.

6.1.1 Free States

In order to simplify the notation, in this chapter we will denote by $\mathcal{H}_n$ a complex Hilbert space of dimension $n$, and by $S_n \equiv S(\mathcal{H}_n)$ the corresponding set of states. Fix a basis $\{|i\rangle\}_{i=1}^n$ in $\mathcal{H}_n$, to be called the incoherent basis. The choice may be dictated by physical considerations (for example, the eigenbasis of a particular observable of interest, or the computational basis in quantum computation).

The set of free states in the resource theory of coherence is the set of incoherent states $\mathcal{I}_n \subset S_n$ defined as

$$\mathcal{I}_n := \left\{ \rho \in S_n : \rho = \sum_{i=1}^n p_i |i\rangle\langle i| \right\}, \quad (6.1)$$

i.e. density matrices which are diagonal in the incoherent basis. Notice that $\mathcal{I}_n$ is the image of $S_n$ under the decohering map, i.e. $\mathcal{I}_n = D(S_n)$, where

$$D(\rho) := \sum_{i=1}^n \langle i|\rho|i\rangle |i\rangle\langle i|. \quad (6.2)$$

Note that the only pure states of $\mathcal{I}_n$ are the elements of the incoherent basis.

6.1.2 Allowed Operations

As said above, there have been several proposals concerning the allowed operations, each yielding a different resource theory. For a thorough analysis of the alternatives, see [29, 28]. Here we focus on three possible choices of free operations, which have the interesting feature of being all characterized by majorization relations when considering conversions between pure states.

Recall that any quantum channel — that is a completely positive and trace preserving (CPTP) map $\mathcal{E} : S_n \to S_n$ — can be characterized in terms of a Kraus representation:

$$\mathcal{E}(\rho) = \sum_{\alpha} \mathcal{K}_\alpha(\rho) = \sum_{\alpha} K_\alpha \rho K_\alpha^\dagger, \quad (6.3)$$

where $\mathcal{K}_\alpha(\cdot) = K_\alpha(\cdot)K_\alpha^\dagger$, and $\{K_\alpha\}$ is a set of (non-uniquely determined) operators on $\mathcal{H}_n$, with $\sum_\alpha K_\alpha^\dagger K_\alpha = \mathbb{1}$. We can then define three classes of CPTP maps on $S_n$, representing three possible choices of free operations.
6.1. Resource theories of (speakable) coherence

**Definition 22 (Incoherent operations).** A quantum channel $\mathcal{E}$ is said to be an incoherent operation (IO) if it can be represented by Kraus operators $\{K_\alpha\}$ such that

$$D(K_\alpha(|i\rangle\langle i|)) = K_\alpha(|i\rangle\langle i|)$$

(6.4)

for all $\alpha$, and for all the elements $|i\rangle$ of the incoherent basis.

Note that if $\mathcal{E}$ is an IO,

$$\rho \in I_n \Rightarrow K_\alpha(\rho) \in \mathbb{R}_+I_n, \quad \text{for all } \alpha.$$  

(6.5)

This restriction guarantees that, even if one has access to individual measurement outcomes $\alpha$ of the instrument $\{K_\alpha\}$, one cannot generate coherent states starting from an incoherent one. Notice that equation (6.4) can be interpreted as a requirement of commutation between the decohering operation $D$ and the operation $K_\alpha$, when acting on the set of incoherent states $I_n$. One can consider a more restricted class of allowed operations by requiring the validity of such commutativity on the whole set of states $S_n$.

**Definition 23 (Strictly incoherent operations).** A quantum channel $\mathcal{E}$ is said to be a strictly incoherent operation (SIO) if it can be represented by Kraus operators $\{K_\alpha\}$ such that

$$D(K_\alpha(\rho)) = K_\alpha(D(\rho))$$

(6.6)

for all $\alpha$, and for all $\rho \in S_n$.

Both of the previous definitions can be expressed through a requirement on the form of the Kraus operators $\{K_\alpha\}$ that characterize the classes IO and SIO.

**Lemma 14 ([28]).** A quantum channel $\mathcal{E}$ is an IO if and only if it can be represented by Kraus operators $\{K_\alpha\}$ of the form

$$K_\alpha = \sum_{j=1}^n c_{\alpha,j} |f_\alpha(j)\rangle\langle j|, \quad c_{\alpha,j} \in \mathbb{C},$$

(6.7)

for some $f_\alpha: \{1,\ldots,n\} \to \{1,\ldots,n\}$. If the functions $f_\alpha$ are permutations, then $\mathcal{E}$ is a SIO.

The previous characterization of IO and SIO in terms of the Kraus operators $\{K_\alpha\}$ can in fact be taken as a definition of these classes (as in [162]).

A third class of incoherent operations has been considered, that satisfy the commutativity relation with $D$ at “global” level rather than at the level of Kraus operator representations.

**Definition 24 (Dephasing-covariant incoherent operations).** A quantum channel $\mathcal{E}$ is said to be a dephasing covariant incoherent operation (DIO) if

$$D(\mathcal{E}(\rho)) = \mathcal{E}(D(\rho))$$

(6.8)

for all $\rho \in S_n$. 

It is clear that $\text{SIO} \subset \text{IO}$ and $\text{SIO} \subset \text{DIO}$, while the classes $\text{IO}$ and $\text{DIO}$ are incomparable \(^{29, 28}\). It has been shown that transformations between pure states (i.e. rank-one projections $\psi = |\psi\rangle\langle\psi|$, with $\|\psi\| = 1$) are fully governed by the same majorization criteria \(^{44, 167, 135}\). Therefore, although the three classes $\text{IO}$, $\text{SIO}$ and $\text{DIO}$ are different from each other, they are operationally equivalent as far as pure-to-pure state transformations are concerned.

### 6.1.3 Convertibility criterion and majorization relation

For pure states $\psi = |\psi\rangle\langle\psi| \in \mathcal{S}_n$, we introduce the coherence vector

$$
\mu(\psi) := (|\psi_1|^2, \ldots, |\psi_n|^2) \in \Delta_{n-1},
$$

where $\psi_j = \langle j | \psi \rangle$, i.e. the diagonal of the density matrix $\psi$, in the (fixed) incoherent basis. Convertibility among pure states can be expressed through a majorization relation between the corresponding coherence vector, as stated in the following theorem, proved by \(^{44, 167}\) for $\text{IO}$ and $\text{SIO}$, and in \(^{135}\) for $\text{DIO}$.

**Theorem 20** \(^{44, 167, 135}\). A pure state $\psi$ can be transformed into a pure state $\psi'$ under $\text{IO}$, $\text{SIO}$ or $\text{DIO}$ if and only if $\mu(\psi) \prec \mu(\psi')$.

Using the same proof leading to Theorem 19, it is possible to prove the counterpart of this theorem for incoherent operations, as stated in the following.

**Theorem 21** \(^{167}\). For two pure states $\psi$ and $\psi'$, the maximal conversion probability under $\text{IO}$ is given by

$$
\Pi(\mu(\psi), \mu(\psi')), \quad (6.10)
$$

with $\Pi(\cdot, \cdot)$ being defined on pairs of probability vectors with $n$ nonzero components as

$$
\Pi(x, y) = \min_{1 \leq k \leq n} \frac{\sum_{j=k}^n x_j^+}{\sum_{j=k}^n y_j^+}. \quad (6.11)
$$

The theorem still holds if the class $\text{IO}$ is replaced by $\text{SIO}$ or $\text{DIO}$ as a consequence of Theorem 20.

Summing up, the three classes of incoherent operations considered are equivalent for manipulation of pure states, and they are all governed by majorization relations. Here, we will only consider pure state transformations. For simplicity we will always refer to the class $\text{IO}$, but all the results also hold for $\text{SIO}$ and $\text{DIO}$.

### 6.2 Random pure states

Let $\psi$ be a random pure state in $\mathcal{S}_n$ distributed according to the unitarily invariant measure. In the incoherent basis $\{|i\rangle\}$,

$$
\psi = \sum_{ij} \psi_{ij} |i\rangle\langle j|,
$$
6.2. Random pure states

where \((\psi_1, \psi_2, \ldots, \psi_n)\) is uniformly distributed in the \(n\)-dimensional complex unit sphere, \(\sum_j |\psi_j|^2 = 1\).

Hence, the random vector \(\mu(\psi) = (|\psi_1|^2, |\psi_2|^2, \ldots, |\psi_n|^2)\) is uniformly distributed in the simplex \(\Delta_{n-1}\),

\[
p_{\mu(\psi)}(x) = (n-1)! 1_{x \in \Delta_{n-1}}. \tag{6.12}
\]

The uniform distribution on the simplex is one of the favourite topics in geometric/integral probability \([9]\), and its relevance in quantum applications has been already highlighted in the past \([12, 91]\).

If \(\mu\) is a uniform point in \(\Delta_{n-1}\), i.e. distributed according to (6.12), then the component \(\mu_k\) is “typically” \(O(1/n)\). The extreme components lie instead on very different scales. The largest components \(\mu_{\downarrow 1}, \mu_{\downarrow 2}, \ldots\) are of size \(\log n/n\) with fluctuations of \(O(1/n)\); the smallest components \(\mu_{\downarrow n}, \mu_{\downarrow n-1}, \ldots\) are on the much smaller scale \(1/n^2\), with fluctuations of \(O(1/n^2)\). We now give a precise asymptotic descriptions of the extreme statistics of the uniform distribution on \(\Delta_{n-1}\): they are distributed as time-homogenous Markov chains with explicit (and remarkably simple) transition densities. The proof is given in Appendix B.3.

**Proposition 10.** Let \(\mu = (\mu_1, \mu_2, \ldots, \mu_n)\) be a uniform point in \(\Delta_{n-1}\). Denote by \(\mu^{\downarrow}\) the decreasing rearrangement of \(\mu\). Then, for any fixed integer \(k \geq 1\), the following hold as \(n \to \infty\):

(i) the rescaled vector of the smallest components \((n^2 \mu_{\downarrow j+1})_{1 \leq j \leq k}\) converges in distribution to \((V_1, V_2, \ldots, V_k)\), where \((V_j)_{j \geq 1}\) is a Markov chain with initial and transition densities given by

\[
\begin{align*}
  f_{V_1}(v) &= \exp(-v)1_{v \geq 0}, \\
  f_{V_{j+1}|V_j}(u|v) &= \exp(v-u)1_{u \geq v};
\end{align*}
\]

(ii) the rescaled vector of the largest components \((n\mu_{\downarrow j} - \log n)_{1 \leq j \leq k}\) converges in distribution to \((W_1, W_2, \ldots, W_k)\), where \((W_j)_{j \geq 1}\) is a Markov chain with initial and transition densities given by

\[
\begin{align*}
  f_{W_1}(w) &= \exp(-e^{-w} - w), \\
  f_{W_{j+1}|W_j}(u|w) &= \exp(e^{-w} - e^{-u} - u)1_{u \leq w}.
\end{align*}
\]

Note that, by the Markov property, we can write the joint density of \((V_1, \ldots, V_k)\),

\[
\begin{align*}
  f_{V_1, \ldots, V_k}(v_1, \ldots, v_k) &= \exp(-v_k)1_{0 \leq v_1 \leq v_2 \leq \cdots \leq v_k}, \tag{6.15}
\end{align*}
\]

and the joint density of \((W_1, \ldots, W_k)\),

\[
\begin{align*}
  f_{W_1, \ldots, W_k}(w_1, \ldots, w_k) &= \\
  &= \exp(-w_1 - w_2 - \cdots - w_k - e^{-w_k})1_{w_1 \geq w_2 \geq \cdots \geq w_k}. \tag{6.16}
\end{align*}
\]
The next Lemma gives a concrete realization of the Markov chains \((V_j)_{j \geq 1}\) and \((W_j)_{j \geq 1}\) in terms of discrete-time continuous random walks. In the following, \((X_j)_{j \geq 1}\) is a sequence of independent exponential random variables with rate 1, i.e. \(P(X_j \leq x) = 1 - \exp(-x)\).

**Lemma 15.** Let \((V_j)_{j \geq 1}\) and \((W_j)_{j \geq 1}\) be the Markov chains defined in (6.13) and (6.14), respectively. Then,

\[
(V_j)_{j \geq 1} \overset{D}{=} (X_1 + \cdots + X_j)_{j \geq 1}, \quad (6.17)
\]

\[
(W_j)_{j \geq 1} \overset{D}{=} (- \log (X_1 + \cdots + X_j))_{j \geq 1}, \quad (6.18)
\]

where \(\overset{D}{=}\) denotes equality in distribution.

See Fig. 6.1 for a pictorial illustration of Proposition 10 and Lemma 15.
6.3 Volume of the set of IO-convertible states

In Chapter 5 we analysed Nielsen’s conjecture about convertibility between bipartite pure states in the limit of large dimensions. We showed that the precise statement of the conjecture is that for two independent random points in the simplex with a distribution of random matrix type, equation (3.25), the probability that they are in majorization relation is asymptotically zero. Here we pose a similar question in the theory of coherence: is it true that ‘most’ pairs of pure \( n \)-dimensional quantum states are not IO-convertible if \( n \) is large? The answer is “Yes”.

6.3.1 Asymptotics \( n \to \infty \)

**Theorem 22.** Let \( \psi \) and \( \psi' \) be independent random pure states in \( S_n \). Then,

\[
\lim_{n \to \infty} P (\psi \prec \psi') = 0.
\]

**Proof.** We use the shorter notation \( \mu \equiv \mu(\psi) \) and \( \mu' \equiv \mu(\psi') \). It turns out to be convenient to write the majorization relation \( \mu \prec \mu' \) as

\[
\sum_{i=n-j+1}^{n} \mu_i^\downarrow \geq \sum_{i=n-j+1}^{n} \mu'_i^\downarrow, \text{ for all } j = 1, \ldots, n,
\]

by using the normalization condition \( \sum_{i=1}^{n} \mu_i^\downarrow = \sum_{i=1}^{n} \mu'_i^\downarrow \). The idea of the proof, inspired by [128], is to show that

\[
\lim_{k \to \infty} \lim_{n \to \infty} P (\mu, \mu' \text{ meet the first } k \text{ conditions in (6.19)}) = 0.
\]

From Proposition 10,

\((\mu_i^\downarrow, \mu_{i-1}^\downarrow, \ldots, \mu_{i-k+1}^\downarrow)\) is asymptotic to \((n^{-2}V_j)_{1 \leq j \leq k}\)

as \( n \to \infty \). By Lemma 15, we have the representation

\[
(V_j)_{j \geq 1} \overset{D}{=} (X_1 + \cdots + X_j)_{j \geq 1}.
\]

Analogous representation holds for the \( k \) smallest components of \( \mu' \) with their own sequence \((V'_j)_{j \geq 1} \overset{D}{=} (X'_1 + \cdots + X'_j)_{j \geq 1}\). Consider the probabilities \((1 \leq k \leq n)\)

\[
\pi_{n,k} := P \left( \sum_{i=n-j+1}^{n} \mu_i^\downarrow \geq \sum_{i=n-j+1}^{n} \mu'_i^\downarrow, \text{ for all } 1 \leq j \leq k \right).
\]

Of course,

\[
\pi_{n,n} = P (\mu \prec \mu'), \quad \text{and} \quad \pi_{n,n} \leq \pi_{n,k}.
\]

By taking the limit \( n \to \infty \) we get
Chapter 6. Resource Theory of Coherence

\[ p_k := \lim_{n \to \infty} \pi_{n,k} = P \left( \sum_{i=1}^{j} V_i \geq \sum_{i=1}^{j} V'_i, \text{for all } 1 \leq j \leq k \right). \]

It is clear that

\[ 0 \leq \limsup_{n} \pi_{n,n} \leq \lim_{k \to \infty} p_k. \]

Hence, to prove that \( \pi_{n,n} \to 0 \) as \( n \to \infty \), it is enough to show that \( p_k \to 0 \) as \( k \to \infty \). The sequence \( \tilde{V}_k := (V_k - V'_k) = \sum_{j=1}^{k} \tilde{X}_j, k \geq 1 \) is a time-discrete continuous random walk with independent steps \( \tilde{X}_j := (X_j - X'_j) \) distributed according to the two-side exponential density \((1/2) \exp(-|x|)\). The process \( I_k := \sum_{j=1}^{k} \tilde{V}_j, k \geq 1 \) is the corresponding integrated random walk (IRW). Hence, \( p_k \) is the so-called persistence probability above the origin of the IRW,

\[ p_k = P \left( \min_{1 \leq j \leq k} I_j \geq 0 \right). \quad (6.21) \]

The proof that the persistence probability of the IRW asymptotically vanishes,

\[ p_k \to 0, \quad \text{as } k \to \infty, \quad (6.22) \]

follows from the Lindeberg-Feller central limit theorem and is given in Appendix B.4. It might seem that, having to deal with i.i.d. variables \( \tilde{X}_j \), the proof that \( p_k \to 0 \) is straightforward. Note however, that the integrated random walk \( (I_k)_{k \geq 1} \) is not Markov \( (I_k \) depends on all variables \( \tilde{X}_j, j \leq k \) and this explains why some analysis is required.

The proof strategy in Theorem 22 is based on bounding \( \pi_{n,n} \) by a sequence \( p_k \) independent on \( n \), and therefore gives no information on the rate of decay of \( \pi_{n,n} \) to zero. Some insights can be obtained from the perspective of persistence probabilities as discussed in the next section.

6.3.2 Majorization, persistence probabilities and the arcsine law

We next turn our attention to the convergence rate of \( P(\psi \prec \psi') \) to 0. For two random pure states \( \psi, \psi' \) in \( S_n \), the vector \( \tilde{\delta} = (\tilde{\delta}_k)_{1 \leq k \leq n} \) with

\[ \tilde{\delta}_k = \mu(\psi')_k^\downarrow - \mu(\psi)_k^\downarrow, \quad (6.23) \]

defines a continuous random walk \( (S_k)_{0 \leq k \leq n} \), starting at \( S_0 := 0 \), with steps \( \tilde{\delta}_j \)’s,

\[ S_k := \sum_{j=1}^{k} \tilde{\delta}_j \quad 1 \leq k \leq n. \quad (6.24) \]

Note that \( S_n = 0 \) (the process is a random bridge). The majorization condition can be interpreted as the persistence (above the origin) of the random
6.3. Volume of the set of IO-convertible states

As we have seen in the previous chapter, persistence probabilities for random processes have been widely studied in statistical physics and probability. Certain exactly solvable models (that include symmetric random walks, classical random bridge, integrated random walks, etc.), and numerical study of many other models, showed that in the general case the persistence probability above the origin decays as $b n^{-\theta}$, for large $n$. It is therefore natural to expect that $P(\psi \prec \psi')$ also decays to zero as a power of $n$. In Figure 6.2 we show the results of numerical estimates of $P(\psi \prec \psi')$ (obtained from $10^6$ re-alisations of $\psi, \psi'$) for increasing values of $n$. The plot in logarithmic scale shows quite convincing evidence that

$$P(\psi \prec \psi') \sim b n^{-\theta}.$$  \hfill (6.25)

The value of the persistence exponent obtained from a numerical fit is $\theta = 0.4052$, with $\sigma_\theta = 0.0028$.

Note that, as already seen in Chapter 5 for the analogous problem of LOCC-conversions between random bipartite pure states, the process $(S_k)_{0 \leq k \leq n}$ is not Markov, and quite different from most familiar discrete-time
random processes (the steps $\tilde{\delta}_k$'s in (6.23) are neither independent and identically distributed, as in a classical random walk, nor distribution-invariant under permutations, as in a classical random bridge). One can nevertheless try to compute certain statistics related to the persistence of $S_k$, for instance the time spent above the origin. Denote this time by $N_n := \# \{k \leq n : S_k \geq 0 \}$. For classical symmetric random walks, the statistics of $N_n$ is universal, and its limit is the well-known arcsine law, see Theorem 18. Surprisingly, numerical results (see Figure 6.3) show that the fraction of time spent above 0 by $S_k$ is also asymptotically described by the arcsine law,

$$\lim_{n \to \infty} P \left( \frac{N_n}{n} \leq t \right) = \frac{2}{\pi} \arcsin \left( \sqrt{t} \right).$$

(6.26)

### 6.3.3 Limit distribution of the maximal success probability of IO-conversion

The maximal success probability of IO-conversion of state $\psi$ into $\psi'$ is $\Pi(\mu, \mu')$, where $\mu = \mu(\psi), \mu' = \mu(\psi')$ are the diagonal entries of $\psi, \psi'$ (Theorem 21). Theorem 22 can be rephrased as the statement that if $\mu, \mu'$ are independent uniform points in $\Delta_{n-1}$, then

$$P(\Pi(\mu, \mu') = 1) \to 0, \quad \text{as } n \to \infty.$$
In Chapter 5 we have seen that the scaling limit of the variable $\Pi(\lambda, \lambda')$, when $\lambda, \lambda'$ are independent spectra of fixed-trace Wishart random matrices is determined by the asymptotic fluctuations of the smallest component of random probability vectors $\lambda, \lambda'$. Translated in this setting, the precise statement is that, if for some scaling constants $a_n, b_n$, the variable

$$a_n \frac{\mu_n}{\mathbb{E}[\mu_n]} + b_n$$

has a nontrivial limit in distribution, then, with the same constants,

$$P(a_n \Pi(\mu, \mu') + b_n \leq p)$$

has a nontrivial limit as $n \to \infty$.

The smallest component $\mu_n = \delta(\psi)^n_{\downarrow}$ has probability density

$$p_n(x) = n^2(1 - nx)^{n-1}1_{0 \leq x \leq 1/n}.$$  

The average and variance of $\mu_n$ are

$$\mathbb{E}[\mu_n] = \frac{1}{n(n+1)}, \quad \text{Var}[\mu_n] = \frac{1}{n(n+1)^2(n+2)}.$$  

Hence, the fluctuations of $\mu_n$ relative to the mean are asymptotically bounded:

$$\frac{\text{Var}[\mu_n]^{1/2}}{\mathbb{E}[\mu_n]} = O(1)$$

and therefore we can take $a_n = 1$ and $b_n = 0$ in (6.27). The conjectural statement (6.28) in this case says that the distribution function

$$F_n(p) = P(\Pi(\mu, \mu') \leq p)$$

should have a scaling limit. Indeed, we found numerically that, for large $n$, the function $F_n(p)$ tends to a limit distribution, as shown in Figure 6.4.

