Probabilistic quantum metrology

Tesis
del programa de Doctorat en Física de la
Universitat Autònoma de Barcelona

Bernat Gendra Casalí

Departament de Física Teòrica: Informació i Fenòmens Quàntics
Universitat Autònoma de Barcelona, 08193 Bellaterra (Barcelona)

realitzada sota la direcció de

Dr. Ramon Muñoz Tapia
Dr. John Calsamiglia Costa

Bellaterra, octubre de 2015

$$a + b = x + n$$

Choose Life. Choose a job. Choose a career. Choose a family. Choose a fucking big television, choose washing machines, cars, compact disc players and electrical tin openers. Choose good health, low cholesterol, and dental insurance. Choose fixed interest mortgage repayments. Choose a starter home. Choose your friends. [...] Choose your future. Choose life...

—*Mark "Rent-boy" Renton*
Transpotting (1996)

Agraïments

Durant el procés de realització d'aquesta tesi he estat acompanyat per moltes persones que m'han recolzat, ajudat i animat. Vull començar aquests agraïments fent especial menció als meus directors de tesi, el Ramon Muñoz i el John Calsamiglia. Sense ells aquesta tesi no hauria estat possible. Ells, conjuntament amb l'Emili Bagan, han estat els exemples a seguir per aconseguir la intuïció, la passió i la rigorositat necessària per portar a terme aquest projecte. I també més enllà dels aspectes purament acadèmics, estic molt content d'haver compartit aquest camí amb ells. Sincerament, el "Burnie" us està molt agraït.

Vull agrair també a tots els que han format part del Grup d'Informació Quàntica amb qui he compartit aquest període. En particular, als dos amb els que he treballat colze amb colze: a l'Elio Ronco, coautor de la major part dels continguts d'aquesta tesi amb qui he viscut experiències molt interessants; i al Gael Sentís, amb qui, a base de conversar i debatre sobre quasi tots els temes possibles, he forjat una gran amistat.

No puc oblidar a tots aquells de fora de l'autònoma. El suport dels meus pares ha estat sempre present, i m'ha ajudat a superar els culs de sac de la vida. Ells, i també els meus germans i resta de la família, han fet que em sentis orgullós del que estava fent i ho afrontés amb força. També a la família extensa, que amb petites accions m'han facilitat molt aquesta feina. Estic molt agraït a tots els amics, especialment als "Pleiers of Mess" (i a totes les sessions lúdiques), i a la junta i resta de membres del Front Diabòlic de la Garriga (i a les carretilles, timbals i malabars).

He reservat aquest espai final per la persona que m'ha acompanyat de més aprop, que ha gaudit les alegries i ha patit els maldecaps, que m'ha ajudat, escoltat i ensenyat: l'Anna. A ella i als meus fills (Xènia i Nil) els dedico aquesta tesi.

Contents

CHAPTER 1

Introduction

One of the most relevant traits of human beings is the deep knowledge of the world around. This has granted us the possibility of adaptation and control over it, and thus it has allowed us to advance technologically and overcome the adversities that nature presents. Long ago we reached such a level of sophistication that simple observation is not enough to collect the information needed to get deeper knowledge. However, knowledge itself provides us with technological tools to construct measurement devices with increasing accuracy. Moreover, advanced statistical tools have been developed in order to infer information from the noisy data and asses error in the estimation. Metrology is the field of research on statistical tools and technological design of measurement devices to infer accurate information about parameters of physical systems. This work is devoted to the theoretical part of this field, in particular in the sub-field of quantum metrology, which has the special feature that the measurement itself has to be theoretically designed. The reason for this is that quantum measurements do not give in general perfect access to the system magnitudes. Instead, their outcomes are randomly distributed among a set of possible values. As a result, quantum metrology attempts to find the measurement that yields a distribution from which the statistical tools achieve the optimal estimation precision.

A process of estimation can be divided into the following parts: first the system is initialised in some probe state. Then the system goes through some interaction process that imprints an unknown parameter on it. This interaction process is what the experimentalist is aimed to study, and the parameter codifies the knowledge that she can learn about it. To this end she performs a measurement over the system and obtains a data sample. A last step of statistical inference extracts from that sample an estimate value for the parameter.

Classical parameter estimation has been widely studied. Its main target is to develop mathematical techniques that most efficiently infer the value of the parameter [Kay, 1993]. Yet, there is still some controversy in a concept that is one of the cornerstones of estimation theory, namely the probability. To state it briefly, the main difference between definitions comes from the role that an agent takes. Typically, the concrete values for probabilities can be based either on the agent beliefs or only on observation. This yields two main schools for parameter estimation which differ in the treatment of the parameter itself as a stochastic or a deterministic variable, the Bayesian and the Frequentist approaches respectively. Within the Bayesian approach, the experimentalist codifies its knowledge about the parameter as a probability distribution and she uses the collected data to update it. The main criticism about this approach is that before obtaining any data a prior probability distribution has to be chosen. This fact can have in principle a significant influence on the estimate. On the other hand, the Frequentist approach define the probability as frequencies in the limit of infinite repetitions. Hence, it considers the parameter to have a fixed (unknown) value.

Classically, a measurement gives access to the magnitudes of a system with a precision limited only by the possible noise caused by experimental imperfections. The only option that the experimentalist has to reduce the effects of such noise is to repeat the experiment. The amount of information gathered reduces the estimation error by dividing it by the squared root of the number of repetitions. On the other hand, in those cases in which the system is governed by the laws of quantum mechanics the choice of the measurement itself becomes another important aspect of metrology. The stochastic nature of quantum measurements adds an inherent noise that can not be reduced by means of technology advances. However, in a fully quantum mechanical setting, the experimentalist has more options than repeating the experiment, namely, the possibility of using entangling operations in the preparation and measurement steps. This can provide a change of scaling the estimation error, i.e. it is divided by the number of used systems (which takes an equivalent role in the quantum setting as the number of repetitions in the classical), instead of its square root. Since this result is a direct consequence of the uncertainty principle takes the name of Heisenberg scaling. Nevertheless, in the last years the agenda of quantum-enhanced metrology has been put under scrutiny by a number of results [Demkowicz-Dobrzański *et al.*, 2012; Escher *et al.*, 2011; Knysh *et al.*, 2011; Aspachs *et al.*, 2009; Huelga *et al.*, 1997] that show that under quite generic experimental noise, the estimation precision do not reach the Heisenberg limit. This has made that a lot of research effort focus on finding alternative schemes that push forward the limits and circumvent or diminish the detrimental effect of noise. This has entailed the of study particular systems with non-trivial noise-models [Chaves *et al.*, 2013; Chin *et al.*, 2012; Jeske *et al.*, 2014; Ostermann *et al.*, 2013; Szańkowski *et al.*, 2014], and non-linear interactions Boixo *et al.* [2007]; Napolitano *et al.* [2011], which enable quantum error-correction codes [Dür *et al.*, 2014; Kessler *et al.*, 2014; Arrad *et al.*, 2014].

In the process of designing an estimation protocol, the experimentalist uses a figure of merit to evaluate the performance of such protocols. This figure of merit allows her to order the different protocols in terms of her needs, taking into account what use will be given to the estimated value. Up until now most quantum metrology schemes and known bounds have been deterministic, that is, they are optimized in order to provide a valid estimate for each possible measurement outcome. In these cases, the used figure of merit is related only with the averaged error of the possible estimates in comparison with the true value of the parameter. This benchmarking of a protocol by its average performance is very natural and convenient, but there can be some scenarios in which this is not enough to express the concrete use that will be given to the obtained value.

For the sake of exemplification, consider a situation in which an experimentalist wants to distinguish between different processes that a system could have gone through. Each of the processes imprints to the system a different value of a parameter that characterise them. Hence, by estimating it she can discriminate between them. In order to be able to decide with some certainty which is the process, the experimentalist needs the estimation error to be smaller than the interval between the possible imprinted values of the parameter. Imagine now that the needed precision surpasses the deterministic limits. This does not mean that the experimentalist will never obtain an outcome that allows her to choose some option with high certainty. There may exist measurements that, although the average precision does not exceed these bounds, a better precision can be associated with some of the outcomes. The average precision tells you very little about the confidence that you have on a particular outcome. Notice that for this to happen there must be other highly imprecise outcomes to compensate the average. It is important to remark that the experimentalist must be able to tell surely if she has obtained one of these outcomes, because if not, the precision that she will have to report will be the average one. In this sense, we say that a probabilistic scheme is heralded. The protocol guarantees a minimal precision upon a heralded outcome. This can be understood as the experimentalist putting forward an extra feature that is always available to her, namely the possibility of post-selecting the outcomes of their measurements and giving with some probability an inconclusive answer.

In this thesis we choose a figure of merit that reflects the maximum precision one can obtain. We optimise the precision of a set of heralded outcomes, and quantify the chance of such outcomes to occur, or in other words the probability that the protocol fails to provide an estimate. With this we open the door to protocols that may provide an estimate with a guaranteed precision, way below the average optimal one. Returning to the example, if the experimentalist wants to make the right choice with high confidence, then it is clear that protocol that provides heralded outcomes might be very interesting to her. Classically this seems a mere characterisation problem, where different figures of merit describe different aspects of a protocol; and in practice the issue is rapidly solved if a complete characterisation is given by labelling every outcome with the precision they provide. However, as mentioned above, in quantum mechanics there are many ways in which data can be read-off from a quantum system. Hence, in order to talk about an optimal measurement we need to fix a particular figure of merit, and typically what optimizes one figure of merit gives a suboptimal performance regarding another figure of merit. We want to capture the fact that some outcomes can be ultra-precise, and hence we characterize the strategies using the optimal guaranteed precision obtained for a given success probability, in contrast to deterministic protocols that are based on averaging over all outcomes, regardless of their precision.

This use of probabilistic strategies in Quantum Information were firstly used precisely in the field of quantum state discrimination. Only recently, it was shown that for quantum parameter estimation schemes with fixed probe state and in the absence of noise, the precision of the favourable outcomes can be greatly enhanced well beyond the limits set for deterministic strategies [Fiurášek, 2006; Marek, 2013], of course at the price of discarding or abstaining on the unfavourable outcomes.

In this work we design probabilistic protocols for phase, direction and reference frame estimation. We show that for pure states post-selection can compensate a bad choice of probe state, or in other words, it can attain optimal precision bounds in situations where the probe state is given. However, probabilistic protocols are more effective where some noise is present in the state system or in the measurement process. In these cases the possibility of abstention can counterbalance the negative effects of noise. In particular, we show that adding the possibility of abstaining in phase estimation in presence of noise can produce an enhancement in precision that overtake the ultimate bound of deterministic protocols (obtained by optimising also the probe state). It is important to remark that the probabilistic protocols do not ensure the experimentalist such good precision for all the measurement instances, only for those cases in which the probabilistic protocol gives a favourable outcome. However, the bound derived is the best precision that can be obtained, and in this sense one can speak of ultimate bound in precision.