Here we push further our previous conjecture and we propose that as $n \to \infty$ the distribution of the random variable $\Pi(\mu, \mu')$ is determined by the asymptotic behaviour of the smallest components of $\mu, \mu'$ only. Any fixed block of the order statistics $n^2\mu_{n-j+1}^\downarrow, j = 1, \ldots, k$ is asymptotic to the first $k$ components of a Poisson process $(V_j)_{j \geq 1}$; similarly, $n^2\mu_{n-j+1}^\downarrow$ is asymptotic to its own independent copy $(V'_{j'})_{j' \geq 1}$. Hence, we conjecture that $\Pi(\mu, \mu')$ is asymptotically distributed as

$$\Pi^\infty(V, V') := \min_{k \geq 1} \frac{\sum_{j=1}^k V_j}{\sum_{j=1}^k V'_{j'}}$$

where $(V_j)_{j \geq 1}$ and $(V'_{j'})_{j' \geq 1}$ are two independent copies of a Poisson process with rate 1 (i.e. point processes with independent exponential spacings). In
Chapter 6. Resource Theory of Coherence

Figure 6.4: Distribution $F_n(p)$ for various values of $n$ vs the limit distribution $F_\infty(p)$ (distribution functions computed from numerical simulations).

formulae, if we denote by $F_\infty(p) := P(\Pi_\infty(V,V') \leq p)$ the distribution function of $\Pi_\infty(V,V')$, we claim that

$$\lim_{n \to \infty} F_n(p) = F_\infty(p). \quad (6.32)$$

Figure 6.4 shows the results of numerical simulations of $\Pi(\mu,\mu')$ and $\Pi_\infty(V,V')$ on samples of $5 \cdot 10^5$ pairs of random probability vectors $\mu, \mu'$ in $\Delta_{n-1}$, and pairs of random processes $V$ and $V'$. The agreement between the corresponding distributions $F_n(p)$ for large $n$, and $F_\infty(p)$ is quite convincing of the correctness of (6.31)–(6.32).

6.4 Concluding remarks

6.4.1 Likelihood of comparability in algebraic combinatorics

In this chapter we proved that the probability that two independent random points uniformly distributed in the unit simplex are in the majorization relation is asymptotically zero as $n \to \infty$. A similar question in the discrete setting was posed back in 1979 by MacDonald [101, Ch.1.1, Ex.18]: for two integer partitions of $n$, chosen uniformly at random, and independently, is it true that the probability that they are in majorization relation (a.k.a. dominance order) is zero as $n \to \infty$? In 1999, Pittel [128] proved the positive answer to Macdonald’s question. In the proof of Theorem 22 we emulated the main ideas exposed in [128]. An interesting remark is that a simplification occurs
in the continuous setting. Pittel considered the first $k$ conditions for majorization involving the $k$ largest components of random integer partitions. They are asymptotic to Markov chains $(W_j)_{j \geq 1}$ of Lemma 15 with double-exponential spacings. In the continuous setting of random points in the simplex, it is fairly easy to obtain the asymptotics of the smallest components. Hence, here we considered instead the first $k$ conditions (6.19) involving the smallest components of $\mu^\downarrow$, and this reduces the problem to the persistence probability of an integrated random walk $I_k = \sum_{j=1}^k \sum_{i=1}^j \tilde{X}_i$, where the increments $\tilde{X}_i$ have two-sided exponential distribution. This choice makes the analysis simpler compared to the discrete setting for integer partitions.

Note also that the proof that $P(\mu \prec \mu') \to 0$ as $n \to \infty$ presented here can be adapted to the case where $\mu, \mu'$ are independent copies of points in the unit simplex picked according to a more general Dirichlet distribution

$$p_{\text{Dir}(\alpha)}(x) = \frac{\Gamma(n\alpha)}{\Gamma(\alpha)^n} \prod_{k=1}^n x_k^{\alpha-1} 1_{x \in \Delta_{n-1}}.$$ 

It would be interesting to see if this distribution appears naturally in the theory of random quantum states.

### 6.4.2 Entanglement theory and LOCC-convertibility

There is a difference between the statement proved in Theorem 22 for the theory of coherence and Nielsen’s conjecture for entanglement theory. The LOCC-convertibility criterion for bipartite systems is the majorization relation for spectra of reduced density matrices. In the random setting, those spectra are not uniformly distributed in the simplex; instead, they follow a random-matrix-type density in $\Delta_{n-1}$ given by (3.25).

In particular, for a point distributed according to $f_{n,m}$ it is no longer true that the largest/smallest components are asymptotically described by point processes with independent spacings (their statistics is given instead by scaling limits at the edges of random matrices, known as Airy and Bessel point processes), and this complicates considerably the analysis, as we have seen in Chapter 5.
Part III

Random states for Quantum Metrology
Chapter 7

Precision Scaling in Quantum Metrology

Quantum metrology is the study of making high-resolution and high-sensitive measurements using Quantum Theory. One of the core tasks in physics is the estimation of physical quantities starting from the experimental data gathered from measurements. The fundamental precision achievable for the estimation of a given parameter can be established using tools from a branch of statistics known as Estimation Theory [79, 85].

This chapter is devoted to the discussion of such fundamental limits when only classical resources are employed, and when quantum resources become available. After a brief introduction on the subject in Section 7.1, Section 7.2 will be devoted to the scaling of the fundamental precision in the number of “probes” employed for the measurement: we will see that if quantum resources are available (such as entanglement among the probes), a better scaling is achievable. Then, in Section 7.3 we will discuss the usefulness of Random Quantum States for Quantum Metrology, considering both the case of generic quantum states sampled from the whole Hilbert space and the case where they are sampled from a suitable subspace.

7.1 Quantum Estimation Theory

7.1.1 Estimators

From a mathematical viewpoint, the measurement results are modeled as a statistical sample \( x = (x_1, \ldots, x_\nu) \) following a distribution with a probability density function \( P(x|\varphi) \) which depends on the value \( \varphi \) of the parameter one wants to estimate. In the following we denote with \( \mathbb{E}_\varphi \) averages computed with respect to this distribution, in order to point out explicitly that they depend on the true value of the parameter \( \varphi \). An estimator describes a strategy for extracting an estimate of the parameter starting from the collected data, and is modeled as a function of the data \( \hat{\varphi} \equiv \hat{\varphi}(x) \). A first requirement for a good estimator is that it gives on average the true value of the parameter:

\[
\mathbb{E}_\varphi[\hat{\varphi}(x)] \equiv \int_{\mathbb{R}^\nu} \hat{\varphi}(x) P(x|\varphi) \, d^\nu x = \varphi.
\]  

(7.1)
Estimators satisfying this condition are said to be unbiased. This requirement alone is not sufficient for the estimator to be good, but it guarantees at least that it is not affected by some systematic error. A natural criterion for establishing the quality of an estimator is to minimize the mean squared error, defined as

$$\text{MSE}[\hat{\varphi}] = \mathbb{E}_\varphi[(\hat{\varphi} - \varphi)^2] = \int_{\mathbb{R}^n} (\hat{\varphi}(x) - \varphi)^2 P(x|\varphi) \, dx,$$  

which measures the mean squared deviation of the estimator from the true value. It can be seen as the sum of two contributions:

$$\text{MSE}[\hat{\varphi}] = \mathbb{E}_\varphi[(\hat{\varphi} - \mathbb{E}_\varphi[\hat{\varphi}])^2] + (\mathbb{E}_\varphi[\hat{\varphi}] - \varphi)^2$$

where $b(\varphi) = (\mathbb{E}_\varphi[\hat{\varphi}] - \varphi)$ is the bias of the estimator and

$$\text{Var}_\varphi[\hat{\varphi}] = \int_{\mathbb{R}^n} \hat{\varphi}(x)^2 P(x|\varphi) dx - \left( \int_{\mathbb{R}^n} \hat{\varphi}(x) P(x|\varphi) dx \right)^2$$

is its variance. For unbiased estimators, $b(\varphi) = 0$, the mean squared error (7.2) coincides with the variance of the estimator, and we say that the uncertainty (or error) in the estimation is $\delta \varphi$ if

$$\sqrt{\text{Var}_\varphi[\hat{\varphi}]} = \delta \varphi.$$  

In principle, we would like to find optimal estimators, e.g. estimators which minimize $\delta \varphi$. In practice, finding the minimum variance unbiased estimator (MVU) can be a very complicated task. In general, an unbiased estimator which minimizes the variance for all $\varphi$ might even not exist, as some explicit examples show [85]. Moreover, even if some MVU exists, we may not be able to find it.

Then, being able to place a lower bound on the variance of an unbiased estimator is extremely important from a practical point of view: in the best case, it can prove that a certain estimator is in fact a MVU estimator; in the worst case, it provides a benchmark against which the performance of a given estimator can be evaluated.

### 7.1.2 Cramér-Rao Bound and Fisher Information

A lower bound on the precision achievable in the estimation of the parameter $\varphi$ can be determined knowing only the probability distribution $P(x|\varphi)$ of the measurement result.
7.1. Quantum Estimation Theory

**Theorem 23 ([85], Theorem 3.1).** Let \( P(x|\varphi) \) be a probability density function satisfying the “regularity” condition

\[
\mathbb{E}_\varphi \left[ \frac{\partial \ln P(x|\varphi)}{\partial \varphi} \right] = 0 \quad \text{for all } \varphi. \tag{7.6}
\]

Then, the variance of any unbiased estimator \( \hat{\varphi} \) must satisfy

\[
\text{Var}_{\varphi}[\hat{\varphi}] \geq \frac{1}{\mathcal{F}(\varphi)}, \tag{7.7}
\]

where

\[
\mathcal{F}(\varphi) = \mathbb{E}_\varphi \left[ \left( \frac{\partial \ln P(x|\varphi)}{\partial \varphi} \right)^2 \right] = \int \text{d}x \ P(x|\varphi) \left( \frac{\partial \ln P(x|\varphi)}{\partial \varphi} \right)^2 \tag{7.8}
\]

is the Fisher Information associated with the probability density function \( P(x|\varphi) \).

The regularity condition 7.6 holds whenever the integration can be exchanged with the differentiation, since

\[
\mathbb{E}_\varphi \left[ \frac{\partial \ln P(x|\varphi)}{\partial \varphi} \right] = \int \frac{\partial \ln P(x|\varphi)}{\partial \varphi} P(x|\varphi) \text{d}x = \int \frac{\partial P(x|\varphi)}{\partial \varphi} \text{d}x = 0 \tag{7.9}
\]

from the normalization condition. This will be essentially true except for some particularly pathological cases, e.g. a probability density function with a bounded domain dependent on the parameter \( \varphi \) ([85], Problem 3.1).

If the estimation is based on multiple independent measurement of the same quantity, then the probability density function factorizes as

\[
P(x|\varphi) = \prod_{j=1}^{\nu} p(x_j|\varphi). \tag{7.10}
\]

As a consequence, the Fisher information is given by

\[
\mathcal{F}(\varphi) = \int \text{d}x \ P(x|\varphi) \left( \sum_{j=1}^{\nu} \left. \frac{\partial \ln p(x_j|\varphi)}{\partial \varphi} \right|_{x_j} \right)^2 = \sum_{j=1}^{\nu} \int \text{d}x_j \ P(x_j|\varphi) \left( \frac{\partial \ln p(x_j|\varphi)}{\partial \varphi} \right)^2, \tag{7.11}
\]

where the mixed terms vanish under the regularity assumption \( \mathbb{E}_\varphi[\partial_\varphi p(x|\varphi)] = 0 \). We see then that in the case of \( \nu \) independent measurements of the same quantity, the global Fisher information is given by

\[
\mathcal{F}(\varphi) = \nu F(\varphi), \tag{7.12}
\]

where

\[
F(\varphi) = \int \text{d}x \ p(x|\varphi) \left( \frac{\partial_\varphi \log p(x|\varphi)}{\partial \varphi} \right)^2 \tag{7.13}
\]
Chapter 7. Precision Scaling in Quantum Metrology

is the Fisher Information of each single measurement and the Cramér-Rao bound reads

$$\text{Var}_\varphi[\hat{\varphi}] \geq \frac{1}{\nu F(\varphi)}. \quad (7.14)$$

Notice that this correspond to the standard result that if $\nu$ independent measurements of the same physical quantity are performed, the best we can hope for the behaviour of the statistical error $\delta\varphi = \sqrt{\text{Var}_\varphi[\hat{\varphi}]}$ is that it decreases as the square root of the number of repetitions:

$$\delta\varphi \propto \frac{1}{\sqrt{\nu}}. \quad (7.15)$$

In the quantum scenario, in order to perform $\nu$ measurements of the same quantity one needs to encode this quantity in the same quantum probe for $\nu$ times. Alternatively, one could use $\nu$ different quantum systems each prepared in the same state to perform $\nu$ independent measurements of the quantity of interest. However, as we will see more precisely later, quantum mechanics allows a more efficient use of the available resources (in this case, the systems to be measured), which can be prepared in a collective quantum state enabling a better scaling in the number of resources employed.

**Example: Fisher Information of a Gaussian probability distribution**

Since we will need this result to derive the results of Chapter 8, let us derive here as an example the Fisher Information of a Gaussian distribution whose average $\mu_\varphi$ and variance $\sigma^2_\varphi$ depend on the parameter $\varphi$ to be determined. Hence, the probability density function reads:

$$p(x|\varphi) = \frac{1}{\sqrt{2\pi\sigma^2_\varphi}} \exp\left[ -\frac{(x - \mu_\varphi)^2}{2\sigma^2_\varphi} \right]. \quad (7.16)$$

Inserting (7.16) into (7.13), a straightforward (but lengthy) computation of the Gaussian integrals leads to:

$$F(\varphi) = \frac{1}{2} \left( \frac{\partial_\varphi \sigma^2_\varphi}{\sigma_\varphi} \right)^2 + \left( \frac{\partial_\varphi \mu_\varphi}{\sigma_\varphi} \right)^2. \quad (7.17)$$

Equation (7.17) shows that both a strong dependence of the variance $\sigma^2_\varphi$ and a strong dependence of the mean $\mu_\varphi$ on $\varphi$ can contribute to a large Fisher Information, i.e. a small fundamental error on the estimation of $\varphi$.

**7.1.3 Quantum Cramér-Rao Bound**

In the following we will use the version (7.14) of the Cramér-Rao bound, denoting with $F(\varphi)$ the Fisher information associated with each single measurement (either involving a single system or multiple systems) and $\nu$ being the number of times the same experiment is repeated. The Fisher information is completely determined by $p(x|\varphi)$, i.e. by the statistics of the measurement
results induced by the value of the parameter \( \varphi \). In quantum mechanics this statistics depends on both the state of the quantum system one is measuring and the particular measurement which is performed. The measurement is described by a Positive Operator Valued Measure (POVM), i.e. a set of operators \( \{ \Pi_x \} \) satisfying

\[
\Pi_x \geq 0 \quad \forall x, \quad \int dx \, \Pi_x = \mathbb{I}.
\]  

(7.18)

The probability distribution of the measurement results is then determined according to the Born rule:

\[
p(x|\varphi) = \text{tr}(\rho_\varphi \Pi_x),
\]

(7.19)

where \( \rho_\varphi \) is the state encoding the parameter, on which the measurement is performed.

For a given encoding of the parameter \( \varphi \) in the state of the system, i.e. for a given \( \rho_\varphi \), we can choose the measurement as we wish. In order to find the ultimate bound on precision achievable on the parameter \( \varphi \), taking into account the freedom in the measurement to perform, we need to maximize the Fisher information (7.8) over all the possible choices of \( \{ \Pi_x \} \). This is done by introducing the Symmetric Logarithmic Derivative (SLD), defined as the self-adjoint operator \( L_\varphi \) satisfying the equation:

\[
\frac{L_\varphi \rho_\varphi + \rho_\varphi L_\varphi}{2} = \frac{\partial \rho_\varphi}{\partial \varphi}.
\]

(7.20)

Then, using the fact that

\[
\partial_\varphi p(x|\varphi) = \text{tr}(\partial_\varphi \rho_\varphi \Pi_x) = \text{Re}[\text{tr}(\rho_\varphi \Pi_x L_\varphi)],
\]

(7.21)

the Fisher Information (7.8) can be rewritten as

\[
F(\varphi) = \int dx \frac{\text{Re}[\text{tr}(\rho_\varphi \Pi_x L_\varphi)]^2}{\text{tr}(\rho_\varphi \Pi_x)}.
\]

(7.22)

A (saturable) upper bound can be found by using the following inequalities:

\[
F(\varphi) \leq \int dx \left| \frac{\text{tr}(\rho_\varphi \Pi_x L_\varphi)}{\sqrt{\text{tr}(\rho_\varphi \Pi_x)}} \right|^2 \leq \int dx \left| \frac{\text{tr}(\rho_\varphi \Pi_x L_\varphi)}{\sqrt{\text{tr}(\rho_\varphi \Pi_x)}} \right|^2 \leq \int dx \text{tr}(\rho_\varphi L_\varphi^2 \Pi_x^2 L_\varphi) = \text{tr}(\rho_\varphi L_\varphi^2).
\]

(7.23)

The first bound is saturated when \( \text{tr}(\rho_\varphi \Pi_x L_\varphi) \) is real for every \( \varphi \). The second bound is based on the Schwartz inequality \( |\text{tr}(A^\dagger B)|^2 \leq \text{tr}(A^\dagger A)\text{tr}(B^\dagger B) \) and is saturated when

\[
\frac{\sqrt{\Pi_x \rho_\varphi}}{\sqrt{\text{tr}(\rho_\varphi \Pi_x)}} = \frac{\sqrt{\Pi_x L_\varphi \rho_\varphi \Pi_x L_\varphi}}{\sqrt{\text{tr}(\rho_\varphi L_\varphi \Pi_x L_\varphi)}}
\]

(7.24)
for each $\varphi$. The condition (7.24) is satisfied when the POVM chosen to perform the estimation is made by the set of projectors over the eigenstates of $L_\varphi$. To sum up, we have the following bound:

$$F(\varphi) \leq \text{tr}(\rho_\varphi L^2_\varphi) =: Q(\rho_\varphi),$$

(7.25)

where $Q(\rho_\varphi)$, called Quantum Fisher Information, represents the Fisher information associated to the optimal measurement. Inserting the bound (7.25) into (7.14) we obtain the ultimate bound on the precision $\delta\varphi = \text{Var}_\varphi[\hat{\varphi}]^{1/2}$ attainable on the estimation of $\varphi$:

$$\delta\varphi \geq \frac{1}{\sqrt{\nu Q(\rho_\varphi)}}. \quad (7.26)$$

Two important properties of the QFI are additivity on product states and convexity, meaning respectively that

$$Q(\rho_\varphi^\otimes N) = NQ(\rho_\varphi) \quad (7.27)$$

and that

$$Q\left( \sum_j p_j \rho_\varphi^{(j)} \right) \leq \sum_j p_j Q(\rho_\varphi^{(j)}), \quad (7.28)$$

i.e. mixing states can never increase their parameter sensitivity above their average sensitivity, which implies that pure states are the optimal ones for the purposes of parameter estimation.

### 7.1.4 Quantum Fisher Information for parameter independent Hamiltonian evolution

In this thesis we will focus on a unitary encoding of the parameter $\varphi$, i.e. when the information on the parameter is imprinted on some initial quantum state $\rho$ via a unitary transformation

$$\rho \mapsto \rho_\varphi = U_\varphi \rho U_\varphi^\dagger. \quad (7.29)$$

In particular, we will focus in this chapter on the case of a unitary with a parameter-independent generator $H$, i.e. $U_\varphi = e^{-i\varphi H}$. One can see from definition (7.25) that the QFI $Q(\varphi)$ depends only on $\rho_\varphi$ and $\partial_\varphi \rho_\varphi$, which translates into a dependence on $\rho$ and $H$ for the particular case under consideration here, where $\rho_\varphi = e^{-i\varphi H} \rho e^{i\varphi H}$. A standard calculation [54, 79] shows that in this case the QFI can be written in terms of the spectral decomposition $\rho = \sum_{j=1}^n p_j |e_j\rangle\langle e_j|$ as

$$Q(\rho, H) = 2 \sum_{j,k: p_j + p_k \neq 0} \frac{(p_j - p_k)^2}{p_j + p_k} |\langle e_j | H | e_k \rangle|^2, \quad (7.30)$$
which for pure states, \( \rho = |\psi\rangle\langle \psi | \equiv \psi \) simplifies to the variance of \( H \), that is:

\[
Q(\psi, H) = 4[\text{tr}(\psi H^2) - \text{tr}(\psi H)^2] = 4\langle (\Delta H)^2 \rangle_{\psi}
\]  

(7.31)

### 7.2 Standard Quantum Limit and Heisenberg Limit

The precision achievable in a measurement when all experimental noise sources are minimized is ultimately determined by the discreteness of all physical phenomena: electronic devices will suffer the discreteness of the electric charge, whereas the quantum nature of light will affect optical devices. Due to this quantum noise, the error in the estimation of a physical parameter \( \varphi \) through a measurement employing \( N \) probes (e.g. photons, electrons) is strongly limited by the so-called “shot noise” factor of \( 1/\sqrt{N} \). This is the uncertainty due to the fluctuation of Poissonian statistics of the particles involved. One can see that this scaling limit is fundamental if the \( N \) probes are employed independently by looking at the Cramer-Rao bound (7.14): if we perform \( \nu = N \) measurements each employing a single probe, we immediately see that

\[
\delta \varphi \geq \frac{1}{\sqrt{N F(\varphi)}},
\]  

(7.32)

where \( \delta \varphi \equiv \sqrt{\text{Var}_{\psi}[\varphi]} \) and \( F(\varphi) \) in this case is the Fisher information related to a single measurement involving a single probe, clearly independent of \( N \).

It turns out that when \( N \) probes are coherently employed, the ultimate precision scaling can improve by a factor of \( \sqrt{N} \), reaching the so-called Heisenberg limit (HL) [63, 64, 43, 62, 42, 18, 161, 100]. To see explicitly how this improved scaling arises, it is better to look at an explicit scheme (which was studied in [64]), and analyze this scheme using the formalism of Quantum Fisher Information presented in the previous section.