1.1 Outline

The main purpose of this thesis is to discuss the role of the abstention in protocols of quantum parameter estimation. We begin with a review of the basics of probability theory in Chapter 2. The different interpretations and definitions of the concept of probability are treated with special interest as they have an impact in the different approaches of parameter estimation. Chapter 2 has also a second part in which we introduce the basic aspects of quantum theory.

Chapter 3 is intended to introduce the reader to metrology tasks. It begins with a review of the classical parameter estimation. It presents both Bayesian and Frequentist approaches and its main results and properties. In the next section of the Chapter we examine quantum state discrimination problems, for the purpose of giving an initial grasp of quantum decisions tasks. We also present an unpublished solution for the discrimination of general symmetric states with a fixed rate of inconclusive results. In the last part of the Chapter we discuss the main topic of this Thesis, namely, the quantum parameter estimation. We review previous results on this field and introduce the theoretical framework of probabilistic protocols that will be used in the following Chapters.

Chapters 4 to 7 present the main contributions of this Thesis. We begin by discussing in Chapter 4 quantum metrology with pure probe states. Within this scheme, the problems of phase, direction and frame estimation are solved. In Chapters 5 and 6 the effect of noise is introduced in direction and phase estimation, respectively. Particularly, in Chapter 6 the ultimate bound for phase estimation is derived. The technical details have been moved to Appendix A.

In Chapter 7 we work out the problem of probabilistic cloning. We show that probabilistic protocols do not challenge the fundamental equivalence between estimation and cloning, i.e. they are asymptotically equivalent. The particular case of quantum clocks is derived explicitly, considering also general Hamiltonians with rationally independent energies. Some of the computations in this Chapter are rather specialise and they are presented in Appendix B. It is worth noticing that some of technical results gathered here may be useful in other contexts.

1.2 List of publications

B. Gendra, E. Ronco-Bonvehí, J. Calsamiglia, R. Muñoz-Tapia, E. Bagan , "Beating noise with abstention in state estimation", *New Journal of Physics* **14**, 105015 (2012).

B. Gendra, E. Ronco-Bonvehí, J. Calsamiglia, R. Muñoz-Tapia, E. Bagan , "Quantum Metrology Assisted by Abstention", *Physical Review Letters* **110**, 100501 (2013).

B. Gendra, E. Ronco-Bonvehí, J. Calsamiglia, R. Muñoz-Tapia, E. Bagan , "Optimal parameter estimation with a fixed rate of abstention", *Physical Review A* **88**, 012128 (2013).

B. Gendra, J. Calsamiglia, R. Muñoz-Tapia, E. Bagan, G. Chiribella, "Probabilistic Metrology Attains Macroscopic Cloning of Quantum Clocks", *Physical Review Letters* **113**, 260402 (2014).

J. Calsamiglia, B. Gendra, R. Muñoz-Tapia, E. Bagan , "Ultimate precision bounds from probabilistic quantum metrology", submitted to *Physical Review X*, preprint in *arXiv:1407.6910v1*.

Beyond the scope of the thesis

G. Sentís, B. Gendra, S. D. Bartlett, and A. C. Doherty, "Decomposition of any quantum measurement into extremals", *Journal of Physics A: Mathematical and Theoretical* **46**, 375302 (2013).

CHAPTER 2

Preliminary concepts

2.1 Basic concepts on Probability theory

Since the work presented in this thesis is built upon Probability Theory, I shall devote the first part of this Chapter to present the basic definitions and properties about probability that will be used throughout. I will also discuss the different interpretations given to this strongly controversial concept. For a further and more rigorous discussion about probability theory and algebra of events see Stirzaker [2003]; Borkar [1995].

2.1.1 The concept of probability: definition and interpretations

One can define experiment as a procedure done without being sure of the end result of it. These final results are called the *outcomes* of the experiment and will be denoted by ω. The set of possible outcomes is the so-called *sample space* and will be denoted by Ω. Any subset A of the sample space Ω is called *event*. Each event can be associated to a binary question about the outcome. For example, the question "Does the outcome have the property a?" will correspond to the subset of all the elements of Ω that present this property. Events can be combined with the logical operators *and*, *or* and *not*, which correspond respectively to the union, intersection and complementary of the subsets. These combinations allow for more complex questions like "Does the outcome have the property a and/or have (not) the property b?".

With this at hand, we can now define mathematically the concept of probability: probability is a function, defined on the set of events of Ω, that fulfils the following properties:

1. Non-negativity: $P(A) \geq 0 \ \ \forall A \in \Omega$;

2. Unitarity: $P(\Omega) = 1$;

3. Aditivity: for any pair $\{A, B\}$ of mutually exclusive events ($A \cap B = \emptyset$) we have $P(A \cup B) = P(A) + P(B)$;

where we have denoted the probability of event A by $P(A)$. A trivial consequence is that $P(A) \leq 1$ for any event $A \in \Omega$. From the mathematical definition above, the probability of an event is a measure of the chance to occur that this event has. But how do we translate "chance" into a number? What is the interpretation of this probability? There is controversy around the meaning of probabilities. In fact, almost every author interprets probabilities in his own way, but we can distinguish two major schools: Bayesianism and Frequentism. The principal difference of interpretation is the role of an *agent*, i.e., the individual that assigns probabilities to events.

The Bayesian approach is a subjective interpretation. It considers that probability codifies the information about the events that the agent has. It represents the *state of belief or knowledge* the agent has about the experiment. Therefore, probabilities are not fixed, they can evolve whenever the agent gets new information. The numerical value can be understood as a measure of the agent disposition to bet in favour of the event. In this sense, different agents can assign different probabilities to the same experiment. Despite the fact that there is no "true probability" and therefore no "right choice" of probability assignments, some will be more reasonable than others in the sense that they represent better the knowledge of the agent. In order to reasonably update the probability, the Bayes rule, which gives the name to this approach, can be applied. It states how the probability of an event may be modified after gathering new information, in terms of the "old" probability of the event in question and the probability of having obtained the gathered information. A formal expression for the Bayes rule is given below.

On the other hand, there are some other interpretations based on objectivity. Namely, they consider the probabilities as physical properties of the systems or *states of reality*. In these approaches the numerical value of probability has definitions like "the ratio between favourable cases and the equally probable possible cases" or "the frequency that the event will have if we repeat independently the experiment infinitely many times", hence the term Frequentism. Despite the objective appearance of these definitions, the impossibility of defining "equally probable cases" or on repeating independently the experiment imply also some subjective elements. A very illustrative example of the problems associated to the Frequentist approach is asking what is the probability of the Sun exploding tomorrow. Before entering deeper in the implications that the choice of approach has in parameter estimation problems I discuss some probability and random variable concepts.

2.1.2 Random variables and Probability Distributions

A *random variable* $X : \Omega \to E$ is a function that associates for every outcome ω, a real number $X(\omega)$ from the set E. If the set E is finite (or infinitely countable) the variable is called a *discrete random variable*. One can associate a probability to each value $x_j \in E$ of X by the trivial relation $p(x_j) = P\left(X^{-1}(x_j)\right)$. The sub-indices notation will be used to label the different values of a discrete random variable. The function $p(x_j)$ is called *Probability Mass Function* and has to be interpreted as the probability of the random variable taking the value x_j.

On the contrary, if the set E is uncountable the random variable is called *continuous*. In this case, due to the large amount of outcomes, each of them has an infinitely small probability. Therefore, the function $f(x) = P\left(X^{-1}(x)\right)$, called the *Probability Density Function*, can not be interpreted directly as a probability. Instead, only events (i.e., subsets of E) have finite probabilities, which take values $P(A \subseteq E) = \int_A f(x)\,dx$. In what follows I will refer to both probability mass functions and probability density functions with the term *Probability Distribution Function* (PDF) and they will be denoted as $p(x)$.It should be clear from the context whether it refers to the discrete or continuous case.

It is worth noticing that the controversy in the definition of probability is translated to the concept of random variable. While the objective view of probability considers as random variables only those quantities that can have different realisations, from the subjective approach any quantity with an unknown value (by the agent) can be treated like that.

Some experiments might be described using a set of n variables. In these cases, the outcomes will be represented by n random variables $X^{(1)}, \ldots, X^{(n)}$, which can be discrete, continuous or a mixture of the two types. In these cases one might define the *joint probability distribution* function as $p(x^{(1)}, \ldots, x^{(n)})$, which contain all the information about the random variables and gives the probability that the variables take the values $\{x^{(1)}, \ldots, x^{(n)}\}$ jointly. In the situations that involve a wider set of random variables but the attention is limited to a reduced number m of those variables, the distribution of the variables of interest can be computed by summing (or integrating) the rest of the variables over all their possible values $x^{(i)} \in E_i$, i.e.,

$$p(x^{(1)}, \ldots, x^{(m)}) = \sum_{x^{(m+1)}} \cdots \sum_{x^{(n)}} p(x^{(1)}, \ldots, x^{(n)}) \,. \tag{2.1}$$

This new distribution is called the *marginal* probability distribution. In doing this marginalisation the information about the $m - n$ marginalised variables is traced out.

Suppose now an experiment that is described by a pair X, Y of random variables with a joint PDF $p(x, y)$. Suppose further that we are told that the second variable has resulted in a value y_0. We can then ask what are the PDF for the first variable with under this condition. The probability distribution then can be updated with the called *conditional probability*

$$p(x \mid y_0) = \frac{p(x, y_0)}{p(y_0)} \tag{2.2}$$

where the denominator $p(y_0) > 0$ ensures that the distribution is normalised. This "new" distribution, or *conditional* distribution, has the same mathematical properties, and hence it must be considered the PDF in the "new world" resulting from the occurrence of the event represented by the realisation of the variable Y taking the value y_0. It may happen that any realisation of a variable Y does not affect the information we have about the variable X, i.e., $p(x \mid y) = p(x) \; \forall y \in E_Y$. From Eq. (2.2) it is easy to see that this implies that the joint probability satisfies $p(x, y) = p(x)\, p(y)$. When this holds the variables X and Y are said to be *independent* random variables. The concept of independence is crucial in the Frequentist definition of probability, inasmuch as the concept of frequency of an event depends on the possibility of the identical repetition of the experiment, that is, independently of the past realisations of it. Moreover, the objective definition of probability can yield to situations in which two or more variables with an unknown PDF are modelled as independent using symmetries of the system. But the realisation of one of them increased the information about the others. Therefore, from the subjective definition, this new information will modify the probability of the remaining variable in such a way that they can not be considered independent. In order to discuss this fact in a deeper way, I need to introduce two important theorems: the Bayes rule and the de Finetti representation theorem.