Let us consider a situation where the information on the parameter \( \varphi \) is encoded in a quantum state of a single probe by a parameter-dependent unitary operation \( u_\varphi \in U(\mathcal{H}) \), with \( \mathcal{H} \) denoting the Hilbert space of the single probe. Since we want to determine the ultimate precision scaling when \( N \) probes are employed, we will consider the evolution \( U_\varphi = u_\varphi^\otimes N \in U(\mathcal{H}^\otimes N) \) acting on the collective system of the \( N \) probes. We will focus here on the case where \( u_\varphi = e^{-i\varphi h} \), where \( h \) is the single-particle Hamiltonian, acting on the corresponding single-particle Hilbert space \( \mathcal{H} \). Then, \( U_\varphi = e^{-i\varphi H} \), with

\[
H = \sum_{j=1}^{N} h_j, \quad h_j = \mathbb{I}^\otimes j-1 \otimes h \otimes \mathbb{I}^\otimes N-j
\]  

(7.33)

representing the Hamiltonian of the evolution of \( N \) non-interacting particles. For a pure quantum state \( \psi = |\psi\rangle\langle \psi | \), we know that the QFI is given by the variance of \( H \), equation (7.31). We will now consider the two cases when the \( N \) probes are used independently, i.e. in a separable state, and the case where they are used coherently, i.e. in an entangled state.
7.2.1 Separable case

Let us denote with \( |\varepsilon_{\text{min}}\rangle, |\varepsilon_{\text{max}}\rangle \in \mathcal{H} \) the eigenvectors of \( h \) corresponding to the minimum and maximum eigenvalue respectively. As a probe state, we consider
\[
|\psi\rangle = |\chi\rangle^\otimes N = \frac{1}{\sqrt{2}}(|\varepsilon_{\text{min}}\rangle + |\varepsilon_{\text{max}}\rangle).
\] (7.34)

Using the additivity of the QFI for separable states, equation (7.27), we have
\[
Q(\psi, H) = NQ(\chi, h) = 4N<\Delta h>^2_\chi = N(\varepsilon_{\text{max}} - \varepsilon_{\text{min}})^2
\] (7.35)

Hence, the Quantum Cramer-Rao bound (7.26) in this case reads
\[
\delta \varphi \geq \frac{1}{\sqrt{Q(\psi, H)}} = \frac{1}{\sqrt{N}} \frac{1}{\varepsilon_{\text{max}} - \varepsilon_{\text{min}}},
\] (7.36)

recovering the standard quantum limit scaling \( 1/\sqrt{N} \) expected when the \( N \) probes are employed independently from each other. Note that this is the same result which would be obtained by repeating \( \nu = N \) times the experiment using a single probe which is prepared always in the state \( |\chi\rangle \). Note also that the state \( |\chi\rangle \) is the one maximizing the QFI for each single probe, since it is the state with maximum variance among the pure states, and by convexity we know that the QFI is maximized for pure states. To summarize, if we insist on using only separable states for the estimation of \( \varphi \) in this scenario, the best precision achievable is given by (7.36).

7.2.2 Entangled case

The state maximizing the variance of the global Hamiltonian \( H \) in equation (7.33) is:
\[
|\psi_{\text{opt}}\rangle = \frac{1}{\sqrt{2}}(|\varepsilon_{\text{min}}\rangle^\otimes N + |\varepsilon_{\text{max}}\rangle^\otimes N),
\] (7.37)

and one can easily see that in such a case
\[
Q(\psi_{\text{opt}}, H) = 4<(\Delta H)^2>_{\psi_{\text{opt}}} = N^2(\varepsilon_{\text{max}} - \varepsilon_{\text{max}})^2,
\] (7.38)

which yields an ultimate bound on precision corresponding to the initial state \( |\psi\rangle_{\text{opt}} \)
\[
\delta \varphi \geq \frac{1}{\sqrt{Q(\psi_{\text{opt}}, H)}} = \frac{1}{N} \frac{1}{\varepsilon_{\text{max}} - \varepsilon_{\text{max}}},
\] (7.39)

i.e. the preannounced Heisenberg scaling \( (\delta \varphi)_{\text{opt}} \propto 1/N \).

Note that in order to obtain the \( N^2 \) scaling of the QFI in (7.38) we employed a state which is entangled. Using the additivity and the convexity of Quantum Fisher Information it is straightforward to conclude that entanglement among the \( N \) probes is necessary if one wants to beat the standard quantum limit. Based on this remark, one may think that entanglement here is the ingredient which plays the essential role. However, note that although the state (7.37) shows entanglement among all the \( N \) probes, the amount of
entanglement among each pair of probes can be relatively small: if the dimension \( d = \text{dim} \mathcal{H} \) of the local Hilbert space pertaining to each probe is greater than 2, the state (7.37) is not maximally entangled; actually, if \( d \) is much greater than 2, it is quite far from maximally entangled states. Furthermore, we know from results discussed in Chapter 3 that random states are typically highly entangled; if entanglement was the essential feature to obtain Heisenberg scaling, one would expect that random quantum states of \( N \) probes would typically yield the Heisenberg scaling. We will momentarily see that this is not the case, and that the essential property giving the Heisenberg scaling for the state (7.37) is instead its bosonic structure.

### 7.3 Typical Sensitivity of Random Quantum States

We will now use some of the concentration results discussed in Chapter 2 to show that typical random states are not useful for quantum metrology, in the sense that they typically lead to a shot-noise limit precision when used for quantum estimation. We will focus here on pure random states, i.e. \( \psi_U = |\psi_U \rangle \langle \psi_U |, |\psi_U \rangle = U |\psi_0 \rangle \), where \( U \in \mathcal{U}(\mathcal{H} \otimes N) \) is a Haar unitary and the choice of \( |\psi_0 \rangle \in \mathcal{H} \otimes N \) is irrelevant due to the unitary invariance of the Haar measure. The Fisher information reads

\[
Q(\psi_U, H) = 4[\text{tr}(\psi_U H^2) - \text{tr}(\psi_U H)^2] = 4\langle (\Delta H)^2 \rangle_U, \tag{7.40}
\]

and note that \( Q(\psi_U, H) = Q(\psi_0, U^\dagger H U) \), which implies that the results obtained as typical sensitivity for random states can also be interpreted as typical sensitivity for a fixed state with random isospectral Hamiltonians. Equation (7.40) tells us that the QFI for random pure states is essentially given by the variance of the observable \( H \) on the random one-dimensional subspace spanned by \( |\psi_U \rangle \). We can apply to this case the concentration results presented in Chapter 2. The average \( \mathbb{E}[Q(\psi_U, H)] \) can be computed using (2.59) with \( A = H \) and \( r = 1 \):

\[
\mathbb{E}[Q(\psi_U, H)] = 4 \mathbb{E}[\langle (\Delta H)^2 \rangle_U] = \frac{4}{(n + 1)} \left[ \text{tr}(H^2) - \frac{\text{tr}(H)^2}{n} \right], \tag{7.41}
\]

where \( n = \text{dim}(\mathcal{H} \otimes N) = d^N \). Taking into account the structure (7.33) of \( H \), we get that:

\[
\begin{align*}
\text{tr}(H) &= \sum_{j=1}^{N} \text{tr}(H_j) = \sum_{j=1}^{N} d^{N-1} \text{tr}(h) \\
&= N d^{N-1} \text{tr}(h), \tag{7.42}
\end{align*}
\]
and

\[
\text{tr}(H^2) = \sum_{j,k=1}^{N} \text{tr}(H_j H_k) = \sum_{j=1}^{N} \text{tr}(H_j^2) + 2 \sum_{j<k} \text{tr}(H_j H_k)
\]

\[
= \sum_{j=1}^{N} \text{tr}(I_j^{j-1} \otimes h^2 \otimes I_j^{N-j}) + 2 \sum_{j<k} \text{tr}(I_j^{j-1} \otimes h \otimes I_j^{k-1} \otimes h \otimes I_j^{N-k})
\]

\[
= N d^{N-1} \text{tr}(h^2) + N(N-1)d^{-2} \text{tr}(h)^2
\]

\[
= N d^{N-1} \left[ \text{tr}(h^2) + \frac{N-1}{d} \text{tr}(h)^2 \right], \tag{7.43}
\]

which inserted into (7.41) yields

\[
\mathbb{E}[Q(\psi_U, H)] = 4N \left( \frac{1 + \frac{1}{d^4}}{d} \right) \left( \frac{\text{tr}(h^2)}{d} - \frac{\text{tr}(h)^2}{d^2} \right), \tag{7.44}
\]

which scales linearly in the number \( N \) of probes employed for the measurement, hence indicating a shot-noise scaling precision on average. As a consequence of the concentration of measure for the variance of a quantum observable, equation (2.81), we can see that the QFI displays a shot-noise scaling not only on average but also typically. More precisely, (2.81) for the particular case \( A = H \) and \( r = 1 \) reads:

\[
\mu_H \left( \left( \langle (\Delta H)^2 \rangle_U - \frac{N}{1 + \frac{1}{d^4}} \left( \frac{\text{tr}(h^2)}{d} - \frac{\text{tr}(h)^2}{d^2} \right) \right) \geq \varepsilon \right) \leq 2 \exp \left( -\frac{d^N}{432 \|H\|^4} \varepsilon^2 \right) \tag{7.45}
\]

Taking into account the structure of \( H \), given by (7.33), we have \( \|H\| = N \|h\| \). Then, using \( Q(\psi_U, H) = 4\langle (\Delta H)^2 \rangle_U \) we get the concentration result for the QFI expressed by:

\[
\mu_H \left( \left| Q(\psi_U, H) - \frac{4N}{1 + \frac{1}{d^4}} \left( \frac{\text{tr}(h^2)}{d} - \frac{\text{tr}(h)^2}{d^2} \right) \right| \geq \varepsilon \right) \leq 2 \exp \left( -\frac{d^N}{C N^4} \varepsilon^2 \right), \tag{7.46}
\]

where \( C = 6912 \|h\|^4 \). Equation (7.46) tells us that the QFI for random pure states typically scales linearly in the number \( N \) of the probes. Strikingly, we can see from (7.46) that the concentration result holds even without considering deviations proportional to \( N \), which would be the scale of the average. In other words, although if the average is of order \( O(N) \), even deviations of order \( O(1) \) are suppressed for a sufficiently large number of probes \( N \), thanks to exponential dependence on \( N \) of the argument of the exponential in the right hand-side of (7.46). Actually, equation (7.46) implies that any deviation \( \varepsilon = O(N^{-\alpha}) \), with \( \alpha > 0 \), is suppressed for large \( N \).

To sum up, typical pure states of \( \mathcal{H}^\otimes N \) are not useful for quantum metrology, since they yield only a shot-noise limit precision, i.e. \( \delta\phi \sim 1/\sqrt{N} \). Due to the convexity of the QFI, one could expect that for mixed quantum states the scaling behaviour can only get worse.

In [123] it was shown that this remains true even if one allows for local
unitary optimization of the input state. By local unitary optimization it is meant that given a quantum state $\rho$, we are allowed to optimize over all the possible local unitary rotations $V\rho V^\dagger$, with $V = V_1 \otimes \cdots \otimes V_N$ and $V_j$ being a unitary acting on the single probe Hilbert space $\mathcal{H}$. This amounts to consider

$$Q^{LU}(\rho, H) := \sup_{V \in LU} Q(V\rho V^\dagger, H), \quad (7.47)$$

where we introduced the set of local unitaries

$$LU := \{V_1 \otimes \cdots \otimes V_N : V_j \in U(\mathcal{H})\}. \quad (7.48)$$

We will report here the result of [123] for completeness. This result was obtained considering unitaries from the special unitary group

$$SU(\mathcal{H}^{\otimes N}) = \{U \in U(\mathcal{H}^{\otimes N}) : \det(U) = 1\}, \quad (7.49)$$

rather that $U(\mathcal{H}^{\otimes N})$.

**Theorem 24** ([123], Theorem 5). Fix a single particle Hamiltonian $h$, local dimension $d = \dim \mathcal{H}$, a pure state $\psi_0 \in S(\mathcal{H}^{\otimes N})$ and denote with $\psi_U = U\psi_0 U^\dagger$, where $U \in SU(\mathcal{H}^{\otimes N})$, the special unitary group. Then, for every $\varepsilon > 0$:

$$\nu_H \left( Q^{LU}(\psi_U, H) \geq 4N\|h\|^2 \left( 1 + \frac{(N-1)d^2}{\sqrt{dN}} \right) + \varepsilon \right) \leq \exp \left( -\frac{d^N}{4096\|h\|^4 N^4 \varepsilon^2} \right), \quad (7.50)$$

and

$$\nu_H \left( Q^{LU}(\psi_U, H) \leq 4N \frac{\text{tr}(h^2)d^N}{d(d^N+1)} - \varepsilon \right) \leq \exp \left( -\frac{d^N}{4096\|h\|^4 N^4 \varepsilon^2} \right), \quad (7.51)$$

where $\nu_H$ denotes the uniform (Haar) measure on $SU(\mathcal{H}^{\otimes N})$.

Introducing the notation $f(N) = \Theta(g(N))$ to express the fact that $f$ and $g$ have the same asymptotic behaviour as $N \to \infty$, i.e:

$$f(N) = \Theta(g(N)) \iff \lim_{N \to \infty} \frac{f(N)}{g(N)} = a, \quad 0 < a < \infty, \quad (7.52)$$

we can set $\varepsilon = 2N\|h\|^2 \left( 1 + \frac{(N-1)d^2}{\sqrt{dN}} \right)$ in (7.50) and $\varepsilon = \frac{2N\text{tr}(h^2)d^N}{d(d^N+1)}$ in (7.51) to write that:

$$\nu_H \left( Q^{LU}(\psi_U, H) = \Theta(N) \right) \geq 1 - 2 \exp \left( -\Theta \left( \frac{d^N}{N^2} \right) \right). \quad (7.53)$$

This inequality expresses formally the fact that for random pure states locally optimized for the metrological task considered, the probability that the optimal precision achievable is not at the Shot-Noise Limit (either better or worse) is exponentially suppressed as $\exp\left( -c\frac{d^N}{N^2} \right)$. 
7.4 Typical Sensitivity of Random Symmetric States

The results discussed in the previous section are about the lack of usefulness of generic random states. In this section we discuss instead a “positive” result, which was derived in [123], about the usefulness of random states selected from a particular subspace of the $N$-particle Hilbert space $\mathcal{H}^\otimes N$, namely the symmetric subspace, defined as:

$$\mathcal{S}_N := \text{Span}\{\psi^\otimes N : \psi \in \mathcal{H}\},$$

(7.54)

which is of dimension

$$\text{dim}(\mathcal{S}_N) = \binom{N + d - 1}{N} = \Theta(N^{d-1}).$$

(7.55)

Here, by “usefulness” of random states for quantum metrology we mean that they typically yield a Heisenberg Scaling QFI, i.e. $Q(\rho, H) = \Theta(N^2)$, implying that an optimal measurement allows to obtain a Heisenberg scaling precision on the estimation of the parameter $\varphi$, i.e. $\delta\varphi = \Theta\left(\frac{1}{N}\right)$. Note that the subspace $\mathcal{S}_N$ contains states which are known to be useful for quantum metrology, such as the optimal state of (7.37), but also other states which naturally appear in experimental setups employing photons and bosonic atoms (whose states need to be from the symmetric subspace).

7.4.1 A particular case

Before discussing the general result, it is instructive to look at a simple example of states belonging to $\mathcal{S}_N$, and to see how they allow to achieve the Heisenberg scaling. Let us consider for example a GHZ state, which is in the same form of the optimal state (7.37):

$$|\psi\rangle = \frac{1}{\sqrt{2}} \left(|\psi_1\rangle^\otimes N + |\psi_2\rangle^\otimes N\right),$$

(7.56)

where $|\psi_1\rangle, |\psi_2\rangle \in \mathcal{H}$ are arbitrary orthogonal states. The variance of the Hamiltonian (7.33) on states in the form (7.56) reads:

$$\langle (\Delta H)^2 \rangle_\psi = \frac{1}{2} \left[ \langle \psi_1^\otimes N | H^2 | \psi_1^\otimes N \rangle + \langle \psi_1^\otimes N | H^2 | \psi_2^\otimes N \rangle + \langle \psi_2^\otimes N | H^2 | \psi_1^\otimes N \rangle + \langle \psi_2^\otimes N | H^2 | \psi_2^\otimes N \rangle \right]$$

$$- \frac{1}{4} \left[ \langle \psi_1^\otimes N | H | \psi_1^\otimes N \rangle + \langle \psi_1^\otimes N | H | \psi_2^\otimes N \rangle + \langle \psi_2^\otimes N | H | \psi_1^\otimes N \rangle + \langle \psi_2^\otimes N | H | \psi_2^\otimes N \rangle \right]^2.$$  

(7.57)
Now, using the fact that for any $|\phi\rangle, |\chi\rangle \in \mathcal{H}$ (not necessarily distinct):

$$
\langle \phi^\otimes N | H | \chi^\otimes N \rangle = \sum_{j=1}^{N} \langle \psi^\otimes j \otimes \mathbb{I}_N^\otimes j \otimes H \otimes \mathbb{I}_N^\otimes N-j | \phi^\otimes N \rangle = 0
$$

$$= N \langle \phi | \chi \rangle^{N-1} \langle \psi | h | \phi \rangle ,
$$

(7.58)

and

$$
\langle \phi^\otimes N | H^2 | \chi^\otimes N \rangle = \sum_{j=1}^{N} \langle \phi^\otimes N | \mathbb{I}_N^\otimes j \otimes h^2 \otimes \mathbb{I}_N^\otimes N-j | \chi^\otimes N \rangle
$$

$$+ 2 \sum_{j<k} \langle \phi^\otimes N | \mathbb{I}_N^\otimes j \otimes h \otimes \mathbb{I}_N^\otimes k-j \otimes h \otimes \mathbb{I}_N^\otimes N-k | \chi^\otimes N \rangle
$$

$$= N \langle \phi | \chi \rangle^{N-1} \langle \psi | h^2 | \phi \rangle + N(N-1) \langle \phi | \chi \rangle^{N-2} \langle \phi | h | \chi \rangle^2 ,
$$

(7.59)

into equation (7.57), we obtain that for orthonormal states $|\psi_1\rangle$ and $|\psi_2\rangle$:

$$
\langle (\Delta H)^2 \rangle_\psi = \frac{N^2}{4} \left[ \langle h | \psi_1 \rangle - \langle h | \psi_2 \rangle \right]^2 + N \left[ (\Delta h)_{\psi_1}^2 + (\Delta h)_{\psi_2}^2 \right].
$$

(6.60)

The Quantum Fisher Information associated with the state (7.56) is then given by:

$$Q(\psi, H) = N^2 \left[ \langle h | \psi_1 \rangle - \langle h | \psi_2 \rangle \right]^2 + 2N \left[ (\Delta h)_{\psi_1}^2 + (\Delta h)_{\psi_2}^2 \right].
$$

(6.61)

Equation (6.61) shows the presence of a quadratic term which is responsible for the Heisenberg scaling, and vanishes when $|\psi_1\rangle$ and $|\psi_2\rangle$ give the same expected value for $h$, and a linear term, which is proportional to the variance of $h$ on $|\psi_1\rangle$ and $|\psi_2\rangle$, and vanishes if these are eigenstates of $h$ (as in (7.37)).

At this point it begins to become more clear what kind of states one needs to employ if one wants to reach the Heisenberg scaling. For example, one could choose the states $|\psi_1\rangle$ and $|\psi_2\rangle$ randomly from the single particle Hilbert space $\mathcal{H}$, and get the Heisenberg scaling. In the next section we will see that this is the case for a more general class of states, which include states in the form (7.56) as a particular case.

### 7.4.2 Usefulness of random symmetric states for quantum metrology

We now turn to the general result that random symmetric states (even mixed ones) are useful for quantum metrology, i.e. they typically yield Heisenberg scaling.

Since this result holds even if the states considered are mixed, it is convenient to fix the spectrum when considering random states: hence, we will talk of isospectral random states, meaning that we consider a density matrix $\rho_0 \in \mathcal{S}_N$ with fixed spectrum $(\lambda_1, \ldots, \lambda_D)$, where $D := \dim \mathcal{S}_N$, and then take the random states obtained as

$$\rho_U = U \rho_0 U^\dagger ,
$$

(7.62)
with $U \in \mathcal{SU}(\mathcal{S}_N)$ sampled according to the Haar measure on $\mathcal{SU}(\mathcal{S}_N)$. It is clear that mixed states will always have worse performances than pure states. In particular, the maximally mixed state over $\mathcal{S}_N$, i.e. $\rho_0 = \sigma_{\text{mix}} = \mathbb{I}_{\mathcal{S}_N}/\dim \mathcal{S}_N$, yields a vanishing QFI for all choices of $U \in \mathcal{SU}(\mathcal{S}_N)$. Let us first look at the average QFI for random states in the symmetric subspace, and later we will turn to its typicality.

**Lemma 16** ([123], Lemma 8). Let $\rho_0 \in \mathcal{S}(\mathcal{S}_N)$ be fixed, and denote with $(\lambda_1, \ldots, \lambda_D)$ its eigenvalues ($D = \dim \mathcal{S}_N$). Let $H$ be an Hamiltonian as in (7.33), and assume for simplicity that $\text{tr}(h) = 0$. Then, for random Haar unitaries $U \in \mathcal{SU}(\mathcal{S}_N)$:

$$
E[Q(U \rho_0 U^\dagger, H)] = \frac{4N(N+d)\text{tr}(h^2)}{d(d+1)} \frac{D}{D+1} \Lambda(\lambda_1, \ldots, \lambda_D),
$$

(7.63)

where the average is computed with respect to the Haar measure on $\mathcal{SU}(\mathcal{S}_N)$ and

$$
\Lambda(\lambda_1, \ldots, \lambda_D) = \sum_{i,j: \lambda_i + \lambda_j \neq 0} \frac{(\lambda_i - \lambda_j)^2}{\lambda_i + \lambda_j}.
$$

(7.64)

**Remark 3.** Note that $\Lambda = 1$ for pure states and $\Lambda = 0$ for the maximally mixed state $\sigma_{\text{mix}}$. For states $\rho_0 \neq \sigma_{\text{mix}}$ it is always non-vanishing and it is bounded from below as [123]:

$$
\Lambda(\lambda_1, \ldots, \lambda_D) \geq \frac{D}{D-1} \frac{d_B(\rho_0, \sigma_{\text{mix}})^2}{2},
$$

(7.65)

where

$$
d_B(\rho, \sigma) := \sqrt{2[1 - \Phi(\rho, \sigma)]}
$$

(7.66)

is the Bures distance between $\rho$ and $\sigma$, and $\Phi(\rho, \sigma) := \text{tr}(\sqrt{\rho^{1/2} \sigma \rho^{1/2}})$ is the fidelity.

Using equation (7.63) we can then claim that on average, for random isospectral states on the symmetric subspace $\mathcal{S}_N$ of $\mathcal{H}^\otimes N$, there exists an optimal measurement allowing for an estimation of the parameter $\varphi$ at the Heisenberg limit. However, as usual, it is not sufficient to know only what happens on average: we need to know also what happens typically.

**Theorem 25** ([123], Theorem 6). With the same notation of Lemma 16, and for every $\varepsilon > 0$:

$$
\nu_H \left( Q(U \rho_0 U^\dagger, H) \leq d_B(\rho_0, \sigma_{\text{mix}})^2 \frac{D^2}{D^2 - 1} \frac{2N(N+d)\text{tr}(h^2)}{d(d+1)} + \varepsilon \right) 
\leq \exp \left( -\frac{D\varepsilon^2}{4096C\|h\|_4^4 N^4} \right)
$$

(7.67)

where $C = \min\{1, 8d_B(\rho_0, \sigma_{\text{mix}})\}$, $D = \dim(\mathcal{S}_N) = \binom{N+d-1}{N} = \Theta(N^{d-1})$ and $\nu_H$ denotes the Haar measure on $\mathcal{SU}(\mathcal{S}_N)$.