2.1.3 Bayes rule and de Finetti theorem

From the definition of conditional probability in (2.2), the joint distribution of two variables can be written as $p(x, y) = p(y \mid x)\, p(x) = p(x \mid y)\, p(y)$. From this relation one can derive the *Bayes rule*, i.e.,

$$p(y \mid x) = \frac{p(x \mid y)\, p(y)}{p(x)} = \frac{p(x \mid y)\, p(y)}{\sum_y p(x \mid y)\, p(y)}. \tag{2.3}$$

This rule can be simply interpreted as a relation between the proportion $p(y \,|\, x)$ of events labelled by y out of the events with a value x, and its counterpart $p(x \,|\, y)$. But within the subjective approach, which receives its name from it, has also a more profound implication: consider a system characterised by two random variables X and Y with a known conditional probability $p(x \,|\, y)$. Before performing any experiment the agent may have some prior knowledge about the possible value of Y, and hence a *a priori* PDF $p(y)$ can be defined. When an experiment is performed and some data x is acquired about the variable X, this new information affects also the knowledge the agent has about Y. Therefore, the PDF must be updated to incorporate it. The Bayes rule tells us that the resulting *a posteriori* PDF is proportional to the a priori PDF $p(y)$ and the conditional distribution $p(x \,|\, y)$. The denominator $p(x)$ ensures that the new distribution is normalised, i.e. $\sum_y p(y \,|\, x) = 1$. Summarising, the Bayes theorem gives us the recipe about how to update our knowledge when we are given some data and it is the cornerstone of the Bayesian methods of parameter estimation.

With the Bayes rule at hand, I now present the de Finetti representation theorem. Its goal is to show light on the concept of independence of events and its repeatability, which is a crucial point in the difference between the two approaches presented. Let me consider the following example[de Finetti, 2008]: we are given N urns labelled by A_i with $i = 1, 2, \ldots, N$. Inside each urn there are black and white balls. The proportion q_i of black balls in each urn A_i is known. An agent chooses secretly an urn and I randomly take one ball, look its color and put it back into the urn. I repeat this process m times. What is the probability of getting a black ball in each round? Or more specifically, are the events of getting a black ball in each shot independent? From an objective point of view, all the events are independent, due to the fact that the reposition ensures that in each shot the proportion of black balls is the same, although unknown. On the contrary, if we take the subjective interpretation, a prior probability is assigned to each urn, taking into account any information we have about the process of election that the agent had followed. Every time we take a ball from the urn, the probability of which urn has been chosen can be updated by the Bayes rule (2.3). Therefore, the probability of getting a black ball in the $k-$th shot will depend on all the previous shots, so the events can not considered independent within this approach. Nevertheless, since only the proportion of black/white balls but not the order in which are taken affects the a posteriori joint probability, it will be invariant under any permutation of the sequence of balls taken. This example yields another important concept related to random variables, proposed by de Finetti [De Finetti, 1990]: a sequence is said to be *exchangeable* if any permutation π of the variables conforming it has the same joint probability distribution, i.e., $p(x_1, x_2, \dots) = p(x_{\pi(1)}, x_{\pi(2)}, \dots)$. Furthermore, if we can extend the sequence adding extra random variables and the sequence remains exchangeable, then is said to be *infinitely/unbounded exchangeable.* De Finetti proved that the joint probability for an infinitely exchangeable sequence $\{x_1, x_2, \dots\}$ is a "mixture" of distributions for sequences of independent random variables, i.e.

$$p(x_1, x_2, \dots, x_m) = \sum_i q_i \prod_k p_i(x_k) \tag{2.4}$$

where q_i must be interpreted as probabilities are hence $\sum_i q_i = 1$. This result is given the name of *de Finetti's representation theorem.* In the example of the urns, consider the random variables X_k that refers to the kth shot and can take the values 1 (black ball) or 0 (white ball). The joint probability of getting a sequence $\{x_{n+1}, \dots, x_m\}$ given that the previous balls taken are described by the sequence $\{x_1, x_2, \dots, x_n\}$ can be written as

$$p(x_{n+1}, \dots, x_m \,|\, x_1, \dots, x_n) = \sum_{i=1}^{N} p(A_i \,|\, x_1, \dots, x_n)\, p(x_{n+1}, \dots, x_m |\, A_i) \tag{2.5}$$

where $p(A_i \,|\, x_1, \dots, x_n)$ is the probability that the agent has taken the urn A_i given that the first n balls taken are $\{x_1, x_2, \dots, x_n\}$, and $p(x_{n+1}, \dots, x_m |\, A_i) = q_i^s (1 - q_i)^{m-n-s}$ is the probability of getting $\{x_{n+1}, \dots, x_m\}$ conditioned to the urn A_i, which only depends in the number of black balls $s = \sum_{k=n+1}^{m} x_k$. Furthermore, de Finetti also shows that in the limit of a large sample, the proportion of black balls $\sum_{k=1}^{m} x_k/m$ will tend, almost surely, to the concrete value q_i. This is known as the de Finetti strong law of large numbers. Given this value q_i, the sequence of observations x_i is said to be *conditionally* independent. This results can be generalised to a more general cases where an unknown source from a given collection produce sequences of independent and identically distributed events. The collection of sources can even be a infinite and hence labelled by a continuous variable. In this case, the sum in (2.4) has to be changed by an integral.

The de Finetti representation theorem motivates the use of *parametric statistical models* to solve statistical inference problems, yet it is not actually used in general for its implementation. In this kind of models, the stochastic behaviour of the system is modelled by some parametric family $\{p(\boldsymbol{x}|\phi) \,|\, \phi \in \Phi\}$, where ϕ is a random variable that codifies the lack of knowledge about the system. This could be the case, for instance, if one wants to infer which urn the agent had taken in the above example, with A_i taking the role of ϕ. After acquiring some new information, the PDF of ϕ is updated by means of the Bayes rule 2.3. However, instead of giving the mixture of distributions as in Eq. 2.4, some value for the parameter is inferred from $p(\phi_i|\, x_1, \ldots, x_m)$. Parameter estimation is discussed with more detail in the next chapter.

2.2 Quantum preliminaries

One of the main goals of this work is to design optimal measurements to reveal some properties of quantum systems. To this end a mathematical description for the state of a system and for the measurements is needed.

2.2.1 The quantum state

The term quantum state refers to a mathematical description of a physical system that is governed by the laws of quantum mechanics. In contrast to classical mechanics, we can not associate a fixed value to all the physical properties of a quantum systems. For the classical systems we can know all the physical magnitudes with certainty, except for the technological limitations in the measuring devices, and therefore they can in principle be distinguished perfectly from each other. In contrast, the quantum states are describe as wave functions, from which one can extract the probability distributions of the possible measurement outcomes [Nielsen and Chuang, 2010]. In fact, as it will discussed in Section 2.2.7, the set of possible outcomes of a measurement depends only on the measurement itself and two different quantum states differ only on how those values are distributed. As a consequence, quantum states can not be distinguished in general by means of measurement. Due to this stochastic nature, the controversy on the interpretation of probabilities translates to quantum states with a debate about the reality of quantum states: are they real or just a description of our knowledge of them?[Plotnitsky and Khrennikov, 2015].

To every physical system is associated a complex vector space with an inner product, a Hilbert space $\mathcal{H}$. Each unit column vector (or *ket*) $|\psi\rangle \in \mathcal{H}$ is called a *state vector*, and it is the mathematical representation of the system state. Its dual (or *bra*) is expressed as $\langle\psi|$. The absolute value of the inner product between two state vectors $|\langle\psi|\varphi\rangle|$, also called its overlap, gives us a measure of how close they are. Since by definition the state vector $|\psi\rangle$ are unit vectors, what is usually called *normalisation* condition, the inner product with itself $\langle\psi|\psi\rangle = 1$. On the contrary, if a pair quantum states are distinguishable, they will be represented by orthogonal state vectors, i.e., with null overlap $|\langle\psi|\varphi\rangle| = 0$. In general, a quantum state can be written as a complex linear combination of a set of other states, i.e.

$$|\psi\rangle = \sum \alpha_i \, |\varphi_i\rangle \, . \tag{2.6}$$

In this sense, the state $|\psi\rangle$ is said to be a *superposition* of the states $\{|\varphi_i\rangle\}$ with the complex numbers α_i being its *amplitudes.* In the case that the set $\{|\varphi_i\rangle\}$ form an orthonormal base for $\mathcal{H}$, the normalisation condition implies $\sum |\alpha_i|^2 = 1$.

But we might not know exactly the state of a system. Instead, we could only know that the system state is one of an ensemble of possible states$\{|\psi_i\rangle\}$, with probabilities $\{p_i\}$ associated to each one. In this cases, the state is described by a *density matrix*, i.e.

$$\rho = \sum_i p_i \, |\psi_i\rangle\langle\psi_i| \, . \tag{2.7}$$

In general, a density matrix is a unit-trace hermitian semidefinite operator, i.e. $\rho \geq 0$ and $\operatorname{tr}\rho = 1$.When the operator is of rank 1the state is called *pure state*, and it represents the state of maximal knowledge about the system. Conversely, higher-rank matrices correspond to *mixed states*, and represent those cases in which the agent has only a partial knowledge about the system. Note that different ensembles of states can produce the same density matrix, therefore, given a density matrix an ensemble cannot be assigned uniquely. Nevertheless, there is a particular ensemble decomposition in which all its the states are orthogonal to each other. This ensemble $\{|\psi_i\rangle \, , p_i\}$ is called the *spectral decomposition*, and the states and probabilities are the eigenvectors and eigenvalues of the density matrix respectively.

2.2.2 Qubits

The simplest and paradigmatic case, which is the main case studied in this thesis, is the two-dimensional Hilbert space. Any two-dimensional state, or *qubit* state, can be expressed as a linear combination of the so-called *computational basis*$\{|0\rangle \, , |1\rangle\}$. Since only the relative phase between amplitudes has a physical meaning, and together with the normalisation condition, the state of a qubit can be parametrised with only two parameter, i.e.

$$|\psi\rangle = \cos\left(\theta/2\right)|0\rangle + \mathrm{e}^{-\phi/2}\sin\left(\theta/2\right)|1\rangle \tag{2.8}$$

with $0 \leq \theta \leq \pi$ and $0 \leq \phi \leq 2\pi$. With this parametrisation, each pure qubit state can be mapped to a point at the surface of an sphere. Furthermore, each qubit density matrix can be written as

$$\rho = \frac{1}{2}\left(\mathbb{1} + \boldsymbol{b}.\boldsymbol{\sigma}\right) \tag{2.9}$$

where $\boldsymbol{\sigma} = (\sigma_x, \sigma_y, \sigma_z)$ with σ_i being the Pauli matrices

$$\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{2.10}$$

and the vectors $\boldsymbol{b} = \langle \boldsymbol{\sigma} \rangle$ are called *Bloch vectors*. In order to maintain the

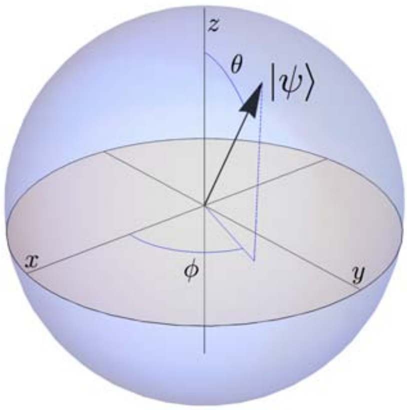


Figure 2.1. Representation of the Bloch sphere.

positivity of ρ, the norm of the Bloch vector, also named the *purity* of the state, must satisfy $r = |\boldsymbol{b}| \leq 1$. Therefore, the possible density matrices are in one-to-one correspondence with the points of an sphere, which is called the *Bloch sphere* (see Fig.2.1). The pure states correspond to those points at the surface ($r = 1$) of the Bloch sphere, while the points in the interior ($r < 1$) tally with the mixed states. The center of the sphere ($r = 0$) correspond the completely mixed state $\rho = \mathbb{1}/2$ that represents the minimal-knowledge state of a qubit system. Note in passing that any vector with $r < 1$ can be expressed as many linear combinations of unit vector. This manifests the idea expressed in the last section of the non uniqueness of ensemble decomposition of mixed states. Nevertheless, one can always decompose the Bloch vector as $\boldsymbol{b} = \frac{1+r}{2}\boldsymbol{b}_+ + \frac{1-r}{2}\boldsymbol{b}_-$ where $\boldsymbol{b}_+ = -\boldsymbol{b}_- = \boldsymbol{b}/r$ are the Bloch vectors of the eigenstates of its spectral decomposition.

2.2.3 Composite systems

In a wide number of situations the system under study is a composite system comprised of two or more subsystems. The state spaces $\mathcal{H}_i$ of the n subsystems are combined through tensor product in order to form the state space of the global system $\mathcal{H}^G = \mathcal{H}^{(1)} \otimes \mathcal{H}^{(2)} \otimes \cdots \otimes \mathcal{H}^{(n)}$. Therefore, the state vector $|\Psi\rangle$ of the global system can be expressed as a superposition of states like $|e_{i_1}\rangle \otimes |e_{i_2}\rangle \otimes \cdots \otimes |e_{i_n}\rangle$, i.e.,

$$|\Psi\rangle = \sum_{i_1, i_2, \ldots, i_n} \alpha_{i_1, i_2, \ldots, i_n} |e_{i_1}\rangle \otimes |e_{i_2}\rangle \otimes \cdots \otimes |e_{i_n}\rangle \tag{2.11}$$

where the sets $\{|e_{i_k}\rangle\}$ are basis of each subsystem $\mathcal{H}^{(k)}$. As in the case of non-composite system, density matrices can be defined through the expression in Eq.(2.7), and they will be all the unit-trace semidefinite operators on $\mathcal{H}^G$.

When dealing with composite systems one can ask about the correlations between its subsystems. Quantum mechanics allows stronger correlations between systems than classical mechanics. We have to distinguish between two types of pure composite systems. In some cases the global state can be factorised in terms of the states of the subsystems, i.e., there exist states $\left|\psi^{(k)}\right\rangle$ for each subsystem $\mathcal{H}^{(k)}$ such that $|\Psi\rangle = \left|\psi^{(1)}\right\rangle \left|\psi^{(2)}\right\rangle \ldots \left|\psi^{(n)}\right\rangle$.[1] Those states are referred to as *product states* and they have no correlation between subsystems. An analogy can be made between product states and independent random variables. In contrast, when no factorisation is possible and, therefore, the subsystems are quantumly correlated, the states are named *entangled states*.

Since mixed states comprise quantum states and classical probability distribution, they can combine also classical and quantum correlations. As in the pure case, when there is no correlation the states are called product states and they can be expressed as tensor product of density matrices, i.e., $\rho^G = \rho^{(1)} \otimes \rho^{(2)} \otimes \ldots \rho^{(n)}$. The composite states that only have classical correlation can be decomposed as convex combination of product states as

$$\rho^G = \sum_k p_k \, \rho_k^{(1)} \otimes \rho_k^{(2)} \otimes \ldots \rho_k^{(n)}. \tag{2.12}$$

When such a decomposition is possible the global state is named *separable state*. Finally, a state that is not separable is called entangled and it presents quantum correlations. It is not at all easy to determine if an state is separable or entangled and it is a current field of research [Gühne and Tóth, 2009].

2.2.4 Identical copies and block decomposition

A very usual paradigm found in quantum estimation tasks is when the system is compounded of identically prepared and uncorrelated subsystems. Hence,

[1]Sometimes the tensor product symbol $\otimes$ is omitted between vectors of different subsystems, or even the notation $\left|\psi^{(1)}, \psi^{(2)}, \ldots, \psi^{(n)}\right\rangle$ is used when no confusion arises.

the states of these systems should be invariant under permutations. This symmetry can be used to derive a particular decomposition that ease the computations. This decomposition was introduced by Vidal *et al.* [1999]; Cirac *et al.* [1999] (See also Bagan *et al.* [2006] for a more detailed presentation).

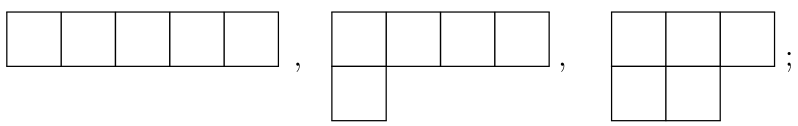
The global state of n qubit systems identically prepared in a state ρ is written as $\rho^{\otimes n}$. The state remains unchanged under the exchange of a pair of subsystems, thus it is invariant under the action of the symmetric group S_n. One can use the group S_n to define the basis $\{|j,m,\alpha\rangle\}_{m=-j}^{j}$ for the $(2j+1)$-dimensional $SU(2)$-invariant subspaces of $(\mathbf{\frac{1}{2}})^{\otimes n}$. Each of these subspaces is labelled by the pair of indices $\{j,\alpha\}$, where $j = \mathrm{mod}_2(j)/2, \ldots, n/2$. The extra index $\alpha = 1, \ldots, \nu_j$ keeps track of the equivalent irreducible representations $\mathbf{j}$ of $SU(2)$ with multiplicity ν_j. The relation between the invariant subspaces and the tensor product representations can be expressed as

$$(\mathbf{\tfrac{1}{2}})^{\otimes n} = \bigoplus \nu_j \, \mathbf{j} \tag{2.13}$$

Therefore, the global state density matrix can be decomposed into a direct sum of $(2j+1)$-dimensional square matrices $\rho^{(j)}$ for each invariant subspaces, i.e.

$$\rho^{\otimes n} = \bigoplus_j \nu_j \, \rho^{(j)}. \tag{2.14}$$

Notice that in each invariant subspace with equal j the state takes the same form independently of α. In order to get a closed expression for ν_j, note that the sets of vectors with fixed j and m form a basis for each of the ν_j-dimensional irreducible representations of the symmetric group S_n. Each of these representations of S_n is associated to a Young diagram of shape $(n/2+j, n/2-j)$, i.e., a set of n boxes arranged in two left-justified rows of lengths $n/2+j$ and $n/2-j$, with the additional condition that the length of any row must be of the same length or shorter than the one above, i.e. $2j \geq n-2j$. For instance, the three possible Young diagrams with $n=5$ that can be constructed are



with $j = 5/2$, $j = 3/2$ and $j = 1/2$ respectively. By filling the Young diagrams with the numbers $1, 2, \ldots, n$ in each box, one might construct the so-called standard Young tableaux, which one of them is associated a different value α. Hence, the multiplicity ν_j corresponds to the number of different standard Young tableaux for this particular value of j. To construct them, the diagram must be filled with the numbers in increasing order from left to right and from top to bottom. Following the above example, the possible Young tableaux for $j = 1/2$ are

$$\begin{array}{|c|c|c|}\hline 1&3&5\\\hline 2&4 \\\cline{1-2}\end{array}\;, \quad \begin{array}{|c|c|c|}\hline 1&3&4\\\hline 2&5 \\\cline{1-2}\end{array}\;, \quad \begin{array}{|c|c|c|}\hline 1&2&5\\\hline 3&4 \\\cline{1-2}\end{array}\;, \quad \begin{array}{|c|c|c|}\hline 1&2&4\\\hline 3&5 \\\cline{1-2}\end{array}\;, \quad \begin{array}{|c|c|c|}\hline 1&2&3\\\hline 4&5 \\\cline{1-2}\end{array}\;.$$

Given a diagram with n boxes defined by the index j, a simple way to compute how many of these tableaux can be constructed is as follows: associate a value of $i_k = +1/2$ ($i_k = -1/2$) to the boxes in the top (bottom) row with k being the corresponding number in it. This particular association is made by considering the multiple-qubit system as spin-1/2 coupled particles, and considering a ordered coupling, i.e. we first couple the first two particles, the a third one is coupled with the first pair and so on. [2] Following this intuition, construct now the sequence of partial sums, i.e. $\boldsymbol{s} = \{s_0, s_1, \ldots, s_n\}$, where $s_0 = 0$ and $s_r = \sum_{k=1}^{r} i_k$ for $1 \le r \le n$. Each s_r corresponds to the intermediate angular momentum when r particles are coupled. For example, the Young tableau $\begin{array}{|c|c|c|}\hline 1&2&5\\\hline 3&4 \\\cline{1-2}\end{array}$, leads to the sequence $(+1/2, +1/2, -1/2, -1/2, +1/2)$, and hence $\boldsymbol{s} = (0, 1/2, 1, 1/2, 0, 1/2)$. It is straightforward to see, by construction of the Young tableaux, that $\boldsymbol{s}$ has non negative entries and $s_n = j$. This agrees with the angular momentum interpretation. Therefore, the number of different Young tableaux related to a given diagram is equal to the number of positive sequences $\boldsymbol{s}$ with $s_n = j$. To simplify the computation, I first count how many sequences can be constructed with $n/2 + j$ positive steps and $n/2 - j$ negative steps, independently of the positivity condition. Then, the multiplicity can be computed by subtracting from them the ones with negatives entries. Although at first sight it seems complicated, there is a clever way to compute this number: Notice that each of the "negative" sequence must take the value $-1/2$ at least at one point. If one vertically reflects with respect to $-1/2$ the points before the first one that take this value (see Fig. 2.2), one gets a sequence starting with -1 and ending with j. This new sequence has one less negative step than the original one. One can see that there is a one-to-one correspondence between the sequences starting with -1 and the "negative" sequences starting with 0. This is the so-called reflection theorem of random walks [Stirzaker, 2003]. Hence, the computation of the two terms involved is reduced to counting the ways to distribute the negative steps into each sequence, given by the binomial coefficients $\binom{n}{n/2-j}$

[2]There are different ways to define the different invariant subspaces, and hence the indices α. However, in the identical-copies case they are equivalent since the state has identical expressions $\rho^{(j)}$ in each invariant subspace due to the invariance under permutations.