An inspection of the right-hand side of (7.67) reveals that deviations $\varepsilon = \Theta(N^2)$ are suppressed for $d > 1$. In fact, for large $N$, the $N$-dependent factor
7.4. Typical Sensitivity of Random Symmetric States

in the exponential is given by

\[
\frac{D}{N^4} = \frac{1}{N^3} \left( \frac{N + d - 1}{N} \right) \simeq \frac{1}{(d - 1)!} \frac{N^{d-1}}{N^4},
\]

which implies that the right-hand side of (7.67) becomes vanishingly small when \( \varepsilon = \Theta(N^2) \) for \( d > 1 \). In particular, by setting

\[
\varepsilon = d_B(\rho_0, \sigma_{\text{mix}}) 2 \frac{N(N + d)\text{tr}(h^2)}{d(d + 1)} \frac{D^2}{D^2 - 1}
\]

we obtain the desired result that the probability of getting \( Q(U\rho_0 U^\dagger, H) \) smaller than a quantity of order \( N^2 \) is exponentially suppressed. Formally, this is expressed by:

\[
\nu_H \left( Q(U\rho_0 U^\dagger, H) \right) \leq d_B(\rho_0, \sigma_{\text{mix}}) \Theta(N^2) \leq \exp \left( -d_B(\rho_0, \sigma_{\text{mix}})^3 \Theta(N^{d-1}) \right)
\]

(7.70)

In other words, we obtain the important result that random states from the symmetric subspace \( \mathcal{S}_N \) of \( \mathcal{H}^\otimes N \) are useful for quantum metrology, in the sense that they typically yields a Quantum Fisher Information which scales quadratically in the number of probes \( N \), which in turn implies that it is possible for typical random symmetric states to find an optimal measurement leading to a Heisenberg scaling precision on the estimation of the parameter \( \varphi \). Remarkably, while the negative result in the previous section was not improved by local unitary optimization, the usefulness of random symmetric states discussed here does not even require local unitary optimization.
Chapter 8

Distributed Quantum Metrology with Gaussian States

8.1 Introduction

Metrological schemes allowing to reach a Heisenberg limited sensitivity have been largely studied in the literature. However, these protocols are usually difficult to implement experimentally due to the convoluted and challenging measurement procedures [PhysRevLett.72.3439, 78, 140] and the fragile quantum coherence needed in the input states [64, 40, 154, 92]. Gaussian states, on the other hand, provide a promising avenue for quantum optical technologies [160, 2], since they are easier to create and manipulate experimentally compared to non-Gaussian ones, such as Fock states. Moreover, they allow a complete analytical treatment from a theoretical point of view [54, 160, 2]. In particular, the squeezing of a Gaussian state, which allows for highly reduced-noise signals, appears to be a valuable tool to reach quantum super-sensitive precision [100]. From a metrological perspective, squeezed states are often used along with Gaussian measurements [110, 106, 121], defined as measurement schemes producing a Gaussian probability distribution of the outcomes for any Gaussian state [160]. Homodyne and heterodyne detection represent paradigmatic examples of Gaussian measurements. It has been shown, both theoretically [163] and experimentally [6], that an adaptive homodyne phase estimation performs better than heterodyne detection, and approaches closer to the intrinsic quantum uncertainty than any previous technique when no prior knowledge of the phase is given. The importance of feedback and adaptivity in quantum estimation protocols has been underlined also in subsequent works [110, 8]. Adaptiveness can be avoided in an optimal protocol (or near optimal) only if some constraint in the range of variation of the parameter is given [57, 58, 149].

Within the domain of quantum optics, photons are sent as probes through an interferometer where a parameter $\varphi$ to be estimated is encoded. The information about the parameter is imprinted then in the output state of the photons, and it can be extracted by a suitable measurement. The situation which has been often considered is the case where $\varphi$ is an optical phase [18, 63, 110, 43, 121] or a phase-like parameter [64, 62]. These results clearly apply also to situations in which other quantities of interest (e.g. a distance) can be
converted into an optical phase \[43\], but they fail to cover more general situations (e.g. the unknown parameter is distributed among several components of the interferometer). In a recent work \[106\], where this generic situation has been addressed, the presence of the unknown parameter \(\varphi\) distributed in multiple modes introduces a non-trivial complication, since it appears that a simultaneous adaptive procedure both in the input probe and in the measurement is needed in order to reach the HL, making the whole scheme quite unfeasible from a practical point of view. Furthermore, the proposed scheme requires an unquantified precision and number of resources in the adaptive procedure.

In this chapter we describe a metrological scheme where all these serious drawbacks are overcome. In particular, we consider a general scenario in which \(\varphi\) can be any parameter embedded into a linear passive \(M\)-modes interferometer: it can be a parameter characterizing any specific component of the interferometer, or distributed among different components of the circuit. We will see that when suitable requirements are satisfied, the Fisher Information associated with the metrological scheme scales quadratically in the (mean) number of photons employed, hence allowing a Heisenberg scaling precision on the estimation of \(\varphi\). After discussing the requirements needed and how they can be achieved, we will examine the constant of proportionality (pre-factor) in the scaling, which depends on how an arbitrary unitary stage is chosen. Using results of concentration of measure in high-dimensional spaces, we will show that distributing the unknown parameter among an high number of modes \(M\) allows this pre-factor to typically take non-vanishing values for random choices of a unitary stage used in the scheme.

In Section 8.2 we introduce the formalism needed to characterize Gaussian Quantum states and Passive Gaussian operations. Then, in Section 8.3 we will introduce the metrological setup under investigation. In Section 8.4 we will derive the Fisher Information associated with the setup and we will prove that the Heisenberg scaling can be achieved under suitable physical conditions. Although the results presented apply to an arbitrary number of modes, we will present an explicit Two-mode example in Section 8.5 to clarify how the setup works in a simple scenario. Then, in Section 8.6 we discuss the typicality of our results for interferometers with a large number of modes. Finally, in Section 8.7 we draw some conclusions and discuss the outlooks.

### 8.2 Gaussian States and Operations

Let us consider a set of \(M\) bosonic modes described by the operators \(\{\hat{a}_j\}_{j=1}^M\) and \(\{\hat{a}^\dagger_j\}_{j=1}^M\) satisfying the canonical commutation relations

\[
[\hat{a}_j, \hat{a}_k] = 0 = [\hat{a}_j^\dagger, \hat{a}_k^\dagger], \quad [\hat{a}_j, \hat{a}^\dagger_k] = \delta_{jk}, \quad (j,k = 1, \ldots, M).
\]
8.2. Gaussian States and Operations

The number operator will be given as usual by:

\[ \hat{N} = \sum_{j=1}^{M} \hat{a}^\dagger_j \hat{a}_j. \] (8.2)

In order to introduce the Gaussian states, it is convenient to introduce the quadrature operators \( \hat{x}_j \) and \( \hat{p}_j \) for each of the \( M \) modes,

\[
\begin{align*}
\hat{x}_j &= \frac{\hat{a}_j + \hat{a}_j^\dagger}{\sqrt{2}} \\
\hat{p}_j &= \frac{\hat{a}_j - \hat{a}_j^\dagger}{\sqrt{2}i}
\end{align*}
\]

\( \iff \)

\[
\begin{align*}
\hat{a}_j &= \frac{\hat{x}_j + i\hat{p}_j}{\sqrt{2}} \\
\hat{a}_j^\dagger &= \frac{\hat{x}_j - i\hat{p}_j}{\sqrt{2}i}
\end{align*}
\] (j = 1, \ldots, M). (8.3)

In terms of the operators \( \hat{x}_j \) and \( \hat{p}_j \), the canonical commutation relations (8.1) read:

\[
[\hat{x}_j, \hat{x}_k] = [\hat{p}_j, \hat{p}_k] = 0, \quad [\hat{x}_j, \hat{p}_k] = i\delta_{jk}. \] (8.4)

If we introduce the \( M \)-dimensional vectors \( \hat{x} \) and \( \hat{p} \) of components \( \hat{x}_j \) and \( \hat{p}_j \) respectively, we can align these vectors in a further column vector of dimension \( 2M \) as:

\[
\hat{z} = \begin{pmatrix} \hat{x} \\ \hat{p} \end{pmatrix} = \begin{pmatrix} \hat{x}_1 \\ \vdots \\ \hat{x}_M \\ \hat{p}_1 \\ \vdots \\ \hat{p}_M \end{pmatrix}
\] (8.5)

so that the above relation (8.3) can be expressed as

\[
\begin{pmatrix} \hat{a} \\ \hat{a}^\dagger \end{pmatrix} = W \begin{pmatrix} \hat{x} \\ \hat{y} \end{pmatrix}
\] (8.6)

where \( W \) is a \( 2M \times 2M \) unitary matrix defined as

\[
W = \frac{1}{\sqrt{2}} \begin{pmatrix} \mathbb{I} & i\mathbb{I} \\ \mathbb{I} & -i\mathbb{I} \end{pmatrix},
\] (8.7)

and \( \mathbb{I} \) is the \( M \times M \) identity matrix. The canonical commutation relations can be then expressed in the following compact form:

\[
[\hat{z}_j, \hat{z}_k] = iJ_{jk}, \quad (j, k = 1, \ldots, 2M),
\] (8.8)

where \( J \) is the \( 2M \times 2M \) real matrix:

\[
J = \begin{pmatrix} 0 & \mathbb{I} \\ -\mathbb{I} & 0 \end{pmatrix}
\] (8.9)

which is usually called standard symplectic form.
8.2.1 Gaussian states

A given state $\hat{\rho}$ of the $M$ bosonic modes is said to be Gaussian if its characteristic function, defined as

$$\chi_{\hat{\rho}}(\eta) = \langle e^{i\eta \cdot \hat{z}} \rangle_{\hat{\rho}} = \text{Tr}(\hat{\rho} e^{i\eta \cdot \hat{z}}),$$

(8.10)

is Gaussian, i.e:

$$\chi_{\hat{\rho}}(\eta) = e^{-\frac{1}{2} \eta^T \Gamma \eta + i d},$$

(8.11)

where $\Gamma$ is the covariance matrix and $d$ is the displacement, respectively given by:

$$\Gamma_{jk} = \frac{1}{2} \langle \{\hat{z}_j, \hat{z}_k\} \rangle - \langle \hat{z}_j \rangle \langle \hat{z}_k \rangle,$$

$$d_j = \langle \hat{z}_j \rangle, \quad j, k = 1, \ldots, 2M,$$

(8.12)

(all the averages $\langle \cdot \rangle$ are taken on the state $\hat{\rho}$).

The average number of photons in a given state can be computed by evaluating the derivatives of the characteristic function at $\eta = 0$:

$$\langle \hat{N} \rangle = \frac{1}{2} \left[ \text{Tr} \left( \Gamma - \frac{1}{2} \right) + d^2 \right].$$

(8.13)

The covariance matrix $\Gamma$ is evidently a real symmetric matrix. It is also positive definite, since for any $\alpha \in \mathbb{R}^{2M}$:

$$\sum_{j,k=1}^{2M} \Gamma_{jk} \alpha_j \alpha_k = \frac{1}{2} \sum_{j,k=1}^{2M} \left( \langle \hat{z}_j \hat{z}_k \rangle + \langle \hat{z}_k \hat{z}_j \rangle \right) \alpha_j \alpha_k - \sum_{j,k=1}^{2M} \langle \hat{z}_j \hat{z}_k \rangle \alpha_j \alpha_k$$

$$= \left\langle \left( \sum_{j=1}^{2M} \alpha_j \hat{z}_j \right)^2 \right\rangle - \left\langle \sum_{j=1}^{2M} \alpha_j \hat{z}_j \right\rangle^2$$

$$= \langle \hat{z}(\alpha)^2 \rangle - \langle \hat{z}(\alpha) \rangle^2$$

where $\hat{z}(\alpha) := \sum_{j=1}^{2M} \alpha_j \hat{z}_j = \hat{z}(\alpha)^\dagger$, so that $\langle \alpha, \Gamma \alpha \rangle$ represents the variance of the hermitian operator $\hat{z}(\alpha)$ on the state $\hat{\rho}$, which is always a positive number.

**Theorem 26 (Williamson).** Given $\Gamma \in \mathcal{M}_{2M}(\mathbb{R})$, $\Gamma = \Gamma^T$, $\Gamma > 0$, there exists $S \in \text{Sp}_{2M}(\mathbb{R})$, and $D \in \mathcal{M}_M(\mathbb{R})$ diagonal and positive such that:

$$\Gamma = S^T \begin{pmatrix} D & 0 \\ 0 & D \end{pmatrix} S.$$

(8.14)

Using the Williamson theorem we can achieve the following canonical decomposition of the covariance matrix:

$$\Gamma = S \Sigma S^T, \quad S = RQR'$$

(8.15)
8.2. Gaussian States and Operations

with

\[ \Sigma = \begin{pmatrix} \sigma & 0 \\ 0 & \sigma \end{pmatrix}, \quad Q = \begin{pmatrix} e^R & 0 \\ 0 & e^{-R} \end{pmatrix}, \]

\[ R = W^\dagger \begin{pmatrix} U & 0 \\ 0 & U^* \end{pmatrix} W, \quad R' = W^\dagger \begin{pmatrix} U' & 0 \\ 0 & U'^* \end{pmatrix} W, \] (8.16)

where:

- \( R = \text{diag}(r_1, \ldots, r_M) \in \mathcal{M}_M(\mathbb{R}) \) and \( \{r_1, \ldots, r_M\} \) are called squeezing parameters;
- \( \sigma = \text{diag}(\sigma_1, \ldots, \sigma_M) \in \mathcal{M}_M(\mathbb{R}) \) and \( \{\sigma_1, \ldots, \sigma_M\} \) are called symplectic eigenvalues of \( \Gamma \);
- \( U \) and \( U' \) are unitary \( M \times M \) matrices.

The symplectic eigenvalues control the purity \( P(\hat{\rho}) \) of the Gaussian state \( \hat{\rho} \) through:

\[ P(\hat{\rho}) = \text{Tr} \hat{\rho}^2 = \frac{1}{\sqrt{\det(2\Gamma)}} = \prod_{j=1}^M \frac{1}{2\sigma_j}. \] (8.18)

Due to the uncertainty principle, the symplectic eigenvalues are bounded below by \( \sigma_j \geq 1/2 \), while the squeezing parameters are all positive. Without loss of generality, we can arrange properly \( R \) and \( R' \) in such a way to realize the ordering

\[ \sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_M \geq \frac{1}{2}, \quad r_1 \geq r_2 \geq \cdots \geq r_M \geq 0. \] (8.19)

The canonical decomposition is such that not only \( S \) is a symplectic matrix, but also \( R, R', Q \) taken individually are symplectic, their symplectic character being determined by:

\[ R^T J R = J, \quad R'^T J R' = J, \quad Q^T J Q = J. \] (8.20)

8.2.2 Passive linear operations

A passive linear network is described by a unitary operator \( \hat{U} \) whose action on the \( M \) modes can be written as

\[ \hat{U}^\dagger \hat{a}_j \hat{U} = \sum_{k=1}^M U_{jk} \hat{a}_k, \] (8.21)

where \( U \in \mathcal{U}(M) \) is a unitary matrix. The unitarity of \( U \) guarantees that the associated operator preserves the total number of photons, since:

\[ \hat{U}^\dagger \hat{N} \hat{U} = \sum_{j=1}^M \sum_{k,k'=1}^M U_{jk}^* U_{jk'} \hat{a}_k \hat{a}_{k'} = \sum_{k=1}^M \hat{a}_k^\dagger \hat{a}_k = \hat{N}, \]
hence the name “passive”: the action of the unitary $\hat{U}$ cannot change the energy contained into the system. By definition, $U$ is the matrix of the single-photon transition amplitudes, i.e. $|U_{jk}|^2$ is the probability that a single boson injected into the $k$-th input channel ends up in the $j$-th output channel due to the action of the network. Now note that from (8.6) and the unitarity of $W$ we get that:

$$
\hat{U}^\dagger \hat{z} \hat{U} = \hat{U}^\dagger W^\dagger \left( \begin{array}{c} \hat{a} \\ \hat{a}^\dagger \end{array} \right) \hat{U} = W^\dagger \left( \begin{array}{c} \hat{U}^\dagger \hat{a} \\ \hat{U}^\dagger \hat{a}^\dagger \end{array} \right) = W^\dagger \left( \begin{array}{cc} U & 0 \\ 0 & U^* \end{array} \right) \left( \begin{array}{c} \hat{a} \\ \hat{a}^\dagger \end{array} \right)
$$

so that in terms of $\hat{z}$ operators the action of a passive linear unitary reads:

$$
\hat{U}^\dagger \hat{z} \hat{U} = R \hat{z},
$$

with

$$
R = W^\dagger \left( \begin{array}{cc} U & 0 \\ 0 & U^* \end{array} \right) W = \left( \begin{array}{cc} \text{Re} \ U & -\text{Im} \ U \\ \text{Im} \ U & \text{Re} \ U \end{array} \right)
$$

being a real $2M \times 2M$ orthogonal symplectic matrix.

By construction, the transformation $\hat{U}$ maps the set of Gaussian states into itself. In particular, given a Gaussian state $\hat{\rho}$, with covariance matrix $\Gamma$ and displacement $d$, the gaussian state $\hat{\rho}' = \hat{U} \hat{\rho} \hat{U}^\dagger$ is characterized by

$$
\Gamma' = R \Gamma R^T, \quad d = Rd.
$$

### 8.3 Setup

We will consider a metrological scheme where the value of the parameter to be estimated is encoded into an $M$-modes passive linear interferometer described by the unitary $\hat{U}_\phi$. At variance with the discussion of Chapter 7, we will not require that the unitary can be written in the form $U_\phi = e^{i\phi H}$. We will instead define the generator of the unitary evolution as the Hermitian matrix:

$$
G_\phi = iU_\phi^\dagger \frac{dU_\phi}{d\phi},
$$

which in the particular case where $G_\phi \equiv H$ does not depend on $\phi$ yields the evolution $U_\phi = e^{i\phi H}$.

The case of a parameter dependent generator allows to consider a more general scenario, where the parameter to be estimated is not just a phase shift, but is encoded in an arbitrary way in the whole interferometer, as long as it corresponds to a passive linear network. Passive linear networks can always be constructed by using beam splitters and phase shifters [145], so the situation we are considering here is the one depicted in Figure 8.1.
8.3. Setup

Figure 8.1: An $M$-mode passive linear interferometer composed of an arbitrary number of beam splitters and phase shifters having different and generic dependencies on a parameter $\varphi$. This scheme may represent for example the situation where $\varphi$ is the temperature or the magnitude of an electromagnetic or a gravitational field affecting the optical properties of the interferometer components.

8.3.1 Metrological scheme

We now investigate an estimation scheme reaching Heisenberg scaling if suitable conditions are satisfied. As shown in Figure 8.2, the preparation of the input probe consists in two steps: first, we inject a single-mode squeezed vacuum in the first port of $\hat{V}_{\text{in}}$. Then, the unitary stage $\hat{V}_{\text{in}}$ is used to scatter the photons injected among all the modes. The input state of the network $\hat{U}_\varphi$ in our protocol is therefore given by

$$|\psi_0\rangle = \hat{V}_{\text{in}} \hat{S}_1(r) |\text{vac}\rangle,$$

(8.27)

where

$$\hat{S}_1(r) = e^{i\hat{a}^2 - \hat{a}^4}$$(8.28)

is a single-mode squeezing operator with squeezing parameter $r > 0$, and

$$|\text{vac}\rangle = |0\rangle \otimes M$$

(8.29)

is the $M$-mode vacuum state. The average number of photons injected in the apparatus is thus $N = \sinh^2 r$. The state $\hat{U}_\varphi |\psi_0\rangle$ at the output of the network $\hat{U}_\varphi$ undergoes the unitary evolution $\hat{V}_{\text{out}}$ which refocuses all the photons into a single mode, namely the first one, where an homodyne measurement of the field quadrature

$$\hat{x}_\theta = e^{i\theta \hat{a}^\dagger \hat{a}} \hat{x}_1 e^{-i\theta \hat{a}^\dagger \hat{a}}$$

(8.30)
Figure 8.2: Block diagram of the investigated setup. A single-mode squeezed vacuum state with real squeezing parameter $r$ is injected into the first preparation stage $V_{\text{in}}$, which inputs the linear network $U_\varphi$ encoding the parameter $\varphi$ to be estimated. After the network, there is a second stage $V_{\text{out}}$ before the measurement. Finally, homodyne detection on the first output port of $V_{\text{out}}$ is performed, and the quadrature field $\hat{x}_\theta$ is measured. In order to reach Heisenberg scaling sensitivity in the estimation of $\varphi$, it suffices to optimize only one of the two auxiliary stages, in such a way that one of the conditions (8.54) or (8.55) holds.

is performed. The homodyne measurement is described by a Positive Operator Valued Measurement (POVM) $\mathcal{M} = \{\hat{\Pi}_x\}$, whose elements are defined by

$$\hat{\Pi}_x = e^{i\hat{a}_1^\dagger \hat{a}_1} |x\rangle \langle x| e^{-i\hat{a}_1^\dagger \hat{a}_1}. \quad (8.31)$$

If the refocusing procedure is not perfect there will be some photons scattered into other channels with probability $1 - P_\varphi$, where

$$P_\varphi = |(V_{\text{out}} U_\varphi V_{\text{in}})_{11}|^2. \quad (8.32)$$

is defined by the the probability amplitude $(u_\varphi)_{11} = (V_{\text{out}} U_\varphi V_{\text{in}})_{11}$ for the transition from the first input to the first output port in the overall interferometer $u_\varphi = V_{\text{out}} U_\varphi V_{\text{in}}$, with $V_{\text{in}}$ and $V_{\text{out}}$ being the single-photon unitary matrix representatives of $\hat{V}_{\text{in}}$ and $\hat{V}_{\text{out}}$ respectively, obtained using (8.21):

$$\hat{V}_{\text{in}}^\dagger \hat{a}_i \hat{V}_{\text{in}} = \sum_{j=1}^M (V_{\text{in}})_{ij} \hat{a}_j, \quad \hat{V}_{\text{out}}^\dagger \hat{a}_i \hat{V}_{\text{out}} = \sum_{j=1}^M (V_{\text{out}})_{ij} \hat{a}_j. \quad (8.33)$$

The probability of obtaining a value $x$ from a measurement of the quadrature $\hat{x}_\theta$ is then given by Born’s rule

$$p(x|\varphi) = \text{Tr} \left( \hat{\Pi}_x \hat{\varphi} \hat{S}_1(r) |\text{vac}\rangle \langle \text{vac}| \hat{S}_1^\dagger(r) \hat{a}_\varphi^\dagger \right), \quad (8.34)$$
over the output state $\hat{u}_\varphi \hat{S}_1(r) |\text{vac}\rangle$ after the overall interferometric evolution $\hat{u}_\varphi = \hat{V}_{\text{out}} \hat{U}_\varphi \hat{V}_{\text{in}}$, which yields (see Appendix D.1)

$$p(x|\varphi) = \frac{1}{\sqrt{2\pi\Delta\varphi}} \exp\left(-\frac{x^2}{2\Delta\varphi}\right),$$

(8.35)

where the variance of the Gaussian distribution

$$\Delta\varphi = \frac{1}{2} \left(1 + |(u_\varphi)_{11}|^2 (\cosh 2r - 1) + \text{Re}[e^{-2i\theta}(u_\varphi)_{11}^2] \sinh 2r\right),$$

(8.36)

encodes the parameter $\varphi$ through the interferometric transition amplitude $(u_\varphi)_{11} = (V_{\text{out}}U_\varphi V_{\text{in}})_{11}$.