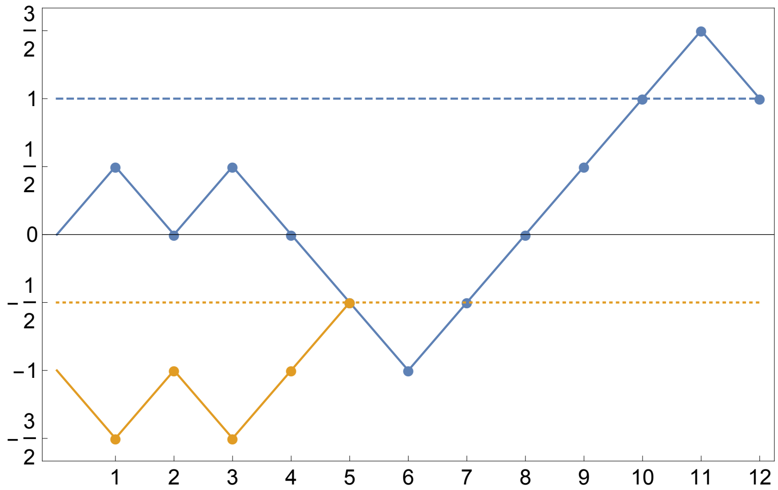


Figure 2.2. Example of sequence with $n = 12$ going from 0 to $j = 1$ in steps of $1/2$ with negative values (blue line). Reflecting with respect $-1/2$ (dotted line) the points on the left of the first negative one, the resulting sequence (orange line) starts at -1 and ends at j .

and $\binom{n}{n/2-j-1}$ respectively. Therefore,

$$\nu_j = \binom{n}{n/2-j} - \binom{n}{n/2-j-1}. \tag{2.15}$$

It only remains to compute an explicit form for the blocks $\rho^{(j)}$. To do so, consider the single system density matrix with Bloch vector $\boldsymbol{b} = \boldsymbol{z}$. The density matrix of such state is diagonal and reads

$$\rho = \tfrac{1}{2}[(1+r)\,|0\rangle\langle 0| + (1-r)\,|1\rangle\langle 1|], \tag{2.16}$$

where r is its purity. Thus, the tensor product leads to diagonal global density matrix too, and only the terms corresponding to $|j,m,\alpha\rangle\langle j,m,\alpha|$ should be computed. Since all the blocks with the same j have identical expressions and, hence, $\rho^{(j)}$ are independent of the particular definition of the subspaces we take, it is convenient to consider the basis for the first subspace[3]

$$|j,m,\alpha=1\rangle = |j,m\rangle \left|\Psi^-\right\rangle^{\otimes n/2-j} \tag{2.17}$$

On one hand, the last $n-2j$ subsystems are, by pairs, in the antisymmetric state $|\Psi^-\rangle = (|01\rangle - |10\rangle)/\sqrt{2}$ and thus each pair contributes with a factor $\frac{1}{4}(1-r^2)$ to each matrix element. On the other hand, the first $2j$ subsystems

[3]The other subspaces can be constructed by imposing orthogonality.

are in the symmetric state $|j,m\rangle$ with $j+m$ subsystems in the state $|0\rangle$ and $j-m$ in the state $|1\rangle$. With this, the matrices $\rho^{(j)}$ take the form

$$\rho_{\boldsymbol{z}}^{\otimes n} = \sum_{j,\alpha} \left(\frac{1-r^2}{4}\right)^{n/2-j} \sum_{m} \left(\frac{1+r}{2}\right)^{j+m} \left(\frac{1-r}{2}\right)^{j-m} |j,m,\alpha\rangle\langle j,m,\alpha| \tag{2.18}$$

Notice that pure states ($r=1$) have only projection to the symmetric invariant subspace with $j=n/2$. For arbitrary Bloch vectors $\boldsymbol{b}$, it suffices to rotate the basis $|j,m,\alpha\rangle$ with the Wigner matrices $D(\boldsymbol{b})$ [Edmonds, 1957] . Hence, a general expression for the global state can be cast as

$$\begin{aligned}
\rho^{\otimes n} &= D(\boldsymbol{b})\, \rho_{\boldsymbol{z}}^{\otimes n}\, D^{\dagger}(\boldsymbol{b}) = \\
&= \sum_{j,\alpha} \left(\frac{1-r^2}{4}\right)^{n/2-j} \sum_{m',m} |j,m,\alpha\rangle\langle j,m',\alpha| \times \\
&\times \sum_{k=-j}^{j} \left(\frac{1+r}{2}\right)^{j+k} \left(\frac{1-r}{2}\right)^{j-k} \mathcal{D}^{(j)}_{m,k}(\boldsymbol{b})\, \mathcal{D}^{*(j)}_{m',k}(\boldsymbol{b})
\end{aligned} \tag{2.19}$$

where $\mathcal{D}^{(j)}_{m',m}(\boldsymbol{b}) = \langle j,m',\alpha|\, D(\boldsymbol{b})\, |j,m,\alpha\rangle$ are the elements of the Wigner matrix $D(\boldsymbol{b})$.

The permutation invariance of the identical copies states translates to the sequence of measurement outcomes being exchangeable. Furthermore, if more identically prepared subsystems are measured, the resulting sequence of outcomes remains exchangeable. Accordingly, the state produces an infinitely exchangeable sequence. In analogy with the classical case, a quantum version of the de Finetti theorem can be obtained. First, a composite state with n subsystems ρ_n is said to be a exchangeable if it is symmetric, i.e., invariant under any permutation of its subsystems, and if, for any $m>0$ there is a symmetric state ρ_{n+m} of $n+m$ subsystems such that the marginal state for the n subsystems is ρ_n, i.e.

$$\mathrm{tr}_{\,m}\rho_{n+m} = \rho_n \tag{2.20}$$

where the partial trace is taken only on the m additional subsystems. This distinction between symmetric and exchangeable states has its analogy in the classical case between exchangeable and infinitely exchangeable sequence. However, in the quantum case the possibility of extend the system plays the role of excluding the entangled states. The quantum de Finetti theorem [Caves *et al.*, 2002] states that any exchangeable state of n subsystems can be uniquely written as a separable state of the form

$$\rho_n = \int_{\mathcal{D}} p(\rho)\, \rho^{\otimes n}\, d\rho \tag{2.21}$$

where $p(\rho)$ is a normalised positive function and $d\rho$ is a suitable measure over the set of all possible density operators $\mathcal{D}$. The function $p(\rho)$ can be interpreted as the probability distribution representing the knowledge the agent has about the state of the subsystems. Like in the classical case, the de Finetti theorem ensures that two different agents, with different initial knowledges and after measuring $m < n$ subsystems, will converge in the states assigned to the remaining $n - m$ subsystems as m grows.

2.2.5 Evolution

So far I have described the quantum states as instant pictures of the systems. Now let me talk about its dynamics. The time evolution of a *closed system* of dimension d, i.e. a system without any interaction with the environment, is described by an hermitian operator H, named Hamiltonian, through the Shcrödinger equation

$$H\,|\psi(t)\rangle = i\hbar\,\frac{\partial}{\partial t}\,|\psi(t)\rangle \tag{2.22}$$

where i is the imaginary unit and $\hbar$ is the Planck constant. For simplicity I will take the units such that $\hbar = 1$. To understand the effect of this equation, let me focus first to the eigenstates of the Hamiltonian. This eigenstates are defined through $H\,|e_k\rangle = e_k\,|e_k\rangle$, where the eigenvalues e_k denote the energy of the eigenstates $|e_k\rangle$. Plugging this into the Schrödinger one gets $|e_k(t)\rangle = \mathrm{e}^{-i\,e_k\,t}\,|e_k(t=0)\rangle$. That is to say, if the system is in an eigenstate of the Hamiltonian, it will remain unchanged since a global phase has not any physical effect. In general, the state will be in a superposition of eigenstates $|\psi\rangle = \sum_k \alpha_k\,|e_k\rangle$ and, because of linearity, each one will get a relative phase, i.e., $|\psi(t)\rangle = \sum_k \alpha_k\,\mathrm{e}^{-i\,e_k\,t}\,|e_k\rangle$. Moreover, the linearity allows us to express the time evolution of the state with a unitary operator which takes a diagonal form in the eigenbasis of the Hamiltonian, i.e.

$$U_t = \sum_{k=1}^{d} \mathrm{e}^{-i\,e_k\,t}\,|e_k\rangle\!\langle e_k| \quad \Rightarrow \quad |\psi(t)\rangle = U_t\,|\psi(0)\rangle \tag{2.23}$$

This unitary formalism is used in the problems of phase estimation. In this kind of problems the phase to be determined corresponds to the factor $\phi = e_k\,t$ of the exponential. Furthermore, by experimentally tuning the Hamiltonian of a system and controlling the interaction time a concrete phase can be codified.

Since the real-world systems are rarely isolated, we have to deal with *open systems*. For instance, this is the case where some environmental noise affects the dynamics of the system. The time evolution of such systems can not be described by unitary transformations. Instead, the general formalism that allows interaction with the environment is that of quantum operations. Mathematically, a quantum operation is described by a linear map $\varepsilon : M_d \rightarrow M_{d'}$ which transforms density matrices into density matrices. In order to be physically realizable, performing a quantum operation on a state must yield another state. Furthermore, since ρ can be extended by the tensor product with some ancillary system, the extended map must also ensure that its images are states. Mathematically it can be summarised by the following properties:

1. Linearity: $\varepsilon(p_1\rho_1 + p_2\rho_2) = p_1\varepsilon(\rho_1) + p_2\varepsilon(\rho_2)$;

2. Trace preserving: $\mathrm{tr}\,[\varepsilon(\rho)] = \mathrm{tr}\,[\rho]$;

3. Positivity: $\rho \geq 0 \Rightarrow \varepsilon(\rho) \geq 0$;

4. Complete positivity: $\varepsilon \otimes \mathbb{1}_k$ must be a positive map $\forall k \geq 1$, where $\mathbb{1}_k$ is the identity map of dimension k.

The *Stinespring dilation theorem* [Kraus, 1971; Choi, 1975] ensures that for any completely-positive trace-preserving(CPTP) map acting on $\mathcal{H}_d$ can be interpreted as a unitary evolution on an extended closed system with $\mathcal{H}_d \otimes \mathcal{H}_{env}$. As a consequence of this theorem, the quantum CPTP maps can be expressed using the *Kraus operators representation*, i.e. for any linear CP map ε there exists a set of operators $\{A_k\}$ such that $\varepsilon(\rho) = \sum_k A_k \rho A_k^\dagger$, where $A_k^\dagger$ is the hermitian conjugate of A_k. The set of operators must satisfy the constraint $\sum_k A_k^\dagger A_k = \mathbb{1}$ in order to the map to be trace preserving. Since there is no need to explicitly define the properties of the environment, this representation simplifies the calculations in situations where we are only interested in the main system. Moreover, since there is no constraint in the effect of the map in the ancillary system $\mathcal{H}_{env}$, there is also freedom in the election of the set of Kraus operators. In fact, a set of operators $\{B_j = \sum_k u_{j,k}\, A_k\}$, with $u_{j,k}$ being a unitary matrix, are an equivalent Kraus representation for the channel.

2.2.6 Examples of noise models for qubits

In this section I present here some usual qubit noise models. This serves as an exemplification of the quantum operations formalism presented above. It also illustrates the mathematical description of the action of noise in physical systems.

The *bit flip* channel flips the state $|0\rangle$ to $|1\rangle$ and vice-versa with some probability $(1-p)$. It can be represented by the Kraus operators:

$$A_0 = \sqrt{p}\,\mathbb{1} \qquad A_1 = \sqrt{1-p}\,\sigma_x \tag{2.24}$$

where σ_x is the Pauli matrix defined in Eq. (2.10) and $0 \leq p \leq 1$ since it is a probability. The pure states with Bloch vectors in the x direction $|\pm\rangle = \frac{1}{\sqrt{2}}\Big(|0\rangle \pm |1\rangle\Big)$ remain unchanged through this channel. A pure state pointing in any other direction will probabilistically suffer a change, thus the state will become a mixed state. A geometrical interpretation of the effect of this channel can be seen using the Bloch sphere: the channel squeeze the y and z directions by a factor $2p-1$. In other words, in terms of the Bloch vectors the channel can be described as the map

$$\varepsilon : \Big(n_x,\, n_y,\, n_z\Big) \to \Big(n_x,\, (2p-1)n_y,\, (2p-1)n_z\Big) \tag{2.25}$$

In the particular case of $p = \frac{1}{2}$, the resulting Bloch vector points to the x direction. This means that all the information we have about the other two directions vanishes.

The *phase flip* channel has a similar effect than the bit-flip one, but it is the z direction that remains unaffected instead of x direction. It can be understood as a flip of the relative phase with probability $(1-p)$. The Kraus operators are the same that in the previous case except for that the Pauli matrix involved is σ_z.

A third obvious case is the *bit-phase flip channel*, which is a combination of the bit flip and phase flip channels, and corresponds to the channel that leaves the y component of the Bloch vector unchanged. In Table 2.1 can be found a summary of this channels.

The three presented channels have in common that they have a privileged basis that remains unchanged. In general, the effect of these *dephasing* channels is to reduce the quantum coherence between the states of this basis, i.e. it reduces the off-diagonal entries of the state density matrix written in this concrete basis.

Another important channel is the *depolarising* channel. In contrast to dephasing channels, it has no privileged direction. Instead, with probability $(1-p)$ the input state is completely depolarised, i.e., it is changed by the completely mixed state $\mathbb{1}/2$. Thus, the output state of the system after the channel can be expressed as

$$\varepsilon(\rho) = p\,\rho + (1-p)\,\frac{\mathbb{1}}{2}. \tag{2.26}$$

Plugging into this expression the Bloch form in Eq.(2.9), it is straightforward to see that the effect of depolarising channel can be seen as a shrinking of the Bloch vector by a factor p.

Channel	**Explicit expression**	**Kraus operators**
Bit flip	$\varepsilon(\rho) = p\,\rho + (1-p)\sigma_x\rho\,\sigma_x$	$A_0 = \sqrt{p}\,\mathbb{1} \qquad A_1 = \sqrt{1-p}\,\sigma_x$
Phase flip	$\varepsilon(\rho) = p\,\rho + (1-p)\sigma_z\rho\,\sigma_z$	$A_0 = \sqrt{p}\,\mathbb{1} \qquad A_1 = \sqrt{1-p}\,\sigma_z$
Bit-phase flip	$\varepsilon(\rho) = p\,\rho + (1-p)\sigma_y\rho\,\sigma_y$	$A_0 = \sqrt{p}\,\mathbb{1} \qquad A_1 = \sqrt{1-p}\,\sigma_y$
Depolarising	$\varepsilon(\rho) = p\,\rho + (1-p)\frac{\mathbb{1}}{2}$	$A_0 = \sqrt{\frac{1+3p}{4}}\,\mathbb{1}$ $A_i = \sqrt{\frac{1-p}{4}}\,\sigma_i \qquad i = x, y, z$

Table 2.1. Summary of typical noise channels for qubits systems.

Using the relation

$$\frac{\mathbb{1}}{2} = \frac{\rho + \sigma_x\,\rho\,\sigma_x + \sigma_y\,\rho\,\sigma_y + \sigma_z\,\rho\,\sigma_z}{4} \tag{2.27}$$

it is straightforward to obtain the Kraus operators of this channel as depicted in Table 2.1. This representation suggest another interpretation for the depolarising channel: the state is left unchanged with some probability $q = (1+3p)/4$ or it suffers a bit flip, a phase flip or both with probability $(1-q)/3$ each.

2.2.7 Measurement

There is a especial case of interaction that a quantum system can suffer, the one that are realised by an agent in order to extract information about the system, namely a measurement. Two main facts cause quantum measurement be a controversial subject: it reveals the quantum stochastic nature of quantum mechanics and substantially affects the state of the measured system. Following the initial formalism of von-Neumann, to every measurable quantity $\mathcal{Q}$ is associated an *observable*, i.e., a linear self-adjoint operator Q acting on the state space $\mathcal{H}$ of the system. The only outcome values that can be obtained by measuring Q coincide with its eigenvalues λ_k. Let $Q = \sum_k \lambda_k\, P_k$ be the spectral decomposition of Q, with P_k being the projector over the eigenspace related to λ_k. The probability of obtaining a concrete value λ_k from a measurement over the state ρ is given by the Born rule

$$p(\lambda_k|\rho) = \mathrm{tr}\,(P_k\,\rho). \tag{2.28}$$

In other words, the state of the system only specifies the PDF of the measurement outcome. Despite this, if the same measurement if sequentially repeated on a quantum system, the same outcome λ_k will be obtained. This is due to the fact that, when a measurement is performed, the state of the system abruptly changes to a projection over the eigenspace corresponding to the obtained outcome, i.e. the state of the system after obtaining outcome λ_k in the measurement is

$$\rho_k = \frac{P_k\,\rho\,P_k}{\mathrm{tr}\,(P_k\,\rho)}. \tag{2.29}$$

This sudden perturbation of the state is known as the *wave function collapse.*

The projective measurements described above, also known as Projection-Valued Measurements (PVM) or von-Neumann measurements, can have at most as many outcomes as the dimension of the Hilbert state. However, one can increase the number of possible outcomes by letting the system interact with an ancillary system. In doing so, it might appear some information flow between the two systems, which can then be extracted by measuring jointly the composite system. As in the case of quantum channels, using the Stinespring dilation theorem, one can define a more general paradigm for quantum measurements. These generalised measurements, also called *Positive Operator-Valued Measurement* (POVM), are represented by a set of positive semi-definite operators $\{\Pi_k \geq 0\}$ that fulfils the completeness relation $\sum_k \Pi_k = \mathbb{1}$. Each of the elements Π_k corresponds to an outcome of the measurement, and they are no necessarily orthogonal. The probability of obtaining each outcome is given by

$$p(k|\rho) = \mathrm{tr}\,(\Pi_k \rho) \tag{2.30}$$

The fact that in these generalised measurements the ancillary system is traced out implies that they are not necessarily related to any physical quantity, and also that the state after the measurement can not be specified uniquely.

The POVM formalism only captures one aspect of quantum measurements, namely the gathering of information, i.e. the relation between the probe state and the outcome probabilities. It is often important to describe the state of the system after the measurement. Since the measurement process is inherently random the map between input and state after the measurement is described by a stochastic quantum operation, labelled by the different outcomes k:

$$\rho'_{|k} = \epsilon_k(\rho) = \sum_l A_l^{(k)}\,\rho\,A_l^{(k)\dagger}/p(k|\rho) \tag{2.31}$$

where the outcome probability is $p(k|\rho) = \mathrm{tr}\left(\sum_l A_l^{(k)}\,\rho\,A_l^{(k)\dagger}\right) = \mathrm{tr}\,(\Pi_k\rho)$ and we have identified the POVM element $\Pi_k = \sum_l A_l^{(k)\dagger} A_l^{(k)}$. These channels describe the stochastic filtering operations that are widely used in this thesis.

As a final remark, consider a situation where the experimentalist first toss a coin and depending on the result applies one POVM or another. Notice that this kind of probability mixtures of two or more POVMs can also be described as a single POVM. Therefore, the set of all POVMs on a Hilbert space is convex. This can be used the other way around, a complex POVM with a large number of outcomes can be decomposed into a mixture of simpler POVMs [Sentís *et al.*, 2013b].

2.2.8 No-cloning theorem and indistinguishability of quantum states

To end this Chapter, here I present two immediate consequences of the quantum formalism that have a crucial effect and pose a genuine handicap in the estimation and discrimination tasks discussed later in this Thesis.

No-cloning theorem

Consider a composite system of three parties: a data system $\mathcal{H}_d$ in which we are given an unknown state $|\psi\rangle$ that we want to clone into a target system $\mathcal{H}_t$, and an auxiliary system $\mathcal{H}_{anc}$. Initially the system is prepared in the state $|\psi\rangle \otimes |0\rangle \otimes |i\rangle$, where $|0\rangle$ and $|i\rangle$ are arbitrary states. A general cloning machine is described then by a unitary operator U acting on $\mathcal{H}_d \otimes \mathcal{H}_t \otimes \mathcal{H}_{anc}$ transforming the initial state as

$$U\Big(|\psi\rangle \otimes |0\rangle \otimes |i\rangle \Big) = |\psi\rangle \otimes |\psi\rangle \otimes |f\rangle \, . \tag{2.32}$$

where $|f\rangle$ is some arbitrary state in which the ancillary system result. But we want the machine to be universal, so for a different initial state

$$U\Big(|\varphi\rangle \otimes |0\rangle \otimes |i\rangle \Big) = |\varphi\rangle \otimes |\varphi\rangle \otimes |g\rangle \, . \tag{2.33}$$

where the ancillary system can result to another state $|g\rangle$. By taking the inner product of the last two equation we get

$$\langle\psi|\varphi\rangle = \langle\psi|\varphi\rangle^2 \langle f|g\rangle \tag{2.34}$$

Taking into account that each of the inner product are upper bounded by 1, the two only solutions are $\langle\psi|\varphi\rangle = 0$ and $\langle\psi|\varphi\rangle = 1$. Therefore, a cloning machine can only be implemented to clone pairs of orthogonal states and no universal one is possible.

Indistinguishability of non-orthogonal quantum states

In order to perfectly distinguish two quantum states $\{\psi_1, \psi_2\}$ we need a POVM with two outcomes $\{\Pi_1, \Pi_2\}$ that identifies unambiguously each state, i.e.

$$p(1|\psi_1) = \langle\psi_1| \Pi_1 |\psi_1\rangle = 1\,; \tag{2.35}$$

$$p(2|\psi_2) = \langle\psi_2| \Pi_2 |\psi_2\rangle = 1\,. \tag{2.36}$$

Let assume Eq. (2.35) as true and see that Eq. (2.36) implies that the states must be orthogonal. From the completeness relation and Eq. (2.35) one can see that $\langle\psi_1| \Pi_2 |\psi_1\rangle = 0$, which in turn implies $\sqrt{\Pi_2}\,|\psi_1\rangle = 0$. Since the Hilbert space spanned by this two states is of dimension 2, one can write $|\psi_2\rangle = \alpha\,|\psi_1\rangle + \beta\,\left|\psi_1^\perp\right\rangle$, where $\left|\psi_1^\perp\right\rangle$ is an state orthogonal to $|\psi_1\rangle$ and $|\alpha|^2 + |\beta|^2 = 1$. With this at hand one can see that

$$\langle\psi_2| \Pi_2 |\psi_2\rangle = |\beta|^2 \left\langle\psi_1^\perp\right| \Pi_2 \left|\psi_1^\perp\right\rangle \leq |\beta|^2 \tag{2.37}$$

Comparing this last expression with Eq. (2.36), we see that in order to $|\psi_1\rangle$ and $|\psi_2\rangle$ be perfectly distinguishable, we must impose $\beta = 1$, which also means that the two states are orthogonal.