### 8.4 Heisenberg scaling

We now want to prove that the described setup allows for an estimation of $\varphi$ with Heisenberg Scaling precision, i.e.

$$\delta\varphi = \Theta \left(\frac{1}{N}\right),$$

(8.37)

where $\delta\varphi = \sqrt{\text{Var}[\hat{\varphi}]}$ is the error associated to the estimator $\hat{\varphi}$. In order to do this, we will show that the Fisher Information associated with the setup scales quadratically in the number $N$ of probes employed.

#### 8.4.1 Fisher Information of the Setup

Using (7.17), we can write the Fisher Information for our setup as

$$F(\varphi) = \frac{1}{2} \left(\frac{\partial_{\varphi}\Delta\varphi}{\Delta\varphi}\right)^2.$$

(8.38)

In order to obtain a more manageable expression of $\Delta\varphi$, let us write

$$(u_\varphi)_{11} = |(u_\varphi)_{11}| e^{i\gamma_\varphi} = \sqrt{P_\varphi} e^{i\gamma_\varphi},$$

(8.39)

where $\gamma_\varphi = \text{arg}(u_\varphi)_{11}$ denotes the phase accumulated through the interferometric evolution $\hat{V}_{\text{out}} \hat{U}_\varphi \hat{V}_{\text{in}}$. In terms of $P_\varphi$ and $\gamma_\varphi$, equation (8.36) reads

$$\Delta\varphi = \frac{1}{2} \left(1 + P_\varphi (\cosh 2r - 1) + P_\varphi \cos(2\gamma_\varphi - 2\theta) \sinh 2r\right),$$

(8.40)

Then, recalling that $\sinh^2 r = N$, we can write the simplified expression

$$\Delta\varphi = \frac{1}{2} + P_\varphi f_\varphi(N),$$

(8.41)

where

$$f_\varphi(N) = N + \cos[2(\gamma_\varphi - \theta)] \sqrt{N(N+1)}.$$

(8.42)
The derivative of (8.41) reads
\[ \partial_{\varphi} \Delta_{\varphi} = (\partial_{\varphi} P_{\varphi}) f_{\varphi}(N) + 2P_{\varphi}(\partial_{\varphi} \gamma_{\varphi}) h_{\varphi}(N), \quad (8.43) \]
where
\[ h_{\varphi}(N) = \sin[2(\gamma_{\varphi} - \theta)] \sqrt{N(N + 1)}. \quad (8.44) \]
Substituting (8.41) and (8.43) into (8.38) we finally obtain the expression of the Fisher Information:
\[ F(\varphi) = 2 \left( \frac{(\partial_{\varphi} P_{\varphi}) f_{\varphi}(N) + 2P_{\varphi}(\partial_{\varphi} \gamma_{\varphi}) h_{\varphi}(N)}{1 + 2P_{\varphi} f_{\varphi}(N)} \right)^2. \quad (8.45) \]

### 8.4.2 Conditions for the Heisenberg Scaling

We can now use the expression of the Fisher Information (8.45) to show that if suitable conditions are met, the metrological setup shown in Figure 8.2 can reach Heisenberg scaling sensitivity.

The first condition is the constraint that the average number of photons scattered into channels which are not measured is a finite quantity \( \ell_{\varphi} = O(1) \), which translates into the condition
\[ P_{\varphi} = 1 - \frac{\ell_{\varphi}}{N} + O\left( \frac{1}{N^2} \right), \quad \ell_{\varphi} \geq 0 \quad (8.46) \]
on the probability \( P_{\varphi} \) in (8.32). Here, \( \ell_{\varphi} \) depends in general on the linear network \( U_{\varphi} \) in which the parameter is embedded, and on the auxiliary stages \( V_{\text{in}} \) and \( V_{\text{out}} \) in Figure 8.2. In subsection 8.4.3 we will show how, for any given arbitrary \( U_{\varphi} \), it is possible to optimize only a single stage so that the probability distribution in (8.32) can be expressed as (8.46). The second condition relates the accumulated phase \( \gamma_{\varphi} = \arg(u_{\varphi})_{11} \) through the whole setup and the phase \( \theta = \theta_{\varphi} \) of the local oscillator determining quadrature \( \hat{x}_\theta \) to be measured, according to
\[ \theta_{\varphi} = \gamma_{\varphi} \pm \frac{\pi}{2} + \frac{k_{\varphi}}{N} + O\left( \frac{1}{N^2} \right), \quad k_{\varphi} \neq 0, \quad (8.47) \]
where \( k_{\varphi} \) can depend on \( \varphi \), but is assumed to be independent of \( N \). In practice, one can even fix \( k_{\varphi} \) to a constant value without using additional resources.

A heuristic explanation behind this condition can be found in Figure 8.3: in order to maximise the ratio in (8.38) while keeping constant \( N = \sinh^2 r \), the choice of the quadrature \( \hat{x}_\theta \) to be measured is a trade-off between two opposite behaviours. One consists in minimizing the variance \( \Delta_{\varphi} \) in the denominator of (8.38), while the other consists in maximizing the sensitivity of the variance with respect to the variations of \( \varphi \), namely choosing \( \theta \) such that \( \partial_{\varphi} \Delta_{\varphi} \) in the numerator is maximal. The former is met for \( \theta \) as close as possible to \( \gamma \pm \pi/2 \), since \( \hat{x}_{\gamma \pm \pi/2} \) are the squeezed quadratures after the rotation in phase space by the phase \( \gamma_{\varphi} \) accumulated through the interferometer; the
latter instead requires a choice of $\theta - \gamma$ far enough from the stationary points of the variance at $\gamma \pm \pi/2$, where $\Delta_x$ and therefore the overall probability distribution $p(x|\varphi)$ are insensitive to variation in the parameter $\varphi$. Noticeably, the larger is $N$, and thus the squeezing parameter, the closer to the squeezed direction the quadrature field should be measured, as can be seen in (8.47).

If the direction of the quadrature being measured satisfies condition (8.47), the asymptotic behaviour of $f_\varphi(N)$ and $h_\varphi(N)$ for large $N$ is given by:

$$f_\varphi(N) = -\frac{1}{2} + \frac{2k_\varphi^2}{N} + \frac{1}{8N} + O\left(\frac{1}{N^2}\right),$$

$$h_\varphi(N) = 2k_\varphi\left(1 + \frac{1}{2N}\right) + O\left(\frac{1}{N^2}\right).$$

Substituting (8.48) and (8.49) into (8.45) and using condition (8.46), one easily obtains

$$F(\varphi) \sim 8\varrho(k_\varphi, \ell_\varphi)(\partial_\varphi \gamma_\varphi)^2 N^2,$$

with

$$\varrho(k, \ell) = \left(\frac{8k}{1 + 16k^2 + 4\ell}\right)^2.$$  

The quadratic scaling in the mean number photons $N$ in (8.50) finally proves that conditions (8.46) and (8.47) allow to reach the Heisenberg scaling.

The asymptotics for Fisher Information carries two pre-factors, $\varrho(k_\varphi, \ell_\varphi)$ and $(\partial_\varphi \gamma_\varphi)^2$. We will briefly discuss here the former, leaving a detailed discussion of the latter to Section 8.6.

We easily notice that the pre-factor $\varrho(k, \ell)$ vanishes only at $k = 0$, and attains its maximum at $\ell = 0, k = \pm 1/4$:

$$\varrho(k, \ell) \leq \varrho(\pm 1/4, 0) = 1,$$

so that, with this choice of the constants $k$ and $\ell$, the Fisher Information asymptotically reads

$$F(\varphi)\Big|_{k=\pm 1/4, \ell=0} \sim 8(\partial_\varphi \gamma_\varphi)^2 N^2$$

Moreover, $\varrho(k, \ell)$ is a decreasing function of $\ell$ independent of $k$, so that $\ell = 0$ is always the best case, meaning that the less photons that are scattered in different channels, the higher the sensitivity in the estimation. Instead, for a fixed arbitrary positive value of $\ell$, the maximum of $\varrho(k, \ell)$ is reached for $k = \pm \sqrt{4\ell + 1}/4$.

### 8.4.3 One-sided adaptivity

Since $P_\varphi = |(V_{\text{out}} U_\varphi V_{\text{in}})_{11}|^2$, condition (8.46) may appear to require a simultaneous optimization of the input $V_{\text{in}}$ and the output $V_{\text{out}}$ in a parameter-dependent way. This two-sided adaptation can be quite difficult to realize in practice.
Chapter 8. Distributed Quantum Metrology with Gaussian States

\[ e^{-r} \sqrt{2} e^{r} \sqrt{2} \]

\[ F(\phi) \sqrt{\Delta \phi} \]

**Figure 8.3:** Phase space representation of the squeezed vacuum state (with squeezing parameter \( r e^{2i\gamma \phi} \)) at the first output channel of the whole setup shown in Figure 8.2 (blue), and of the Fisher information (8.45) (four red lobes). We have considered for simplicity the case where all the photons are refocused in the first output channel (when condition (8.46) reduces to \( P_{\phi} = 1 \)). Given any axis at an angle \( \theta \) with respect to the horizontal axis, the distance between its intersections with the ellipse and the origin represents the standard deviation \( \sqrt{\Delta \phi} \) of the quadrature \( \hat{x}_\theta \). In other words, the blue graph is the polar plot of \( \sqrt{\Delta \phi} \) shown in (8.36) as a function of \( \theta \). A polar plot of the Fisher Information is overlaid in red. The Fisher Information takes vanishing values if the minimum-variance quadratures are measured, namely for \( \theta_{\text{min}} = \gamma \phi \pm \pi/2 \). This happens because the variance \( \Delta \phi \) of the quadrature along \( \theta_{\text{min}} \) is locally insensitive to the variations of \( \phi \). Thus, one needs to get far enough from \( \theta_{\text{min}} \) to achieve a suitable high value of the Fisher Information. In particular, we have shown that for a large number \( N \) of photons it is enough to move from \( \theta_{\text{min}} \) of an additional angle of the order \( 1/N \) as in (8.47) to reach the Heisenberg scaling in the measure of the parameter \( \phi \).

However, we are going to show that in fact conditions (8.46) and (8.47) can always be satisfied with just a one-sided parameter-dependent adaptation, which can be performed either at the input or at the output of the network. Remarkably, this adaptation can be accomplished by performing a preliminary classical, shot-noise limited, estimation of \( \phi \).

In particular, one can choose to adaptively optimize only \( V_{\text{out}} \) and fix \( V_{\text{in}} \) to an arbitrary parameter-independent unitary stage: in this case, one can set...
the parameter-dependent condition

\[(V_{\text{out}})_{1i} = (V_{\text{in}}^\dagger U_\varphi)_{1i} + O\left(\frac{1}{\sqrt{N}}\right), \quad (8.54)\]

Alternatively, it is possible to adaptively optimize only \(V_{\text{in}}\) with the condition

\[(V_{\text{in}})_{i1} = (U_\varphi^\dagger V_{\text{out}}^\dagger)_{i1} + O\left(\frac{1}{\sqrt{N}}\right), \quad (8.55)\]

while \(V_{\text{out}}\) can be arbitrarily chosen.

Equations (8.54) and (8.55) imply that an error \(O(1/\sqrt{N})\) is allowed to prepare the optimized stage to reach condition (8.46). To show that both equations (8.54) and (8.55) satisfy condition (8.46), we notice that \(P_\varphi\) can be expressed as a transition probability

\[P_\varphi = \left|\langle V_{\text{out}} U_\varphi V_{\text{in}} \rangle_{11}\right|^2 = \left|\langle v_{\text{out}} | v_{\text{in}} \rangle\right|^2, \quad (8.56)\]

between the two normalized vectors \(|v_{\text{in}}\rangle = U_\varphi V_{\text{in}} |e_1\rangle\) and \(|v_{\text{out}}\rangle = V_{\text{out}}^\dagger |e_1\rangle\), with \(|e_1\rangle = (1, 0, \ldots, 0)^T\). Then, equation (8.54) translates into

\[|v_{\text{out}}\rangle = |v_{\text{in}}\rangle + |\delta v\rangle = e^{i\varepsilon H} |v_{\text{in}}\rangle, \quad \varepsilon = O\left(\frac{1}{\sqrt{N}}\right), \quad (8.57)\]

with some \(H = H^\dagger\), by unitarity. Therefore, we can see that

\[
P = \left|\langle v_{\text{out}} | v_{\text{in}} \rangle\right|^2 = \left|\langle v_{\text{in}} | e^{-i\varepsilon H} |v_{\text{in}}\rangle\right|^2
\]

\[
= \left|1 - i\varepsilon \langle v_{\text{in}} | H |v_{\text{in}}\rangle - O(\varepsilon^2)\right|^2 = 1 - O(\varepsilon^2)
\]

\[
= 1 - O\left(\frac{1}{N}\right), \quad (8.58)
\]

In both equations (8.54) and (8.55), no assumption on the non-optimized stage is made, so that its choice is completely arbitrary. This freedom affects the precision of the estimation of \(\varphi\) through the \(N\)-independent pre-factor \((\partial_\varphi \gamma_\varphi)^2\) which appears in the asymptotic behaviour of \(F(\varphi)\), equation (8.50).

At this point, one may argue that this pre-factor may be vanishing if a poor choice for the non-adapted unitary is made. Remarkably, in Section 8.6 we will show that the pre-factor is typically non-vanishing for random choices of the non-adapted stage and suitably well-behaved given linear networks \(U_\varphi\).

### 8.5 A two-mode example

We show how our results, which are valid for an arbitrary \(M\)-port interferometer, can be applied to a particular example of a parameter \(\varphi\) distributed over a 2-channel interferometer, shown in the red box in Figure 8.4. In this setup, both the reflectivity \(\sin \eta_\varphi\) of a beam splitter, and the optical path lengths \(\lambda_\varphi\)
and $\lambda_\varphi'$ in the two arms depend on the parameter to be estimated: we can think of the parameter $\varphi$ as the magnitude of an external field, or of a characteristic of the environment, say the temperature, which in turn influences the optical properties of the devices. The functional dependence of $\eta_\varphi$, $\lambda_\varphi$ and $\lambda_\varphi'$ on $\varphi$ is assumed to be smooth. The distributed nature of $\varphi$ prevents us from thinking of it as a generalized phase, a case commonly studied in literature. The unitary matrix describing phase shifts $\lambda$ and $\lambda'$ on the two arms is the $2 \times 2$ diagonal matrix

$$U_{PS}(\lambda, \lambda') = \begin{pmatrix} e^{i\lambda} & 0 \\ 0 & e^{i\lambda'} \end{pmatrix}$$

while the action of a beam splitter with reflectivity $\sin \eta$ is given by

$$U_{BS}(\eta) = e^{i\sigma_y} = \begin{pmatrix} \cos \eta & \sin \eta \\ -\sin \eta & \cos \eta \end{pmatrix}$$

with $\sigma_y$ being the second Pauli matrix. Thus, the interferometer in Figure 8.4 is described by

$$U_\varphi = U_{PS}(\lambda_\varphi, \lambda_\varphi') U_{BS}(\eta_\varphi) = \begin{pmatrix} e^{i\lambda_\varphi} \cos \eta_\varphi & e^{i\lambda_\varphi'} \sin \eta_\varphi \\ -e^{i\lambda_\varphi'} \sin \eta_\varphi & e^{i\lambda_\varphi} \cos \eta_\varphi \end{pmatrix}.$$  

As previously discussed, Heisenberg scaling can be achieved by suitably adapting one of the two passive linear optical stages $\hat{V}_\text{in}$ and $\hat{V}_\text{out}$. Condition (8.46) is satisfied here with the arbitrary choice of $\hat{V}_\text{in}$ which is shown in Figure 8.4. It consists of a balanced beam splitter, followed by two $\pm \pi/4$-phase shifts, one on each arm, and thus is described by

$$V_\text{in} = U_{PS}(\pi/4, -\pi/4) U_{BS}(\pi/4).$$

The stage $\hat{V}_\text{out}$, which will have to be adapted, consists of two phase shifts, $\mp \alpha$, followed by another balanced beam splitter, and corresponds to the unitary matrix

$$V_\text{out} = U_{BS}(\pi/4) U_{PS}(-\alpha, +\alpha).$$

A direct computation of the matrix element $(V_\text{out} U_\varphi V_\text{in})_{11}$ gives for this scheme the transition probability (see equation (8.32)):

$$P_\varphi = \frac{1}{2} \left( 1 + \sin(\lambda_\varphi - \lambda_\varphi' - 2\alpha) \right),$$

and the accumulated phase (defined by (8.39))

$$\gamma_\varphi = \frac{\lambda_\varphi + \lambda_\varphi'}{2} + \eta_\varphi + \frac{\pi}{2}.$$  

The adaptive procedure in this example can be accomplished by simply tuning the phase shifts $\pm \alpha$ (see Figure 8.4) to $\pm \alpha_\varphi$, with $\alpha_\varphi = (\lambda_\varphi - \lambda_\varphi')/2 - \pi/4$, so that $P_\varphi = 1$.

Of course, tuning $\alpha$ requires a prior knowledge of the parameter we want
8.6 Typical sensitivity

to estimate. However, as discussed above, for any arbitrary given network $U_\varphi$, by denoting with $\delta \varphi = \varphi_{cl} - \varphi$ the difference between a previous coarse estimation $\varphi_{cl}$ and the true value $\varphi$ of the parameter, a classical precision $\delta \varphi = O(1/\sqrt{N})$ is sufficient to reach Heisenberg scaling. Indeed, by tuning the phase shifters in the output stage according to the coarse estimation of the parameter, equation (8.64) reads

$$P_\varphi = \frac{1 + \cos (\lambda_\varphi - \lambda_{\varphi_{cl}} - \lambda'_\varphi + \lambda'_{\varphi_{cl}})}{2}.$$  \hfill (8.66)

Thus, a Taylor expansion for small values of $\delta \varphi$ shows that

$$P_\varphi \sim 1 - \frac{1}{4} \left( \frac{\partial (\lambda_\varphi - \lambda_{\varphi_{cl}})}{\partial \varphi} \right)^2 \delta \varphi^2.$$  \hfill (8.67)

It is clear from this expression that it is possible to satisfy equation (8.46) with $\delta \varphi = O(1/\sqrt{N})$, which is achievable with a classical strategy employing $\alpha N$ photons for a measurement at the shot-noise limit, and $(1 - \alpha)N$ photons for the homodyne estimation.

Finally, in accordance with equations (8.47) and (8.65), the phase $\theta$ of the local oscillator in the homodyne detection must then be tuned according to the value

$$\theta_\varphi \sim \frac{\lambda_\varphi + \lambda'_\varphi}{2} + \eta_\varphi + \frac{k_\varphi}{N}.$$  \hfill (8.68)

We notice that, although not appearing in $\hat{V}_{out}$, the value of the unknown reflectivity $\sin \eta_\varphi$ influences the quadrature field to be measured.

8.6 Typical sensitivity

In this section we will address in more detail the study of the pre-factor $(\partial_\varphi \gamma_\varphi)^2$ in the Fisher Information (8.50), clarifying under what circumstances it can be safely considered non-vanishing and characterizing its magnitude for random choices of the non-optimized stage. First of all, we can link $(\partial_\varphi \gamma_\varphi)^2$ to the derivative of the matrix element $(u_\varphi)_{11} = (V_{out} U_\varphi V_{in})_{11} = \sqrt{P_\varphi} e^{i\gamma_\varphi}$:

$$|(\partial_\varphi u_\varphi)_{11}|^2 = \left| (\partial_\varphi \sqrt{P_\varphi} + i(\partial_\varphi \gamma_\varphi) \sqrt{P_\varphi}) e^{i\gamma_\varphi} \right|^2 = (\partial_\varphi \sqrt{P_\varphi})^2 + (\partial_\varphi \gamma_\varphi)^2 P_\varphi.$$  \hfill (8.69)

If condition (8.46) is satisfied, equation (8.69) simplifies to

$$(\partial_\varphi \gamma_\varphi)^2 = |(\partial_\varphi u_\varphi)_{11}|^2 + O \left( \frac{1}{N} \right),$$  \hfill (8.70)

so that the two quantities are equal up to order $1/N$. 
Chapter 8. Distributed Quantum Metrology with Gaussian States

Now, if the adaptation is performed in the output, i.e. we choose an arbitrary $V_{in}$ and adapt $V_{out}$ according to equation (8.54), we see that

$$|⟨(\partial_\varphi u_{\varphi})_{11}⟩|^2 = (V_{in}^\dagger G_{\varphi} V_{in})_{11}^2 + O\left(\frac{1}{N}\right),$$

(8.71)

where $G_{\varphi}$ is the generator of $U_{\varphi}$, equation (8.26). If, on the other hand, condition (8.46) is realized through an adaptation on the input while taking an arbitrary $V_{out}$, then equation (8.55) implies that

$$|⟨(\partial_\varphi u_{\varphi})_{11}⟩|^2 = (V_{out} U_{\varphi} G_{\varphi} U_{\varphi}^\dagger V_{out}^\dagger)_{11}^2 + O\left(\frac{1}{N}\right),$$

(8.72)

Using equations (8.70)-(8.72), we can finally rewrite the asymptotic expression of the Fisher information (8.50) as

$$F(\varphi) \sim \varrho(k_{\varphi}, \ell_{\varphi}) f(U, G_{\varphi}) N^2,$$

(8.73)

as $N \to \infty$, where

$$f(U, G_{\varphi}) = (U^\dagger G_{\varphi} U)_{11}^2,$$

(8.74)

with $U = V_{in}$ if the optimization is performed on the output, while $U = U_{\varphi}^\dagger V_{out}^\dagger$ if the optimization is carried out on the input. It is worth to emphasize that the pre-factor $f(U, G_{\varphi})$ is completely independent on the choice of the optimized stage.
The maximization of the prefactor (8.74) can be realized, for example, if $U = V_\varphi$ is some unitary diagonalizing $G_\varphi$ i.e. satisfying equation $V_\dagger G_\varphi V_\varphi = D_\varphi$ with $D_\varphi = \text{diag}(g_1, g_2, \ldots, g_M)$ being the diagonal matrix of the eigenvalues of $G_\varphi$, ordered in such a way that $|g_1| = \|G_\varphi\|$ is the maximum eigenvalue in absolute value [106]. Actually, it is not necessary to take a diagonalizing unitary to maximize (8.74), since only the first column of $U$ enters in the definition of $f(U, G_\varphi)$; hence, to maximize $f(U, G_\varphi)$ it is sufficient to require this column to be the eigenvector of $G_\varphi$ corresponding to the maximum eigenvalue $\|G_\varphi\|$. However, even that requirement would necessitate the complete knowledge of $G_\varphi$, which in general depends on the unknown parameter $\varphi$. Therefore, it is more relevant to consider arbitrary choices of the non-adapted network (the unitary $U$) independently of $\varphi$ in order to determine the practical advantages of the obtained Heisenberg scaling precision for finite values of $N$ and only one (classically) adapted stage.