No cloning and indistinguishability are closely related, if cloning were possible one could produce infinitely many copies of a state and, by measuring them, perfectly identify the state.

CHAPTER 3

Basics on quantum metrology

In this Chapter I introduce quantum metrology tasks. One of the main goals of metrology is to infer unknown properties from physical systems from gathered data. Some questions arise: which is the best way to gather such information? Which procedure has to be followed in order to extract most efficiently the information contained in the data about the sought property? To answer these questions, one has to firstly fix a criterion to evaluate the possible options that will help to take a decision. We will see that this criterion can have a significant effect in the final choice. In this sense, choosing an estimator or optimising a quantum measurement can be cast as a decision problem.

I begin this Chapter by briefly introducing such decision problems, and then I move to metrology tasks. First I consider classical estimation theory and I discuss how the controversy in the definition of probability presented in the previous Chapter leads to different estimation methods. Secondly, the two paradigmatic quantum metrology problems of quantum state discrimination and quantum parameter estimation are discussed. I close the Chapter by introducing probabilistic quantum state estimation protocols, which is the main topic of this thesis.

3.1 Decision Problems

Decision-making is present in almost any aspect of life. Although in everyday situations it is based in intuition, some problems can be tackled by means of logic and statistical methods. In fact, there are a lot of cases in which the logical and rigorous methods shows that intuition has a powerful A decision problem can be described by $\{\mathcal{A}, \mathcal{E}, \mathcal{C}, U\}$ [Bernardo and Smith, 1994], where

(i) $\mathcal{A} = \{a_i\}$ is a set of possible actions;

(ii) $\mathcal{E} = \{e_{i,j}\}$ is the set of possible outcomes of taking any action $a_i \in \mathcal{A}$. We can decompose $\mathcal{E} = \bigcup_i \mathcal{E}_i$, where each subset $\mathcal{E}_i$ contains the possible outcomes of each action a_i;

(iii) $\mathcal{C} = \{c_{i,j}\}$ is the set of consequences, or decisions, based on the taken action a_i and its resulting outcome $e_{i,j}$;

(iv) and $U : \mathcal{A} \to \mathbb{R}$ is a utility function, or a figure of merit, that assigns a number to each action, depending through its set of consequences. This figure of merit U defines an order of preference in $\mathcal{A}$, and thus, an optimal action to be taken.

For sake of exemplification, consider a physical system with some unknown properties. In this case the set of actions is the possible measurements that can be performed to the system and the set $\mathcal{E}$ contains its possible outcomes. The possible values that can be inferred for the properties under study form the set $\mathcal{C}$. The function $\mathcal{G} : \mathcal{E} \to \mathcal{C}$ that maps each outcome $e_{i,j}$ to an inferred value for the desired quantity is called *guessing rule* or *guessing function*. Finally, a figure of merit U evaluates the mean accuracy of each possible measurement together with the guessing rule.

I will refer to a statistical protocol as a pair made up of a measurement $\mathcal{M}$ (an action a_i with its set of outcomes $\{e_{i,j}\}$) and a guessing function $\mathcal{G}$. In order to find the optimal protocol these two ingredients have to be taken into account when evaluating the figure of merit.

3.2 Classical Parameter Estimation

The aim of classical parameter estimation is to determine a set of unknown parameters $\phi = (\phi_1, \phi_2, \ldots, \phi_k)$ that characterises a given probability density function (PDF) $p(\mathrm{X} \,|\, \phi)$ for a random variable X. The set of actions in this case has only one element: to get a data set $\boldsymbol{x} = \{x_1, x_2, \ldots, x_\nu\}$ consisting on ν realisations of X. Therefore, our goal is to find a guessing function $\hat{\phi}(\boldsymbol{x})$, called in this framework *estimator*, that outputs the most accurate estimate for the parameters ϕ given the data $\boldsymbol{x}$. For the sake of simplicity, I will consider the case with only one parameter.

The problem of estimation can be tackled using different approaches depending on the interpretation of probabilities (see Chapter 2): typically, the Bayesian and the Frequentist approaches. The former considers the probabilities as a representation of the knowledge (or ignorance) about the random variable to which is concerned, and thus the parameter itself is considered as random. Hence, after acquiring more information (e.g. through a measurement) the PDF for ϕ is updated. On the other hand, the Frequentist point of view interprets probabilities as the frequencies of each event in the limit of a infinite sized sample. For this reason, the parameter ϕ is considered to have a fixed value and it is not treated as a random variable, in contrast to Bayesian approach. This difference in interpretation leads to very distinct estimation methods, which are introduced in the following sections.

3.2.1 Bayesian estimation

Due to the Bayesian interpretation of probabilities every unknown quantity can be characterised by a probability distribution. Hence, the parameter ϕ itself is treated as a random variable and *a priori* PDF $p(\phi)$ is assigned to it, representing the initial lack of knowledge of the experimentalist about its value. Therefore, the estimator must be optimised for all possible values of the parameter, taking into account which of these are more probable.

When some data $\boldsymbol{x}$ is obtained, the information we have about the parameter ϕ in general increases. Therefore, the a priori PDF is updated to a posteriori PDF following the Bayes rule (2.3), which now take the form

$$p(\phi \,|\, \boldsymbol{x}) = \frac{p(\boldsymbol{x} \,|\, \phi)\, p(\phi)}{p(\boldsymbol{x})} \tag{3.1}$$

where $p(\boldsymbol{x}) = \int d\phi\; p(\boldsymbol{x} \,|\, \phi)\, p(\phi)$ is the marginal probability of obtaining the data $\boldsymbol{x}$. Since in most cases the gathered data can be separated in conditionally independent parts, i.e. $\boldsymbol{x} = (x_1, x_2)$ with $p(x_1, x_2|\phi) = p(x_1|\phi)\, p(x_2|\phi)$, this last expression can be rewritten as

$$p(\phi\,|x_2, x_1) = \frac{p(x_2|\phi)\, p(\phi|x_1)}{\int d\phi\, p(x_2|\phi)\, p(\phi|x_1)} \tag{3.2}$$

Therefore, Eq.(3.2) can be interpreted as a progressive updating of the PDF for the parameter: initially we have $p(\phi)$; we get the first data x_1 and update the parameter PDF to $p(\phi|x_1)$, which at the same time will be used a prior distribution when more data x_2 is obtained.

Both the prior and the gathered information about the value of ϕ are then codified in the posterior PDF $p(\phi \,|\, \boldsymbol{x})$. Therefore, from the Bayesian point of view, it suffices to give this a posteriori distribution because it contains all the information collected so far. However, sometimes one may be asked to give a single value for the parameter and thus to define an estimator $\hat{\phi}_\nu(\boldsymbol{x})$. The election of such function should be made by means of a cost function $C\left(\hat{\phi}(\boldsymbol{x}), \phi\right)$ that assigns a value to each estimate to represent its quality. The estimators can be qualified using as figure of merit the averaged cost $\mathcal{C}_\nu(\hat{\phi})$, i.e., the average of the cost function over the possible values of the parameter and the possible realisation $\boldsymbol{x}$:

$$\mathcal{C}_\nu(\hat{\phi}) = \int d\phi\, p(\phi) \int dx^\nu\, C\left(\hat{\phi}(\boldsymbol{x}), \phi\right)\, p(\boldsymbol{x}\,|\,\phi)\,. \tag{3.3}$$

The optimal estimator would be the one with the minimum $\mathcal{C}_\nu$. In some texts, a *utility function* is defined instead of a cost function, the only difference is that it must be maximised instead of minimised.

The most used figure of merit to quantify the accuracy of an estimate is the squared error with respect the true value of the parameter, i.e., $\sigma_{\hat{\phi}_\nu(\boldsymbol{x})} = (\hat{\phi}_\nu(\boldsymbol{x}) - \phi)^2$. In this case, the estimator will be the one that minimise the *Averaged Mean Square Error*

$$\overline{\mathrm{MSE}}(\hat{\phi}_\nu) = \int d\phi \int dx^\nu \left(\hat{\phi}_\nu(\boldsymbol{x}) - \phi\right)^2 p(\boldsymbol{x}, \phi)\,, \tag{3.4}$$

where the average runs over all possible values of ϕ and $\boldsymbol{x}$ weighted with the joint probability $p(\boldsymbol{x}, \phi) = p(\phi\,|\,\boldsymbol{x})p(\boldsymbol{x})$, thus

$$\overline{\mathrm{MSE}}(\hat{\phi}_\nu) = \int dx^\nu\, p(\boldsymbol{x}) \int d\phi \left(\hat{\phi}_\nu(\boldsymbol{x}) - \phi\right)^2 p(\phi|\boldsymbol{x})\,. \tag{3.5}$$

The minimal $\overline{\mathrm{MSE}}$ will be attained by the estimator that minimise the second integral for all $\boldsymbol{x}$. Therefore, by taking the derivative with respect to $\hat{\phi}(\boldsymbol{x})$ one can show that the optimal estimator is equal to the expected value of ϕ computed with respect to the conditional PDF $p(\phi\,|\,\boldsymbol{x})$, i.e.,

$$\hat{\phi}^{\mathrm{MMSE}}(\boldsymbol{x}) = \langle\phi\rangle_{p(\phi\,|\,\boldsymbol{x})} = \int d\phi\, p(\phi\,|\,\boldsymbol{x})\, \phi \tag{3.6}$$

which is called the *Minimum Mean Squarred Error* (MMSE) estimator.

In some cases the MSE is not a good choice as figure of merit. This is so because the integral in Eq. (3.3) might not respect the possible symmetries of the parameter. It is the case for example of angular parameters with circular symmetry $\phi = \phi + 2\pi k$ with $k \in \mathbb{Z}$, which is the case treated in this thesis. In order to correctly account for this symmetry, the cost function should satisfy:

- symmetric $C(\hat{\phi}, \phi) = C(\phi, \hat{\phi})$;

- periodic $C(\hat{\phi}, \phi) = C(\hat{\phi} + 2\pi k, \phi)$ with $k \in \mathbb{Z}$;
- group invariant $C(\hat{\phi}, \phi) = C(\hat{\phi} + \theta, \phi + \theta)$;
- and it has to reach its minimum value at $\hat{\phi} = \phi$ and raise monotonically in $|\phi - \hat{\phi}|$ to its maximum value at $\hat{\phi} = \phi \pm \pi$.