For this reason, we will perform now a statistical analysis on the typical values which can be assumed by the prefactor $f(U, G_\varphi)$ for random choices of the unitary $U$ according to the Haar measure on $\mathcal{U}(M)$.

### 8.6.1 Average Behaviour

Denoting with $\{|j\rangle\}_{j=1}^M$ the canonical basis of $\mathbb{C}^M$, the prefactor can be written as

$$f(U, G_\varphi) = \langle 1 | U^\dagger G_\varphi U | 1 \rangle^2 = \text{tr}(G_\varphi U P U^\dagger)^2 \quad (8.75)$$

where $P = \langle 1 \rangle \langle 1 \rangle$. Using the notation of Chapter 2 (see equations (2.56) and (2.57)), we have:

$$f(U, G_\varphi) = \langle G_\varphi \rangle_U^2 = \langle G_\varphi^2 \rangle_U - \langle (\Delta G_\varphi)^2 \rangle_U \quad (8.76)$$

Then, using Lemma 10 (with $n = M$ and $r = \dim P = 1$), the average of $f(U, G_\varphi)$ is given by:

$$E[f(U, G_\varphi)] = E[\langle G_\varphi^2 \rangle_U] - E[\langle (\Delta G_\varphi)^2 \rangle_U] \quad (8.77)$$

$$= \frac{\text{tr}(G_\varphi^2)}{M} - \frac{1}{M + 1} \left[ \text{tr}(G_\varphi^2) - \frac{\text{tr}(G_\varphi)^2}{M} \right] \quad (8.78)$$

which simplifies to:

$$E[f(U, G_\varphi)] = \frac{\text{Tr}(G_\varphi^2) + \text{Tr}(G_\varphi)^2}{M(M + 1)} \quad (8.79)$$

In the trivial case of a generator proportional to the identity, $G_\varphi = \|G_\varphi\| I$, which corresponds to the case of a network $U_\varphi = e^{i\|G_\varphi\| I}$ acting as a global phase shift dependent on $\varphi$, we have $\text{Tr}(G_\varphi)^2 = M^2 \|G_\varphi\|^2$ and $\text{Tr}(G_\varphi^2) = M \|G_\varphi\|^2$, so that the average value of the pre-factor equals the maximum one, $f_{\text{max}} = \|G_\varphi\|^2$, in accordance with the fact that in this particular case every unitary in $\mathcal{U}(M)$ diagonalizes $G_\varphi$. 

In general, we are interested in determining the conditions which make this average value in (8.79) as large as possible. First of all, we can note directly from expression (8.79) that eigenvalues of opposite signs can have a detrimental effect on this average, since they lower the value of $\text{Tr}(G_\varphi)$. In general, we can find a lower bound on the average value (8.79) using Jensen’s inequality

$$E[f(U, G_\varphi)] \geq E[(U^\dagger G_\varphi U)_{11}]^2 \geq \left[\frac{\text{Tr}(G_\varphi)}{M}\right]^2,$$

where we used again Lemma 10 for the average of $(U^\dagger G_\varphi U)_{11} = \langle G_\varphi \rangle_U$.

Notice that the right-hand side of this inequality is nothing but the square of the average between $G_\varphi$’s eigenvalues. Hence, if we have some degree of control on the eigenvalues, we can achieve a result which is a certain fraction $\alpha$ of the maximum value $f_{\text{max}}$ if the average of $G_\varphi$’s eigenvalues is at least a fraction $\sqrt{\alpha}$ of the maximum eigenvalue, namely:

$$\left[\frac{\text{Tr}(G_\varphi)}{M}\right]^2 \geq \alpha \|G_\varphi\|^2 \quad \Rightarrow \quad E[f(U, G_\varphi)] \geq \alpha f_{\text{max}}.$$

### 8.6.2 Typicality of the prefactor

Equation (8.79) determines the average prefactor for random choices of the non-adapted stage. Using the results discussed in Chapter 2 we can show now that for large interferometers (i.e. with many ports $M$), this average is typical with respect to random choices of the non-adapted stage. This is a consequence of the concentration of measure over the unitary group, Theorem 7. In order to apply Theorem 7 to $f(U, G_\varphi)$, we need a suitable upper bound on the Lipschitz constant of $f(U, G_\varphi)$. This is immediately obtained by Lemma 11, upon noting that $f(U, G_\varphi)$ can be written as (8.75), which gives:

$$|f(U, G_\varphi) - f(V, G_\varphi)| \leq 4\|G_\varphi\|^2 \|U - V\|_{\text{HS}}.$$  

Then, we can apply Theorem 7 to obtain:

$$\mu_H \left( \left| f(U, G_\varphi) - \frac{\text{Tr}(G_\varphi^2) + \text{Tr}(G_\varphi^2)}{M(M + 1)} \right| \geq \varepsilon \right) \leq 2 \exp \left( - \frac{M}{C \|G_\varphi\|^2 \varepsilon^2} \right).$$

where $C = 192$. This result tells us that for large interferometers it is extremely unlikely to obtain a pre-factor sensibly different from its average, since for large values of $M$ the probability of $f(U, G_\varphi)$ being different from its average is exponentially suppressed. Thus, for any well-behaved linear network $\hat{U}_\varphi$ such that the expectation value in (8.79) is far enough from zero, any random choice of the non-adapted stage in the proposed interferometric setup typically yields a Heisenberg-scaling precision for the estimation of $\varphi$ if the number $M$ of interferometric channels is large enough.
8.6. Typical sensitivity

Figure 8.5: Histograms of the random variable \( f(U, G_\varphi) = \langle U^\dagger G_\varphi U \rangle^2 \) numerically obtained with \( 10^5 \) samplings of \( U \) from the unitary group with Haar measure and choosing \( G_\varphi \) as a diagonal matrix with half entries \( 3 \)s and half entries made of \( 1 \)s. The normalization is chosen in such a way that the total area under the curve equals 1. For these particular cases an explicit analytic expression of the distribution is achievable, and is given by the orange curves. The derivation is shown in Section 8.6.3.
8.6.3 Analytic distribution of the pre-factor for generators with only two distinct eigenvalues

The concentration of the pre-factor around its average can be seen also from the exact distribution of the pre-factor computed for some particular cases, which is shown in Figure 8.5. We will derive here this distribution in the particular case of a generator $G_\varphi$ having only two distinct eigenvalues. First of all, note that the distribution of $f(U, G_\varphi)$ depends only on the eigenvalues of $G_\varphi$, which we denote with $g = (g_1, \ldots, g_M)$, dropping the $\varphi$ subscript for notational simplicity. This can be seen using the spectral decomposition of $G_\varphi = V_\varphi^\dagger D_\varphi V_\varphi$, where $D_\varphi = \text{diag}(g)$, yielding:

$$U^\dagger G_\varphi U = (V_\varphi U)^\dagger D_\varphi (V_\varphi U) \overset{D}{=} U^\dagger D_\varphi U,$$

where in the last step we used the invariance property of the Haar measure and we used the notation $\overset{D}{=}$ to say that the two random variables have the same distribution. In light of this remark, we have that

$$f(U, G_\varphi) \overset{D}{=} f(U, D_\varphi) = (U^\dagger D_\varphi U)^2_{11} = \left( \sum_{j=1}^{M} |u_j|^2 g_j \right)^2,$$

having defined the random vector $u = U e_1$ obtained by the application of the random matrix $U \in U(M)$ to the fixed basis vector $e_1 = (1, \ldots, 0)^T \in \mathbb{C}^M$. We see that $f(U, D_\varphi)$ can be interpreted as a weighted average of the eigenvalues of $G_\varphi$ with random weights; these weights are given by the square modulus of the components of a random vector drawn from the unit sphere in $\mathbb{C}^M$ with the Haar measure. The distribution of this random variable can be quite complicated for a generic choice of the $G_\varphi$’s eigenvalues $g = (g_1, \ldots, g_M)$. We will consider here the situation in which there are at most two distinct eigenvalues $g_1 \geq g_2 \geq 0$, i.e:

$$g = \underbrace{(g_1, \ldots, g_1)}_{k} \underbrace{, g_2, \ldots, g_2}_{M-k} \quad (8.85)$$

so that

$$f(U, D_\varphi) = \left( \sum_{j=1}^{k} |u_j|^2 + g_2 \sum_{j=k+1}^{M} |u_j|^2 \right)^2$$

$$= \left[ (g_1 - g_2) \sum_{j=1}^{k} |u_j|^2 + g_2 \right]^2 \quad (8.86)$$

where we used the normalization constraint

$$\sum_{j=1}^{M} |u_j|^2 = 1. \quad (8.87)$$
In order to get the distribution of (8.86), let us first consider the random quantity defined by the sum inside the brackets, namely:

\[ \tau(U) = \sum_{j=1}^{k} |u_j|^2. \]  

(8.88)

We start from the distribution \( q(t) \) of \( \tau(U) \), defined in such a way that \( q(t)dt \) is the probability to have

\[ t \leq \tau(U) = \sum_{j=1}^{k} |u_j|^2 \leq t + dt \]  

(8.89)

or, defining \( x_{2j-1} : = \text{Re} \ u_j \) and \( x_{2j} : = \text{Im} \ u_j \), the probability to have:

\[ t \leq \sum_{j=1}^{k} \left( x_{2j-1}^2 + x_{2j}^2 \right) \leq t + dt. \]  

(8.90)

This probability can be interpreted as the geometrical surface of a \( 2k \)-dimensional hyperspherical cap of a \((2M - 1)\)-dimensional hypersphere sitting in \( \mathbb{R}^{2M} \). Using this interpretation, one then finds that [116, 115]:

\[ q(t) = \frac{(M-1)!}{(k-1)!(M-k-1)!} t^{k-1} (1-t)^{M-k-1} \chi_{[0,1]}(t), \]  

(8.91)

where

\[ \chi_{[0,1]}(t) = \begin{cases} 1 & 0 \leq t \leq 1, \\ 0 & \text{otherwise}. \end{cases} \]  

(8.92)

Starting from the distribution (8.91) of \( \tau(U) \), the probability density function \( p(x) \) of

\[ f(U, \text{diag}(g)) = \left[ (g_1 - g_2)\tau(U) + g_2 \right]^2, \]  

(8.93)

can be found with a change of variables to be:

\[ p(x) = \frac{1}{2\sqrt{x}\Delta g} q \left( \frac{\sqrt{x} - g_2}{\Delta g} \right), \]  

(8.94)

where \( \Delta g := g_1 - g_2 \). We then have explicitly:

\[ p(x) = C \frac{(g_1 - \sqrt{x})^{M-k-1}(\sqrt{x} - g_2)^{k-1}}{\sqrt{x}} \chi_{[g_2,g_1]}(\sqrt{x}) \]  

(8.95)

where \( C \) is a normalization constant given by:

\[ C = \frac{1}{2(\Delta g)^{M-1}} \frac{(M-1)!}{(k-1)!(M-k-1)!}. \]  

(8.96)

This distribution is valid whenever \( G_\rho \) has only two distinct positive eigenvalues \( g_1 \geq g_2 \geq 0 \). Numerical results are compared with the probability
density function (8.95) in Figure 8.5. We can see the concentration of measure in action for this explicit example, since as $M$ increases, the distribution concentrates around its average.

### 8.7 Conclusions

We provided a metrological scheme that yields a Heisenberg scaling precision for the estimation of a generic parameter $\varphi$ encoded into a $M$-mode network. Our scheme can be applied regardless of the nature of the parameter, which can even be distributed among several components of the network. In particular, the proposed scheme makes use of a single-mode squeezed state as probe, scattered throughout all the modes by means of an auxiliary passive linear stage. Once the information on the parameter is gathered by the probe, this gets refocused on a single output channel by a second auxiliary stage, and then detected with homodyne measurement. The analysis of the Fisher Information associated with such scheme reveals that, if a constant average number of photons (not scaling with the total number of photons injected) is scattered into channels different from the measured one, due to an imperfect refocusing procedure, the Heisenberg limit can be asymptotically reached, provided that the homodyne detection is performed with a sufficient resolution. For a distributed parameter, the refocusing is generally parameter-dependent, implying some sort of adaptive procedure in order to correctly refocus the probe. However, we have shown that all the dependence on the parameter can be entirely bounded to only one of the two auxiliary stages, while the other only affects the estimation through a multiplicative pre-factor. Moreover, we have also shown that all the information on the parameter needed to sufficiently refocus the probe can be obtained with a classical shot-noise precision, meaning that the number of resources required to adaptively optimize the auxiliary stages is not detrimental for the Heisenberg scaling precision. Finally, we have shown that, for a large number of modes, Heisenberg scaling is typically obtained by an arbitrary non-adapted stage, with an exponentially suppressed probability of failure.
Conclusion and Outlook

Typical properties of random states have been discussed in this thesis, considering in particular their behaviour from a resource theoretic point of view and their usefulness for quantum metrology.

In Part I the essential techniques employed in the investigation have been introduced, in particular the concentration of measure phenomenon, its validity for the unitary group endowed with the uniform (unitarily invariant) Haar measure, and techniques from Random Matrix Theory (specifically about Wishart matrices) which prove to be useful in the investigation of the behaviour of bipartite quantum systems.

In Part II, after a brief introduction on the subject of quantum resource theories, we focused on conversions between pure bipartite states using only LOCC operations in Chapter 5 and conversions among pure states using only incoherent operations in Chapter 6. We have provided extensive numerical evidence that the relative volume of LOCC-convertible pairs of states vanishes in the limit of large dimensional systems. Numerical investigations about the statistical properties of the maximal success probability in stochastic conversions revealed that the limit behaviour can be characterized with a suitable rescaling which depends only on the fluctuations of the smallest eigenvalues of Wishart matrices. This connection between state conversions and extreme value statistics has been confirmed and exploited in Chapter 6, where it has been proven analytically that in the case of the resource theory of coherence, most pairs of quantum states are indeed incomparable (in the sense that they cannot be converted into each other); furthermore, numerical evidence shows that the distribution of the maximal success probability for stochastic conversions can be entirely described in terms of a Markovian process which describes the smallest components of the coherence vector of random (Haar) states. The common features displayed by the two resource theories of entanglement and coherence is a consequence of the fact that in both cases state conversions are completely characterized by a majorization relation (among the Schmidt coefficients of the states for entanglement, and among the coherence vectors for the resource theory of coherence); from a probabilistic point of view, the only difference between the two cases is in the probability distribution of the vectors involved in the majorization relation, which in one case (entanglement) are determined by the eigenvalues of Wishart matrices, while in the other (coherence) are given by the uniform distribution on the simplex.

In Part III, we investigated the subject of quantum metrology, and analysed the usefulness of random states for achieving a precision in the estimation of a physical parameter at the Heisenberg scaling, i.e. with an uncertainty of order $1/N$ in the number $N$ of probes employed for the
estimation. We briefly discussed the state of the art in Chapter 7, using some of the techniques introduced in Chapter 2 to show that generic random states are not useful for quantum metrology, since they typically yield a fundamental limit on precision which scales as \(1/\sqrt{N}\); on the contrary, generic states from the symmetric subspace, i.e. states having a bosonic structure, are useful for quantum metrology, i.e. they typically yield an uncertainty which scales as \(1/N\) in the number of resources employed. Then, in Chapter 8 we focused on the performances of particular states of \(M\) bosonic modes, namely the Gaussian ones, which are known to be relevant from an experimental point of view since they are easy to prepare and manipulate. We provided a scheme employing the squeezing as a resource to achieve a measurement which allows an estimation at the Heisenberg limit, \(\delta \varphi = O(1/N)\) in the mean number of photons \(N\). Then, using again the techniques presented in Chapter 2, we analysed in detail the performances for random choices of the unitaries involved in the setup, proving that the prefactor (constant of proportionality in the scaling) exhibits a typical behaviour for a large number of modes \(M\).

Results presented in Chapter 5, based on the methods discussed in Chapter 3, are published in [37]. The findings of Chapter 6 are published in [36]. The results presented in Chapter 8 are reported in [70] and [71].

8.8 Outlooks

Majorization and its generalizations appear to play an important role in determining state conversions in several resource theories [67, 81, 98]. It would be interesting to generalize the techniques employed here to these other scenarios. The difference would consist in the distribution of the vectors involved in the majorization relation for the particular resource theory considered. If this distribution is not too much complicated (e.g. a Dirichlet distribution), techniques employed in Chapter 6 could be easily generalized.

We focused our analysis on the investigation of deterministic conversions and stochastic conversions. These are both cases in which the final state is exactly the desired one. One could also consider approximate conversions, which up to now have been considered only in the asymptotic scenario of many independent copies. It would be interesting to understand what happens for approximate conversions among single pairs of states taken from a large dimensional system. What is the distribution of the maximal fidelity achievable in a state conversion between random initial and target states?

Another interesting line of research could be at the intersection of the resource theories with the field of quantum control, which I have investigated as part of my research activity, finding some results (not discussed in this thesis) which are published in [69, 24]. It would be interesting to build a framework to investigate control techniques with some restriction on the resources available, e.g. we want to tailor the temporal evolution of a quantum system according to some desired fashion, but only some kind of operations are possible. Another interesting development on this line would be to consider stochastic controls instead of deterministic ones.
The metrological scheme analysed in Chapter 8 involves only Gaussian states and Gaussian measurements, which are easily implementable experimentally. Hence, an experimental realization, even as a proof of principle, could provide interesting developments. It would also be interesting to understand how the results presented in Chapter 8 can be generalized to the multi-parameter scenario, where one wants to estimate multiple parameters, or some functions of them.

A further development of the research presented in this thesis lies at the crossover between the two main topics, resource theory and quantum metrology: it would be interesting to develop a resource theory for quantum metrology. As a matter of fact, some progress in this direction is already coming into light [90, 59], but state conversions have not been considered yet. It would be interesting to investigate if majorization could play some role also for state conversions in a resource theory of quantum metrology.
Appendix A

Log-Sobolev inequalities and Concentration of measure

In this appendix we discuss the concentration of measure for log-Sobolev spaces. The result discussed here are used in Chapter 2 to derive the concentration of measure on the unitary group.

A.1 Laplace bounds on the concentration function

From the results discussed in Chapter 1, it appears evident that if one is able to efficiently bound the concentration function, it is possible to use it to control deviations of any sufficiently regular function on the probability space. One way to derive quite good bounds is based essentially on the same idea which is behind the Chernoff bound (1.10), and it uses the Laplace functional of $\mu$ on $(X, d)$, which is defined as:

$$E_{(X,d,\mu)}(\lambda) = \sup \{ \mathbb{E}[e^{\lambda F}] : F \text{ bounded } , \| F \|_{\text{Lip}} = 1 , \mathbb{E}[F] = 0 \}, \quad (A.1)$$

for $\lambda \geq 0$.

**Proposition 11.** With the previous notation, one has

$$\alpha_{(X,d,\mu)}(r) \leq \inf_{\lambda \geq 0} e^{-\lambda r/2}E_{(X,d,\mu)}(\lambda) \quad (A.2)$$

In particular, if

$$E_{(X,d,\mu)}(\lambda) \leq e^{\frac{\lambda^2}{2}}, \quad \lambda \geq 0, \quad (A.3)$$

then, every 1-Lipschitz function $F : X \to \mathbb{R}$ is integrable and for every $r \geq 0$,

$$\mu(F \geq \mathbb{E}[F] + r) \leq e^{-cr^2/2} \quad (A.4)$$

and $(X, d, \mu)$ has normal concentration

$$\alpha_{(X,d,\mu)}(r) \leq e^{-cr^2/8} , \quad r \geq 0. \quad (A.5)$$

If $E_{(X,d,\mu)}(\lambda_0) < \infty$ for some $\lambda_0 > 0$, then $(X, d, \mu)$ has exponential concentration.
A.2 Concentration from Logarithmic Sobolev inequalities

In the following, we will consider a metric space \((X, d)\) equipped with a Borel probability measure \(\mu\). Let us denote entropy of a measurable function \(f : X \to [0, \infty)\) with respect to \(\mu\) as

\[
\text{Ent}(f) = \mathbb{E}[f \log(f)] - \mathbb{E}[f] \log(\mathbb{E}[f]).
\] (A.6)

Note that the entropy is homogeneous of degree 1, i.e. for \(c > 0\), \(\text{Ent}(cf) = c\text{Ent}(f)\) and that by Jensen’s inequality \(\text{Ent}(f) \geq 0\). A completely generic metric space \((X, d)\) may not have a smooth structure, in the sense that the gradient of functions may not be defined. However, the concept of length of the gradient may be extended in the following sense: for a locally Lipschitz function \(f : X \to \mathbb{R}\), we define

\[
|\nabla f|(x) := \limsup_{y \to x} \frac{|f(y) - f(x)|}{d(y, x)}.
\] (A.7)

For smooth functions \(\phi : \mathbb{R} \to \mathbb{R}\), the following chain rule is satisfied:

\[
|\nabla \phi(f)| \leq |\phi'(f)||\nabla f|.
\] (A.8)

An argument due to Herbst allows to derive the concentration of measure starting from log-Sobolev inequalities, i.e:

\[
\text{Ent}(f^2) \leq 2C\mathbb{E}[|\nabla f|].
\] (A.9)

**Theorem 27.** Let \((X, d, \mu)\) satisfy a log-Sobolev inequality with constant \(C > 0\). Then, every Lipschitz function \(F : X \to \mathbb{R}\) with \(\|F\|_{\text{Lip}} = 1\) has \(\mathbb{E}[F] < \infty\), and:

\[
\mu(|F - \mathbb{E}[F]| \geq r) \leq 2e^{-\frac{r^2}{2C}}.
\] (A.10)

In particular, the concentration function satisfies

\[
\alpha(r) \leq e^{-\frac{r^2}{2C}}, \quad r > 0.
\] (A.11)

**Proof.** The proof will make use of Proposition 11, and specifically we will show that the Laplace functional of \(\mu\) on \((X, d)\) can be bounded as:

\[
\mathbb{E}_{(X, d, \mu)}(\lambda) \leq e^{C\lambda^2/2}.
\] (A.12)

In order to prove (A.12), we want to show that for any bounded \(1\)-Lipschitz function \(F : X \to \mathbb{R}\), the following inequality holds:

\[
\mathbb{E}[e^{\lambda F}] \leq e^{C\lambda^2/2}.
\] (A.13)
We will show this by using the log-Sobolev inequality on the function $f$ defined by

$$f^2 := e^{\lambda F}.$$  \hfill (A.14)

For notational convenience, let us denote with $H(\lambda) := \mathbb{E}[e^{\lambda F}]$. We have that:

$$\text{Ent}(f^2) = \mathbb{E}[\lambda Fe^{\lambda F}] - H(\lambda) \log H(\lambda) = \lambda H'(\lambda) - H(\lambda) \log H(\lambda),$$ \hfill (A.15)

and

$$|\nabla f(x)| \leq \frac{\lambda}{2} e^{\lambda F(x)/2} |\nabla F|(x) \leq \frac{\lambda}{2} e^{\lambda F(x)/2},$$ \hfill (A.16)

where the last inequality follows from $|\nabla F| \leq \|F\|_{\text{Lip}} = 1$. Taking the square and the expectation of the inequality (A.16), we get:

$$\mathbb{E}[|\nabla f|^2] \leq \frac{\lambda^2}{4} H(\lambda).$$ \hfill (A.17)

Using (A.15) and (A.17), the log-Sobolev inequality (A.9) implies:

$$\lambda H'(\lambda) - H(\lambda) \log H(\lambda) \leq C \frac{\lambda^2}{2} H(\lambda),$$ \hfill (A.18)

which can be rearranged as

$$\frac{H'(\lambda)}{\lambda H(\lambda)} - \frac{\log H(\lambda)}{\lambda^2} \leq \frac{C}{2}.$$ \hfill (A.19)

Noting that

$$\frac{H'(\lambda)}{\lambda H(\lambda)} - \frac{\log H(\lambda)}{\lambda^2} = \frac{d}{d\lambda} \left( \frac{\log H(\lambda)}{\lambda} \right),$$ \hfill (A.20)

an integration in $\lambda$ of equation (A.19) with the initial condition $H(0) = 1$ yields:

$$\frac{\log H(\lambda)}{\lambda} \leq \frac{C\lambda}{2},$$ \hfill (A.21)

which implies the desired bound:

$$H(\lambda) = \mathbb{E}[e^{\lambda F}] \leq e^{\frac{C\lambda^2}{2}}.$$ \hfill (A.22)

\hfill \Box
Appendix B

Miscellaneous results on joint statistics

B.1 Eigenvalues of Wishart and fixed-trace Wishart matrices

In this section, we prove that if $W$ is a Wishart random matrix of dimension $n$ and parameter $m$, as defined in Definition 6, the joint eigenvalue distributions of the fixed trace ensemble $W/\text{tr}(W)$ is given by the probability density (3.25). Denoting with $(x_1, \ldots, x_n)$ the spectrum of $W$, note that the spectrum of $W/\text{tr}(W)$ is simply given by

$$\lambda_j = \frac{x_j}{\sum_{k=1}^n x_k}, \quad j = 1, \ldots, n. \quad (B.1)$$

which can be seen as a change of variables. Hence, the proof consists simply in showing that if $(x_1, \ldots, x_n)$ is distributed according to $G_{n,m}$ of equation (3.27), the vector $(\lambda_1, \ldots, \lambda_n)$ is distributed according to $g_{n,m}$ equation (3.25). This is a standard calculation of multivariate statistics. Considering the change of variables

$$(x_1, \ldots, x_n) \mapsto (\lambda_1, \ldots, \lambda_{n-1}, s) = (x_1/s, \ldots, x_{n-1}/s, s), \quad (B.2)$$

where $s = \sum_{k=1}^n x_k$, the Jacobian of the transformation is

$$\begin{vmatrix}
  s & 0 & \ldots & \lambda_1 \\
  0 & s & \ldots & \lambda_2 \\
  \vdots & \vdots & \ddots & \vdots \\
  -s & -s & \ldots & 1 - \sum_{k=1}^{n-1} \lambda_k
\end{vmatrix}
= s^{n-1}, \quad \begin{vmatrix}
  s & 0 & \ldots & \lambda_1 \\
  0 & \ddots & \vdots & \vdots \\
  \vdots & s & \ddots & \vdots \\
  0 & 0 & \ldots & 1
\end{vmatrix}
= s^{n-1}, \quad (B.3)$$

where in the second equality we have added to the last row the first $n-1$ rows, an operation that does not change the determinant. If we denote $\lambda_n = x_n/s = 1 - \sum_{k=1}^{n-1} \lambda_k$, we see that the probability density function of
\((\lambda_1, \ldots, \lambda_{n-1}, s)\) factorises
\[
\tilde{G}_{n,m}(\lambda_1, \ldots, \lambda_{n-1}, s) = C_{n,m}e^{-s}s^{n-1} \prod_{1 \leq i < j \leq n} (s\lambda_i - s\lambda_j)^2 \prod_{k=1}^{n} (s\lambda_k)^{m-n} = C_{n,m}e^{-s}s^{nm-1} \left( \prod_{1 \leq i < j \leq n} (\lambda_i - \lambda_j)^2 \prod_{k=1}^{n} \lambda_k^{m-n} \right). \tag{B.4}
\]

Integrating over \(s\) we obtain the distribution of the vector \((\lambda_1, \ldots, \lambda_{n-1})\). The integral gives the constant
\[
C_{n,m} \left( \int_{0}^{\infty} e^{-s}s^{nm-1} \, ds \right) = C_{n,m} \Gamma(nm) = c_{n,m}. \tag{B.5}
\]

Therefore, the \(n\)-tuple \((\lambda_1, \ldots, \lambda_{n-1}, \lambda_n)\), with \(\lambda_n = 1 - \sum_{k=1}^{n-1} \lambda_k\) has density
\[
g_{n,m}(\lambda_1, \ldots, \lambda_n) = c_{n,m} \prod_{1 \leq i < j \leq n} (\lambda_i - \lambda_j)^2 \prod_{k=1}^{n} \lambda_k^{m-n} \Delta_{n-1}(\lambda). \tag{B.6}
\]

with respect to the Lebesgue measure on the simplex.

### B.2 Order statistics of i.i.d. random variables and Markov property

We collect here a series of more or less known results about order statistics of independent random variables. In the following, \(X_1, X_2, \ldots\) are independent and identically distributed (i.i.d.) random variables with distribution function \(F(x) := \mathbb{P}(X_1 \leq x)\). We always assume that they have a density \(f(x) = F'(x)\).

For a finite family \(X_1, X_2, \ldots, X_n\), the order statistics \(X_{1:k}, k \leq n\), are the rearrangements of the variables in nonincreasing order, i.e. \(X_{1:k} \geq X_{2:k} \geq \cdots \geq X_{n:k}\). Of course, the order statistics are not i.i.d. random variables.

Under the previous assumptions on the distribution of the \(X_i\)'s, the order statistics have a density. The following lemma gives the explicit formulæ that we need for our calculations.

**Lemma 17.** Let \(X_1, X_2, \ldots, X_n\) as above. Then,

1) The density of \(X_{1:k}\) is
\[
f_{X_{1:k}}(x_k) = \frac{n!}{(n-k)!(k-1)!} F(x_k)^{n-k} f(x_k)(1 - F(x_k))^{k-1}; \tag{B.7}
\]
[(B.8)]

\[
\frac{\prod_{i=1}^{l} F(x_i)^{n_i} f(x_i)(F(x_i) - F(x_{i-1}))^{l_i-1} f(x_{i-1})(1 - F(x_{i-1}))^{k_i-1}}{(n-l)!(l-k-1)!(k-1)!}
\]

for \(x_k \geq x_i\), and zero otherwise;

[(B.9)]

\[
\frac{n!}{(n-k)!} f(x_1) f(x_2) \cdots f(x_k) F(x_k)^{n-k},
\]

for \(x_1 \geq x_2 \geq \cdots \geq x_k \geq 0\), and zero otherwise;

[(B.10)]

\[
\frac{n!}{(n-k)!} (1 - F(x_{n-k+1}))^{n-k} f(x_{n-k+1}) \cdots f(x_{n-1}) f(x_n),
\]

for \(x_{n-k+1} \geq x_{n-k+2} \geq \cdots \geq x_n \geq 0\), and zero otherwise.

**Proof.** The proof is rather elementary (see, e.g. [7]). We sketch only the proof of Part i) to give a flavour of the type of arguments involved. The probability that \(X_k^+\) is in \(x_k\), is the probability that, among \(X_1, \ldots, X_n\): one is in \(x_k\) (this gives a factor \(f(x_k)\)); exactly \((k-1)\) are larger than \(x_k\) (this gives the factor \((1 - F(x_k))^{k-1}\)); the remaining \((n-k)\) variables are smaller than \(x_k\) (corresponding to the factor \(F(x_k)^{n-k}\)). There are \(n!\) ways to partition the \(n\) variables in that manner.

**Proposition 12.** Let \(X_1, X_2, \ldots, X_n\) be as above. Then,

i) The vector \((X_1^+, X_2^+, \ldots, X_n^+)\) forms an inhomogeneous (finite) Markov chain with initial density

\[
f_{X_1^+}(x) = n F(x)^{n-1} f(x),
\]

and transition densities given by

\[
f_{X_k^+ | X_k^+} (y|x) = \begin{cases} 
(n - k) \frac{F(y)^{n-k-1}}{F(x)^{n-k}} f(y) & \text{if } y \leq x \\
0 & \text{otherwise}.
\end{cases}
\]

ii) The vector \((X_1^+, X_{n-1}^+, \ldots, X_n^+)\) forms an inhomogeneous (finite) Markov chain with initial density

\[
f_{X_k^+}(x) = n(1 - F(x))^{n-1} f(x),
\]
and transition densities given by

\[
f_{X^i_{n-k} | X^i_{n-k+1}}(y|x) = \begin{cases} (n-k) \frac{(1 - F(y))^{n-k-1}}{(1 - F(x))^{n-k}} f(y) & \text{if } y \geq x \\ 0 & \text{otherwise.} \end{cases} \quad (B.14)
\]

**Proof.** We prove Part i). The density (B.11) is a specialisation of (B.7). For \( k \leq l \), the conditional density of \( X^i_l \) given \( X^i_k \) is

\[
f_{X^i_l | X^i_k}(x_l | x_k) = \frac{f_{X^i_l, X^i_k}(x_l, x_k)}{f_{X^i_k}(x_k)} = \frac{(n-k)!}{(n-l)!(l-k)!} F(x_l)^{n-l} (F(x_k) - F(x_l))^{l-k-1} f(x_l),
\]

for \( x_k \geq x_l \), and zero otherwise. In particular, for \( l = k + 1 \), we get (B.12). Similarly, from (B.9), we have, for all \( x_1 \geq x_2 \geq \cdots \geq x_k \),

\[
f_{X^i_{k+1} | X^i_k, X^i_2, \ldots, X^i_1}(x_{k+1} | x_1, x_2, \ldots, x_k) = \frac{f_{X^i_{k+1}, X^i_k, X^i_2, \ldots, X^i_1}(x_{k+1}, x_1, x_2, \ldots, x_k)}{f_{X^i_k, X^i_2, \ldots, X^i_1}(x_1, x_2, \ldots, x_k)} = (n-k) \frac{F(x_{k+1})^{n-k-1}}{F(x_k)^{n-k}} f(x_{k+1}),
\]

for \( x_{k+1} \leq x_k \), and zero otherwise. Hence, we have proved that

\[
f_{X^i_{k+1} | X^i_k, X^i_2, \ldots, X^i_1}(x_{k+1} | x_1, x_2, \ldots, x_k) = f_{X^i_{k+1} | X^i_k}(x_{k+1} | x_k).
\]

**Mutatis mutandi** we can prove Part ii).

**Remark 4.** The previous formulae for the densities can be rephrased in terms of the distribution functions:

\[
F_{X^i_1}(x) = \int_{-\infty}^{x} f_{X^i_1}(z) dz = F(x)^n, \quad (B.18)
\]

\[
F_{X^i_{k+1} | X^i_k}(y | x) = \int_{-\infty}^{y} f_{X^i_{k+1} | X^i_k}(z | x) dz = \left( \frac{F(\min(y, x))}{F(x)} \right)^{n-k}, \quad (B.19)
\]

\[
F_{X^i_1}(x) = 1 - \int_{x}^{\infty} f_{X^i_1}(z) dz = 1 - (1 - F(x))^n, \quad (B.20)
\]

\[
F_{X^i_{n-k} | X^i_{n-k+1}}(y | x) = 1 - \int_{y}^{\infty} f_{X^i_{n-k} | X^i_{n-k+1}}(z | x) dz = 1 - \left( \frac{1 - F(\max(y, x))}{1 - F(x)} \right)^{n-k}. \quad (B.21)
\]
B.3 Proof of Proposition 10

Proof. We first recall a standard representation for the uniform distribution in \( \Delta_{n-1} \) in terms of i.i.d. exponential random variables, and the classical asymptotic distributions of the extreme values for exponential random variables.

**Lemma 18.** Let \( X_1, X_2, \ldots \) be independent exponential random variables with rate 1, i.e. \( P(X \leq x) = 1 - e^{-x} \). Then, the vector

\[
(\mu_1, \mu_2, \ldots, \mu_n) := \left( \frac{X_1}{\sum_{i=1}^{\infty} X_i}, \frac{X_2}{\sum_{i=1}^{\infty} X_i}, \ldots, \frac{X_n}{\sum_{i=1}^{\infty} X_i} \right)
\]  

(B.22)

is uniformly distributed in \( \Delta_{n-1} \).

**Lemma 19.** If \( F(x) = 1 - e^{-x} \), then

\[
1 - (1 - F(u/n))^n = 1 - \exp(-u) \quad \text{(exponential distribution).} \\
\lim_{n \to \infty} F(\log n + u)^n = \exp(-e^{-u}) \quad \text{(Gumbel distribution)}
\]  

(B.23) (B.24)

Let \( \mu = (\mu_1, \mu_2, \ldots, \mu_n) \) defined as in (B.22) be a uniform point on \( \Delta_{n-1} \). Combining Proposition 12, Remark 4, and formula (B.23), we see that for any fixed \( k \geq 1 \),

\[
\left( nX_{n,1}^\downarrow, nX_{n-1,1}^\downarrow, \ldots, nX_{n-k+1}^\downarrow \right)
\]

converges in distribution to the first \( k \) components \( (V_1, V_2, \ldots, V_k) \) of the time-homogeneous Markov chain \( (V_j)_{j \geq 1} \) with density of \( V_1 \) and transition density

\[
f_{V_1}(v) = \exp(-v)1_{v \geq 0}, \quad f_{V_{j+1}|V_j}(u|v) = \exp(v-u)1_{u \geq v},
\]

respectively. To show the convergence for the order statistics of \( \mu \), we simply observe that the vector

\[
\left( n^2 \mu_{n-j+1}^\downarrow \right)_{1 \leq j \leq k} = \left( \frac{n}{\sum_{i=1}^{\infty} X_i} nX_{n-j+1}^\downarrow \right)_{1 \leq j \leq k}
\]

has the same limit distribution of \( \left( nX_{n-j+1}^\downarrow \right)_{1 \leq j \leq k} \). (Recall that \( \mathbb{E}[\sum_{i=1}^{\infty} X_i] = n \); hence the factor \( n^{-1} \sum_{i=1}^{\infty} X_i \) converges to 1 by the law of large numbers.)

Similarly, from Proposition 12, Remark 4, and the asymptotic formula (B.24), we deduce that

\[
\left( X_1^\downarrow - \log n, X_2^\downarrow - \log n, \ldots, X_k^\downarrow - \log n \right)
\]

converges in distribution to the first \( k \) components \( (W_1, W_2, \ldots, W_k) \) of the time-homogeneous Markov Chain \( (W_j)_{j \geq 1} \) with density of \( W_1 \) and transition density

\[
f_{W_1}(w) = \exp(-e^{-w} - w), \quad f_{W_{j+1}|W_j}(u|w) = \exp(e^{-w} - e^{-u} - u)1_{u \leq w},
\]
respectively. Denote by $\mu^\downarrow$ the decreasing rearrangement of $\mu$. For any $k$, we want to show that

$$\left(n\mu_1^\downarrow - \log n, n\mu_2^\downarrow - \log n, \ldots, n\mu_k^\downarrow - \log n\right)$$

converges in distribution to $(W_1, W_2, \ldots, W_k)$. We write

$$\left(n\mu_j^\downarrow - \log n\right)_{1 \leq j \leq k} = \left(\sum_{i=1}^{n} X_i - \log n \right)_{1 \leq j \leq k},$$

and we want to show that this vector has the same limit distribution of $(X_1^\downarrow - \log n)_{1 \leq j \leq k}$, as $n \to \infty$. The factor $n^{-1} \sum_{i=1}^{n} X_i$ converges to 1 by the law of large numbers. For all $\epsilon > 0$,

$$P\left(\left|\sum_{i=1}^{n} X_i - 1\right| > \frac{\epsilon}{\log n}\right) = P\left(\sum_{i=1}^{n} X_i < \frac{n \log n}{\log n + \epsilon} \quad \text{or} \quad \sum_{i=1}^{n} X_i > \frac{n \log n}{\log n - \epsilon}\right)$$

$$= P\left(\sum_{i=1}^{n} X_i < n - \frac{n\epsilon}{\log n + \epsilon} \quad \text{or} \quad \sum_{i=1}^{n} X_i > n + \frac{n\epsilon}{\log n - \epsilon}\right)$$

$$\leq P\left(\sum_{i=1}^{n} X_i - n > \min\left\{\frac{n\epsilon}{\log n + \epsilon}, \frac{n\epsilon}{\log n - \epsilon}\right\}\right).$$

Recall that $\text{Var}[\sum_{i=1}^{n} X_i] = n$. Assuming $n > \exp(\epsilon)$, and using Chebyshev’s inequality we can estimate

$$P\left(\left|\sum_{i=1}^{n} X_i - 1\right| > \frac{\epsilon}{\log n}\right) \leq \frac{\text{Var}\left[\sum_{i=1}^{n} X_i\right]}{(n\epsilon/\log n + \epsilon)^2} = \frac{1}{n\epsilon^2} (\log n + \epsilon)^2.$$

Hence, $\left(\sum_{i=1}^{n} X_i - 1\right) \log n$ converges to 0 in probability as $n \to \infty$.  

\[\Box\]

### B.4 Proof of Claim (6.22) in Theorem 22

We want to prove that the persistence probability asymptotically vanishes,

$$\lim_{k \to \infty} P\left(\min_{1 \leq j \leq k} I_j \geq 0\right) = 0.$$
Notice that one has
\[
\lim_{k \to \infty} P \left( \min_{1 \leq j \leq k} I_j \geq 0 \right) = P \left( \inf_{k \geq 1} I_k \geq 0 \right) \leq P \left( \liminf_{k \to \infty} I_k \geq 0 \right) = P \left( \liminf_{k \to \infty} \frac{I_k}{k \log k} \geq 0 \right) .
\] (B.25)

Therefore, it is sufficient to show that
\[
P \left( \liminf_{k \to \infty} \frac{I_k}{k \log k} \geq 0 \right) = 0,
\] (B.28)

and this follows from the Lindeberg-Feller central limit theorem as we outline now.

Denote by \( A \) the event in (B.28).

Claim 1. \( P(A) \in \{0, 1\} \).

The proof of the Claim is almost verbatim the proof given by Pittel [128]. For the event
\[
A = \left\{ \liminf_{k \to \infty} \frac{I_k}{k \log k} \geq 0 \right\} ,
\]
we want to show that \( P(A) \in \{0, 1\} \). The key observation here is that the probability of the event \( A \) does not depend on the variables of \( \tilde{X}_1, \tilde{X}_2, \ldots, \tilde{X}_J \), no matter how large, albeit finite, \( J \) is. Indeed, let \( \tilde{V}_k(J) = \sum_{j=J+1}^{k} \tilde{X}_j \) for \( k > J \) and \( I_k(J) = \sum_{j=J+1}^{k} \tilde{V}_j(J) \) for \( k > J \) as well. Then,
\[
I_k - I_k(J) = \sum_{j=1}^{k} \tilde{V}_j - \sum_{j=J+1}^{k} \tilde{V}_j(J) = \sum_{j=1}^{k} j \tilde{X}_{k-j+1} - \sum_{j=1}^{k-J} j \tilde{X}_{k-j+1} = \sum_{j=k-J+1}^{k} j \tilde{X}_{k-j+1} .
\]

Therefore, almost surely
\[
\lim_{k \to \infty} \frac{1}{k \log k} \left| I_k - I_k(J) \right| = 0, \quad \text{for all} \ J.
\]

So, denoting
\[
A_J = \left\{ \liminf_{k \to \infty} \frac{I_k(J)}{k \log k} \geq 0 \right\} ,
\]
we can write for the symmetric difference \( A \triangle A_J \) of the events \( A \) and \( A_J \),
\[
P(A \triangle A_J) = 0, \quad \text{for all} \ J.
\]

Now observe that \( A_J \) is measurable with respect to \( (\tilde{X}_j)_{j>J} \). (Informally, the event \( A_J \) does not involve the first \( J \) variables \( \tilde{X}_1, \ldots, \tilde{X}_J \)). Then, writing a.a.
for “almost always” and i.o. for “infinitely often”,

\[ A_\infty = \liminf J A_J = \bigcup_{J \geq 1} \bigcap_{m \geq J} A_m = \{ A_J \text{ a.a.} \} \]

is a tail-event, and

\[ P(A \triangle A_\infty) = P(A \cap A_\infty^c) + P(A^c \cap A_\infty) = P(A \cap A_j^c \text{ i.o.}) + P(A^c \cap A_J \text{ a.a.}) \]

\[ \leq \sum_{J \geq 1} [P(A \cap A_J) + P(A^c \cap A_J)] = \sum_{J \geq 1} P(A \triangle A_J) = 0. \]

By the Kolmogorov 0-1 law, \( P(A_\infty) \in \{0, 1\} \), so from the previous calculation we obtain \( P(A) \in \{0, 1\} \), as well.

Given Claim 1, we can now complete the proof if we show that \( P(A) < 1 \). By the definition of \( A \), to do so it suffices to show that

\[ \lim_{k \to \infty} P\left( \frac{I_k}{k \log k} \geq -b \right) < 1, \quad \text{(B.29)} \]

for a constant \( b > 0 \). Writing \( I_k = \sum_{j=1}^k j \tilde{X}_{k-j+1} \) it is a routine matter to compute

\[ E[I_k] = 0, \quad \text{Var}[I_k] = 2 \sum_{j=1}^k j^2 = \frac{k(k+1)(2k+1)}{3} = O(k^3). \]

From this, one can check that the sequence \( I_k \) satisfies the Lindeberg-Feller conditions, and thus \( I_k / \sqrt{\text{Var}[I_k]} \) converges in distribution to the standard Gaussian variable as \( k \to \infty \). Hence,

\[ P\left( \frac{I_k}{k \log k} \geq -b \right) = \left( \frac{\sqrt{\text{Var}[I_k]}}{k \log k} \frac{I_k}{\sqrt{\text{Var}[I_k]}} \geq -b \right) \xrightarrow{k \to \infty} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-x^2/2} dx = \frac{1}{2} < 1. \]

\[ \square \]
Appendix C

Mathematical characterization of LOCC operations

Despite having a quite simple physical description, the precise mathematical characterization of LOCC is notoriously challenging [14, 41]. A first issue is due to the fact that the particular choice of local measurements to perform at some stage of a LOCC protocol depends on all the measurement results previously obtained (and shared between the parties). A further complication arises from the fact that the number of communication rounds is potentially unbounded [26]. A detailed mathematical description of the full LOCC class has been presented in [31] adopting the quantum instruments formalism. A full treatment of the subject is beyond the scope of this thesis, and the interested reader is referred to [31] and references therein. The aim of this section is to provide, some minimal background useful for the problems studied here.

C.1 Finite-round LOCC in bipartite systems

A thorough definition of finite-round LOCC has been presented in [41] in terms of the operator sum representation of the operations belonging to the class. I will present here a brief summary of this formalization.

First of all, one defines the one-way LOCC operations, in which two parties $A$ and $B$ perform local quantum operations on their respective quantum systems and there is a unidirectional communication of measurements results from $A$ to $B$, who can decide which operations to perform on its system based on the measurement results obtained by $A$. This is formalized as follows.

**Definition 25 (One-way LOCC).** A quantum operation $\Lambda \in \mathcal{L}(\mathcal{H}_1^A \otimes \mathcal{H}_1^B, \mathcal{H}_2^A \otimes \mathcal{H}_2^B)$ is called a one-way LOCC from $A$ to $B$ if it can be written in the form

$$\Lambda(\rho) = \sum_{\alpha=1}^{K} \sum_{\beta=1}^{L} (\mathbb{I}_2^A \otimes W_{\beta\alpha}^B)(V_{\alpha}^A \otimes \mathbb{I}_1^B)\rho(V_{\alpha}^A \otimes \mathbb{I}_1^B)\dagger(\mathbb{I}_2^A \otimes W_{\beta\alpha}^B)\dagger,$$
for all \( \rho \in S(\mathcal{H}_1^A \otimes \mathcal{H}_1^B) \) and some sequences of operators \( V_\alpha^A : \mathcal{H}_1^A \to \mathcal{H}_2^A \) and \( W_\beta^B : \mathcal{H}_1^B \to \mathcal{H}_2^B \) satisfying the completeness relations

\[
\sum_{\alpha=1}^{K} V_\alpha^A \dagger V_\alpha^A = \mathbb{I}_1^A \quad \text{and} \quad \sum_{\beta=1}^{L} W_\beta^B \dagger W_\beta^B = \mathbb{I}_1^B \quad \text{for each} \quad \alpha = 1, \ldots, K, \tag{C.2}
\]

where \( \mathbb{I}_k^S \) are the identity operators acting on \( \mathcal{H}_k^S \), for \( k = 1, 2 \) and \( S = A, B \).

The operational meaning behind the mathematical objects involved in Definition 25 is the following: the index \( \alpha \) labels all the possible outcomes resulting from the operations performed by party \( I \); the resulting global state if outcome \( \alpha \) was obtained is \( (V_\alpha^A \otimes I_B)\rho(V_\alpha^A \otimes I_B)^\dagger/p_\alpha \), where \( p_\alpha = tr(V_\alpha^A V_\alpha^A \rho) \) is the probability of obtaining that outcome when the initial state is \( \rho \). At this point party \( A \) communicates their results to party \( B \), who performs their operations, represented by \( \{W_\beta^B\}_\beta \), based on the outcome \( \alpha \) which was obtained by \( A \).

With this operational interpretation in mind we can go to the next level and define the more general finite round (bidirectional) LOCC transformations on a bipartite system, without feeling (too) lost by the cumbersome notation.

**Definition 26 (Finite round LOCC).** A quantum operation

\[
\Lambda \in \mathcal{L}(\mathcal{H}_A^A \otimes \mathcal{H}_B^B, \mathcal{H}_A^A \otimes \mathcal{H}_B^B) \tag{C.3}
\]

is called a finite-round LOCC if there exists an integer \( n > 0 \) and sequences of Hilbert spaces \( (\mathcal{H}_k^A \mathcal{H}_k^B)_{k=1}^{n+1} \) with \( \mathcal{H}_1^A = \mathcal{H}_1^B \) and \( \mathcal{H}_n^A = \mathcal{H}_n^B \) \((S = A, B)\), such that \( \Lambda \) can be written as

\[
\Lambda(\rho) = \sum_{\alpha_1, \ldots, \alpha_2n=1}^{K_1, \ldots, K_{2n}} V_{\alpha_1, \ldots, \alpha_2n}^{AB} \rho V_{\alpha_1, \ldots, \alpha_2n}^{AB} \dagger \tag{C.4}
\]

for all \( \rho \in S(\mathcal{H}_A^A \otimes \mathcal{H}_B^B) \), where \( V_{\alpha_1, \ldots, \alpha_2n}^{AB} : \mathcal{H}_A^A \otimes \mathcal{H}_B^B \to \mathcal{H}_k^A \otimes \mathcal{H}_k^B \) is given by

\[
V_{\alpha_1, \ldots, \alpha_2n}^{AB} := (\mathbb{I}_{n+1}^A \otimes W_{2n}^{\alpha_2n, \ldots, \alpha_1})(V_{2n-1}^{\alpha_2n-1, \ldots, \alpha_1} \otimes \mathbb{I}_n^B) \cdots (\mathbb{I}_2^A \otimes W_{2}^{\alpha_2, \alpha_1})(V_{1}^{\alpha_1} \otimes \mathbb{I}_1^B) \tag{C.5}
\]

with families of operators

\[
V_{2k-1}^{\alpha_2k-1, \ldots, \alpha_1} : \mathcal{H}_k^A \to \mathcal{H}_{k+1}^A, \quad k = 1, \ldots, n \tag{C.6}
\]

\[
W_{2k-1}^{\alpha_2k, \ldots, \alpha_1} : \mathcal{H}_k^B \to \mathcal{H}_{k+1}^B, \quad k = 1, \ldots, n \tag{C.7}
\]

such that for \( k = 1, \ldots, n \):

\[
\sum_{\alpha_2k-1=1}^{K_{2k-1}} (V_{2k-1}^{\alpha_2k-1, \ldots, \alpha_1})^\dagger V_{2k-1}^{\alpha_2k-1, \ldots, \alpha_1} = \mathbb{I}_k^A \tag{C.8}
\]
for each sequence of outcomes \((\alpha_{2k-2}, \ldots, \alpha_1)\) and:

\[
\sum_{\alpha_{2k}=1}^{K_{2k}} (W_{2k}^{\alpha_{2k-2}, \ldots, \alpha_1})^\dagger W_{2k}^{\alpha_{2k-2}, \ldots, \alpha_1} = \Pi_k^B, \tag{C.9}
\]

for each sequence of outcomes \((\alpha_{2k-1}, \ldots, \alpha_1)\).

In Definition 26, indices \(\alpha_{2k-1}\) with odd suffix denote the outcomes of measurements performed by party \(A\), and indices \(\alpha_{2k}\) denote the outcomes obtained by \(B\). At the \(k\)-th round, the operations performed by one party depend on all previous outcomes obtained by both parties.

### C.2 Simplification in the pure bipartite case

As proved in [97], any LOCC transformation on a pure bipartite state can be performed in a single-step one-way-communications LOCC protocol, i.e. a single generalized measurement performed by one party followed by the communication of the result to the other party and finally by a local unitary transformation based on the measurement outcome. Thus, the most general LOCC transformation \(\Lambda\) on a pure bipartite state \(|\psi\rangle\) in \(\mathcal{H}_A \otimes \mathcal{H}_B\) can be written as

\[
\Lambda(|\psi\rangle) = \sum_{\alpha} (M_\alpha \otimes U_\alpha) |\psi\rangle (M_\alpha^\dagger \otimes U_\alpha^\dagger), \tag{C.10}
\]

where the operators \(M_\alpha\), with \(\sum_\alpha M_\alpha^\dagger M_\alpha = \mathbb{I}_A\), describe a generalised measurement on party \(A\) with outcomes labelled by \(\alpha\), while \(U_\alpha\) is the unitary performed on party \(B\) when the result of the measurement is \(\alpha\). As a particular case of the more general discussion after equation (4.18), we note that the operation (C.10) is also a convex combination of the form \(\Lambda(|\psi\rangle) = \sum_\alpha p_\alpha \rho_\alpha\), where

\[
p_\alpha = \text{Tr}(M_\alpha^\dagger M_\alpha |\psi_A\rangle \langle \psi_A|), \quad \rho_\alpha = (M_\alpha \otimes U_\alpha) |\psi\rangle (M_\alpha^\dagger \otimes U_\alpha^\dagger)/p_\alpha, \tag{C.11}
\]

with \(\rho_\alpha\) being the final state when the protocol is completed, given that the outcome \(\alpha\) is obtained during the measurement stage. Therefore, the probability of success \(p\) of the conversion \(|\psi\rangle \overset{p}{\rightarrow} |\varphi\rangle\) by a LOCC of the form (C.10) is given by

\[
p(\psi; \varphi; \Lambda) = \sum_{\alpha: \rho_\alpha = \varphi} p_\alpha. \tag{C.12}
\]

Thus, the maximal conversion probability introduced in Definition 21 can be written using this notation as

\[
\Pi(\psi \rightarrow \varphi) = \max_{\Lambda \in \text{LOCC}} p(\psi; \varphi; \Lambda) \tag{C.13}
\]

As pointed out in [31], the set of LOCC is convex but it is not topologically closed, in the sense that there are some convergent sequences of LOCC protocols (e.g. approaching some probability of success \(p\) as more
rounds of communications are allowed) which are not LOCC-feasible in the limit [27]. However, this does not pose a problem in the definition of the maximal success probability $\Pi(\psi \rightarrow \varphi)$, since the sets of finite-round LOCC protocols (such as the one-round LOCC protocols we are interested here) are closed [31].
Appendix D

Homodyne detection

D.1 Probability distribution of the homodyne detection

In order to evaluate the probability distribution associated with the measure of the quadrature field $\hat{x}_\theta$, it is natural to proceed with the phase-space formalism for Gaussian quantum optics. For a systematic overview of this topic, several works and reviews can be found in literature [137, 136, 160]. In this appendix, we will briefly introduce only the concepts and tools needed to obtain the expression (8.35) for the probability distribution $p(x|\psi)$ of measuring the quadrature value $x$ at the output of the proposed interferometer in Figure 8.2. We recall first the definition (8.34)

$$p(x|\psi) = \langle \text{vac} | \hat{S}_1^\dagger(r) \hat{u}_\psi \hat{a}_1^\dagger \hat{a}_1 | x \rangle e^{-i \hat{x}_1 \cdot \hat{z}'},$$  \hspace{1cm} (D.1)

where $\hat{u}_\psi$ describes the overall interferometric evolution of the single mode squeezed state $\hat{S}_1(r) | \text{vac} \rangle$. To evaluate (D.1), it is useful to firstly recover its Fourier transform,

$$\chi(\xi|\psi) = \int dx \ p(x|\psi) e^{-ix\xi}$$

$$= \langle \text{vac} | \hat{S}_1^\dagger(r) \hat{u}_\psi^\dagger e^{i \theta \hat{a}_1^\dagger \hat{a}_1} e^{-i \hat{x}_1 \cdot \hat{z}'} | \text{vac} \rangle. $$ \hspace{1cm} (D.2)

It is possible to write this characteristic function in a more canonical way. Indeed, first notice that we can write (the derivation is given in Appendix D.1.1)

$$e^{i \theta \hat{a}_1^\dagger \hat{a}_1} e^{-i \hat{x}_1 \cdot \hat{z}'} = \hat{D}(\xi_\theta),$$ \hspace{1cm} (D.3)

where

$$\hat{D}(\xi_\theta) = e^{-i \xi_\theta \cdot \hat{z}'}$$ \hspace{1cm} (D.4)

is the displacement operator, with

$$\xi_\theta = \begin{pmatrix} \xi \cos \theta \\ \xi \sin \theta \end{pmatrix}, \quad \hat{z}' = \begin{pmatrix} \hat{x}_1 \\ \hat{p}_1 \end{pmatrix}. $$ \hspace{1cm} (D.5)
Then, using equation (D.3) we can write the characteristic function (D.2) as

\[
\chi(\xi|\varphi) = \text{Tr} \left[ \hat{D}(\xi) \hat{u}_\varphi \hat{S}_1(r) \ket{\text{vac}}\bra{\text{vac}} \hat{S}_1^\dagger(r) \hat{u}_\varphi^\dagger \right],
\]  
(D.6)

Due to the Gaussian nature of the squeezed vacuum state and the linearity of the interferometric setup, the characteristic function (D.6) is a Gaussian bivariate function centred at zero, of the form \[160\]

\[
\chi(\xi|\varphi) = e^{-\frac{1}{2} \xi^\dagger \Gamma' \xi},
\]  
(D.7)

where \(\Gamma'\) is the \(2 \times 2\) covariance matrix of the whole interferometer output state \(\hat{u}_\varphi \hat{S}_1(r) \ket{\text{vac}}\), reduced to the first mode.

The covariance matrix \(\Gamma_0\) of the input state \(\hat{S}_1(r) \ket{\text{vac}}\) is given by

\[
\Gamma_0 = \frac{1}{2} \begin{pmatrix} e^{2R} & 0 \\ 0 & e^{-2R} \end{pmatrix},
\]  
(D.8)

\(R = \text{diag}(r, 0, \ldots, 0) \in M_M(\mathbb{R})\). After the action of \(\hat{u}_\varphi = \hat{V}_\text{out} \hat{U}_\varphi \hat{V}_\text{in}\), the covariance matrix transforms into

\[
\Gamma_\varphi = R_\varphi \Gamma_0 R_\varphi^T,
\]  
(D.9)

where \(R_\varphi\) is the orthogonal and symplectic matrix associated with the interferometer unitary matrix \(u_\varphi\)

\[
R_\varphi = W \begin{pmatrix} u_\varphi & 0 \\ 0 & u_\varphi^\dagger \end{pmatrix} W = \begin{pmatrix} \text{Re}(u_\varphi) & -\text{Im}(u_\varphi) \\ \text{Im}(u_\varphi) & \text{Re}(u_\varphi) \end{pmatrix},
\]  
(D.10)

so that \(\Gamma_\varphi\) in (D.9) reads

\[
\Gamma_\varphi = \begin{pmatrix} \Delta X_\varphi^2 & \Delta X P_\varphi \\ (\Delta X P_\varphi^\dagger)^T & \Delta P_\varphi^2 \end{pmatrix},
\]  
(D.11)

where we have defined the \(M \times M\) matrices

\[
\Delta X_\varphi^2 = \frac{1}{2} \left[ \text{Re}(u_\varphi) e^{2R} \text{Re}(u_\varphi^\dagger) - \text{Im}(u_\varphi) e^{-2R} \text{Im}(u_\varphi^\dagger) \right]
\]  
\[
= \frac{1}{2} \left[ \text{Re}(u_\varphi \cosh(2R) u_\varphi^\dagger) + \text{Re}(u_\varphi \sinh(2R) u_\varphi^\dagger) \right],
\]  
(D.12)

\[
\Delta P_\varphi^2 = \frac{1}{2} \left[ -\text{Im}(u_\varphi) e^{2R} \text{Im}(u_\varphi^\dagger) + \text{Re}(u_\varphi) e^{-2R} \text{Re}(u_\varphi^\dagger) \right]
\]  
\[
= \frac{1}{2} \left[ \text{Re}(u_\varphi \cosh(2R) u_\varphi^\dagger) - \text{Re}(u_\varphi \sinh(2R) u_\varphi^\dagger) \right],
\]  
(D.13)

\[
\Delta X P_\varphi = \frac{1}{2} \left[ -\text{Re}(u_\varphi) e^{2R} \text{Im}(u_\varphi^\dagger) - \text{Im}(u_\varphi) e^{-2R} \text{Re}(u_\varphi^\dagger) \right]
\]  
\[
= \frac{1}{2} \left[ -\text{Im}(u_\varphi \cosh(2R) u_\varphi^\dagger) + \text{Im}(u_\varphi \sinh(2R) u_\varphi^\dagger) \right].
\]  
(D.14)

In the second line of each of the previous expressions, we have exploited the fact that \(R\) is real. We are interested in the evaluation of the covariance matrix
reduced to the first mode, which we denote with
\[ \Gamma' = \begin{pmatrix} (\Delta X^2)_{11} & (\Delta X P)_{11} \\ (\Delta X P)_{11} & (\Delta P^2)_{11} \end{pmatrix}, \quad (D.15) \]
and insert in (D.7). Our final step is to invert the Fourier transform to finally get the expression of the probability distribution \( p(x|\varphi) \) given in (8.35). In order to do that, we introduce the \( 2 \times 2 \) orthogonal matrix
\[ O_\theta = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix}, \quad (D.16) \]
such that \( \xi_\theta = O_\theta \xi_0 \), with \( \xi_0 = (\xi, 0)^T \). Then, the characteristic function (D.7) can be written in a more convenient way, namely
\[ \chi(\xi|\varphi) = e^{-\frac{1}{2} \xi^T \Gamma'_{\varphi} \xi_0} = e^{-\frac{1}{2} (O^T_{\varphi} \Gamma'_{\varphi} O_\theta)_{11} \xi^2}. \quad (D.17) \]
Exploiting the fact that that by the definition of \( R \):
\[
(u_\varphi \cosh(2R) u_\varphi^T)_{11} = \cosh(2r) |(u_\varphi)_{11}|^2 + \sum_{i=2}^{M} |(u_\varphi)_{i1}|^2 \\
= \cosh(2r) |(u_\varphi)_{11}|^2 + (1 - |(u_\varphi)_{11}|^2), \quad (D.18) \\
(u_\varphi \sinh(2R) u_\varphi^T)_{11} = \sinh(2r) |(u_\varphi)_{11}|^2, \quad (D.19)
\]
and some elementary trigonometry, the term \((O^T_{\theta} \sigma_{\varphi} O_\theta)_{11}\) can be further manipulated to match the expression of \( \Delta_{\varphi} \) given in (8.36). In fact
\[
(O^T_{\theta} \sigma_{\varphi} O_\theta)_{11} = \sum_{i,j=1,2} (O_\theta)_{1i} (O_\theta)_{j1} (\sigma_{\varphi})_{ij} \\
= \cos^2 \theta \ (\Delta X^2)_{11} + \sin^2 \theta \ (\Delta P^2)_{11} + 2 \cos \theta \sin \theta \ (\Delta X P)_{11} \\
= \frac{1}{2} \cos^2 \theta \ (\cosh(2r) |(u_\varphi)_{11}|^2 + \sinh(2r) \Re((u_\varphi)_{11}^2)) + \\
+ \frac{1}{2} \sin^2 \theta \ (\cosh(2r) |(u_\varphi)_{11}|^2 - \sinh(2r) \Re((u_\varphi)_{11}^2)) + \\
+ \frac{1}{2} \left( \cos^2 \theta + \sin^2 \theta \right) (1 - |(u_\varphi)_{11}|^2) + \\
+ \cos \theta \sin \theta \sinh(2r) \Im[(u_\varphi)_{11}^2] \\
= \frac{1}{2} \left( 1 + |(u_\varphi)_{11}|^2 (\cosh(2r) - 1) + \\
+ \sinh(2r) \left( \cos(2\theta) \Re((u_\varphi)_{11}^2) + \sin(2\theta) \Im((u_\varphi)_{11}^2) \right) \right) \\
= \frac{1}{2} \left( 1 + |(u_\varphi)_{11}|^2 (\cosh(2r) - 1) + \\
+ \Re[e^{-2i\theta} (u_\varphi)_{11}^2 \sinh(2r)] \right) \equiv \Delta_{\varphi}. \quad (D.20)
\]
After applying the inverse Fourier transformation on the Gaussian characteristic function \((D.7)\), the probability distribution reads
\[
p(x|\varphi) = \frac{1}{2\pi} \int d\xi \chi(\xi|\varphi)e^{i2\xi} = \frac{1}{\sqrt{2\pi\Delta\varphi}} \exp \left( -\frac{x^2}{2\Delta\varphi} \right)
\]  
(D.21)
as displayed in \((8.35)\).

### D.1.1 Derivation of \((D.3)\)

Exploiting the unitarity of \(e^{i\theta\hat{a}_1^\dagger}\hat{a}_1\), we can write
\[
e^{i\theta\hat{a}_1^\dagger}\hat{a}_1e^{-i\theta\hat{a}_1^\dagger}\hat{a}_1 = \exp \left( -i\xi e^{i\theta\hat{a}_1^\dagger}\hat{x}_1e^{-i\theta\hat{a}_1^\dagger}\hat{a}_1 \right).
\]  
(D.22)

By using the definition \(\hat{x}_1 = (\hat{a}_1 + \hat{a}_1^\dagger)/\sqrt{2}\), the first-mode quadrature operator along \(\theta\) reads
\[
e^{i\theta\hat{a}_1^\dagger}\hat{x}_1e^{-i\theta\hat{a}_1^\dagger}\hat{a}_1 = \frac{1}{\sqrt{2}}(\hat{a}_1(\theta) + \hat{a}_1(\theta)^\dagger),
\]  
(D.23)

where \(\hat{a}_1(\theta) = e^{i\theta\hat{a}_1^\dagger}\hat{a}_1e^{-i\theta\hat{a}_1^\dagger}\hat{a}_1\) is the first-mode annihilation operator at time \(\theta\) in the Heisenberg picture. The Heisenberg equation is obtained by taking the derivative of \(\hat{a}_1(\theta)\) with respect to \(\theta\), and reads
\[
\frac{d\hat{a}_1(\theta)}{d\theta} = i[\hat{a}_1(\theta)^\dagger\hat{a}_1(\theta), \hat{a}_1(\theta)] = -i\hat{a}_1(\theta),
\]  
(D.24)
since \([\hat{a}_1(\theta), \hat{a}_1(\theta)^\dagger] = [\hat{a}_1, \hat{a}_1^\dagger] = 1\) and \([\hat{a}_1(\theta), \hat{a}_1(\theta)^\dagger] = [\hat{a}_1, \hat{a}_1^\dagger] = 0\). Therefore,
\[
\hat{a}_1(\theta) = e^{-i\theta}\hat{a}_1, \quad \hat{a}_1(\theta)^\dagger = e^{-i\theta}\hat{a}_1^\dagger,
\]  
(D.25)
the second equality being obtained by taking the adjoint of the first. By plugging \((D.25)\) into \((D.23)\) one gets
\[
\xi e^{i\theta\hat{a}_1^\dagger}\hat{x}_1e^{-i\theta\hat{a}_1^\dagger}\hat{a}_1 = \xi \cos \theta \hat{x}_1 + \xi \sin \theta \hat{p}_1 = \xi_\theta \cdot \hat{z},
\]  
(D.26)
where the vectors \(\xi_\theta\) and \(\hat{z}\) are given in \((D.5)\). Inserting this expression into \((D.22)\), we finally obtain \((D.3)\).
Bibliography