A general function that fulfils the above conditions can be written as a series of cosines of the difference angle $\theta = \hat{\phi} - \phi$, i.e. $C(\theta) = \sum_{k=0}^{\infty} \beta_k \cos(k\,\theta)$ with $\beta_k \geq 0$. In terms of a utility function instead of a cost function, one has only to invert the positions of the extremal points in the last of the conditions. In what follows I use the utility function $f(\theta) = \frac{1}{2}\,[1 + \cos(\theta)]$ which takes values in the range $[0, 1]$.

The main criticism about Bayesian estimation methods is related to the prior distribution about the parameter. It is true that some choices of prior distributions can have a big impact in the estimated value, especially with small samples. But this is not wrong by itself if this prior distribution represents reliably a true information about the parameter. In fact, one of the most used prior distribution is a uniform one, i.e., the distribution that express no preferences about any value.

As an example [Kay, 1993] consider a sample $\{x_1, \ldots, x_\nu\}$ with each x_i distributed following a normal distribution $p(x_i|\mu) = \mathcal{N}(\mu, \sigma)$ from which we want to estimate its mean μ. Our previous knowledge is that its value is around some μ_0, and for this reason we assign it a prior normal distribution $p(\mu) = \mathcal{N}(\mu_0, \sigma_\mu)$. The variance σ_μ represents the uncertainty we have about the true value of μ, i.e., the larger σ_μ the less informative is the prior PDF. The posterior PDF will be proportional to the probability of the sample conditioned by the value of μ times the prior assigned to this parameter, i.e.

$$p(\mu|\boldsymbol{x}) \propto p(\boldsymbol{x}|\mu)\,p(\mu) \propto \exp\left[-\frac{(\mu - \mu_0)^2}{2\sigma_\mu^2} - \sum_{k=1}^{\nu} \frac{(x_k - \mu)^2}{2\sigma^2}\right],, \tag{3.7}$$

up to a normalisation factor that does not depend on μ. By applying some algebra, it is straightforward to write the posterior PDF also as a normal distribution with mean

$$\mu_{|\boldsymbol{x}} = \alpha\,\langle x\rangle + (1-\alpha)\,\mu_0 \quad \text{with} \quad \alpha = \frac{\sigma_\mu}{\sigma_\mu + \sigma/\nu} \quad \text{and} \quad \langle x\rangle = \frac{1}{\nu}\sum_{k=1}^{\nu} x_k \tag{3.8}$$

and variance

$$\sigma^2_{\mu|\boldsymbol{x}} = \left(\frac{1}{\sigma_\mu^2} + \frac{\nu}{\sigma^2}\right)^{-1} = \frac{\sigma^2\,\sigma_\mu^2}{\sigma^2 + \nu\,\sigma_\mu^2}. \tag{3.9}$$

Notice that the mean of this posterior distribution, which coincides with the MMSE estimator, is a convex combination (since $0 \leq \alpha \leq 1$) of the a priori mean μ_0 and the sample mean $\langle x \rangle$, with weights α that depend on the sample size ν. For small samples ($\sigma^2/\nu \gg \sigma_\mu^2$) the weight $\alpha \approx 0$, hence the posterior PDF is almost equal to the prior distribution, i.e. $\mu_{|\boldsymbol{x}} \approx \mu_0$ and $\sigma^2_{\mu|\boldsymbol{x}} \approx \sigma^2_\mu$. In this case the form of the prior information is very relevant in the estimation. On the contrary, for large samples ($\sigma^2/\nu \ll \sigma^2_\mu$), the estimated value tend to the mean of the sample $\langle x \rangle$, which will coincide with μ as ensured by the law of large numbers. The information codified in the prior PDF plays less and less important role. Summarising, notice that the coefficient α indicates the relative confidence we have between the initial knowledge (σ_μ) and that acquired experimentally (σ/ν).

As a final remark, note that any prior information increase the accuracy of the estimation. This can be seen by computing explicitly Eq. (3.5) for this concrete example. Since the MMSE estimator is the mean of the posterior distribution [see Eq. (3.6)], the second integral in Eq. (3.5) coincides by definition with the posterior variance and the integral with respect to x^ν can be done straightforwardly because the variance in Eq. (3.9) does not depend on $\boldsymbol{x}$. The $\overline{\text{MSE}}$ is equal in this case to the variance in Eq. (3.9), which can be written as

$$\overline{\text{MSE}}(\hat{\phi}_\nu^{MMSE}) = \frac{\sigma^2}{\nu}\left(\frac{\sigma^2_\mu}{\sigma^2_\mu + \sigma^2/\nu}\right) \leq \frac{\sigma^2}{\nu}, \tag{3.10}$$

where the equality hold only for the complete uninformative prior with $\sigma_\mu \to \infty$. Therefore, any prior information increases the accuracy of the MMSE estimator.

3.2.2 Frequentist estimation

In contrast to the Bayesian point of view, the Frequentist approach assumes that the parameter is deterministic, i.e. it has a fixed value ϕ. However the estimator must be treated as a random variable since it is a function of the random realisations $\boldsymbol{x}$. Hence, the statistical properties of the estimator, such as mean or variance, can still be defined, i.e.

$$\langle \hat{\phi}_\nu \rangle_\phi = \int \hat{\phi}_\nu(\boldsymbol{x})\, p(\boldsymbol{x};\phi)\, d^\nu x\,, \tag{3.11}$$

$$\text{Var}(\hat{\phi}_\nu)_\phi = \int \left(\hat{\phi}_\nu(\boldsymbol{x}) - \langle \hat{\phi} \rangle_\phi\right)^2 p(\boldsymbol{x};\phi)\, d^\nu x\,, \tag{3.12}$$

where the semicolon in the argument of the PDF indicates that the quantities at the right are parameters and not variables. The first requirement that this approach demands to the estimator is to be *consistent*, i.e., the probability distribution of the estimated value has to become infinitely peaked around the true value ϕ as the sample size ν grows. To put it mathematically,

$$\lim_{\nu\to\infty} \langle\hat{\phi}_\nu\rangle_\phi = \phi \tag{3.13}$$

$$\lim_{\nu\to\infty} \mathrm{Var}(\hat{\phi}_\nu)_\phi = 0. \tag{3.14}$$

This requirement emphasises two important aspects of the Frequentist approach: the fixed-value character of the parameter and the definition of probability as frequencies of infinitely sized samples.

The performance of an estimator is evaluated by means of the MSE

$$\mathrm{MSE}(\hat{\phi}_\nu)\Big|_\phi = \int d^\nu x \,\left(\hat{\phi}_\nu(\boldsymbol{x}) - \phi\right)^2 p(\boldsymbol{x};\phi) \tag{3.15}$$

which is to be minimised to find the optimal estimator. In general, the optimal estimator could be different depending on the value of ϕ. Such *local* optimal estimators require the estimated parameter to be known before the estimation procedure itself. This fact seems to imply that they are useless for practical purposes. However, since any other estimator will perform worse than them, they can be used to derive a lower bound on the estimation accuracy. Moreover, in some cases, it can be found a *global* optimal estimator, that is an estimator that performs optimally for any value of ϕ.

One of the most important results of the Frequentist estimation theory is the Cramér-Rao lower bound (CRLB). This bound refers to *unbiased estimators*, i.e. estimators that yield on average to the true value of the parameter,

$$\langle\hat{\phi}\rangle_\phi = \phi \qquad \forall\phi. \tag{3.16}$$

For these, the MSE in Eq.(3.15) and the variance in Eq.(3.9) are equivalent. In fact, defining the bias of an estimator as $b(\hat{\phi}) = \langle\hat{\phi}\rangle_\phi - \phi$, the variance and the MSE are related by

$$\mathrm{MSE}(\hat{\phi}) = \mathrm{Var}(\hat{\phi}) + b^2(\hat{\phi}). \tag{3.17}$$

Then, for unbiased estimators minimising its variance and its MSE are equivalent. In general, the Frequentist approach only deals with unbiased estimators. However, it may exists a biased estimator with better accuracy than unbiased ones, or even be the only one that minimise the MSE. But, as an estimator is required to be consistent, in the assymptotic limit of large sample size, which is the regime that the Frequentist approach is designed for, the estimator must be also unbiased.

Going back to the CRLB, it can be easily derived from the unbiased condition in Eq. (3.16)[Kay, 1993]. Some regularity conditions are also asked to the PDF:

- for all values of x with non-zero probability, $\frac{\partial \log p(\boldsymbol{x};\phi)}{\partial \phi}$ must exist and be finite;
- the operations of integration with respect to x and differentiation with respect to ϕ must be interchangeable.

These conditions are in general fulfiled, except for those PDF that have a support (the set of values of x with non-zero probability) whose limits depend on the parameter ϕ. With these conditions at hand, the Eq. (3.16) can be differentiated with respect to ϕ to give

$$\begin{aligned} 1 &= \frac{\partial}{\partial \phi}\langle\hat{\phi}\rangle = \int \hat{\phi}(\boldsymbol{x})\, \frac{\partial p(\boldsymbol{x};\phi)}{\partial \phi}\, d^\nu x = \\ &= \int \hat{\phi}(\boldsymbol{x})\, \frac{\partial \log p(\boldsymbol{x};\phi)}{\partial \phi}\, p(\boldsymbol{x};\phi)\, d^\nu x \end{aligned} \tag{3.18}$$

where log refers to the natural logarithm. In the same way, by using the normalisation of the PDF one can see that

$$\int \phi\, \frac{\partial \log p(\boldsymbol{x};\phi)}{\partial \phi}\, p(\boldsymbol{x};\phi)\, d^\nu x = 0 \tag{3.19}$$

Then, by substracting (3.18) and (3.19), one gets

$$\int \left(\hat{\phi}(\boldsymbol{x}) - \phi\right)\, \frac{\partial \log p(\boldsymbol{x};\phi)}{\partial \phi}\, p(\boldsymbol{x};\phi)\, d^\nu x = 1 \tag{3.20}$$

and using the Cauchy-Schwarz inequality

$$\left(\int w(x)\, g(x)\, h(x)\, dx\right)^2 \leq \int w(x)\, g^2(x)\, dx \int w(x)\, h^2(x)\, dx \tag{3.21}$$

with $w(x) \geq 0$. The equality holds if and only if $g(x) = C\, h(x)$, with C an arbitrary constant independent of x. Doing the associations $w(x) = p(\boldsymbol{x};\phi)$, $g(x) = \left(\hat{\phi}(\boldsymbol{x}) - \phi\right)$ and $h(x) = \frac{\partial \log p(\boldsymbol{x};\phi)}{\partial \phi}$ in Eq. (3.20) one gets

$$\int \left(\hat{\phi}(\boldsymbol{x}) - \phi\right)^2 p(\boldsymbol{x};\phi)\, d^\nu x \int \left(\frac{\partial \log p(\boldsymbol{x};\phi)}{\partial \phi}\right)^2 p(\boldsymbol{x};\phi)\, d^\nu x \geq 1 \tag{3.22}$$

$$\Rightarrow \mathrm{Var}(\hat{\phi})\Big|_\phi \geq \left[\int \left(\frac{\partial \log p(\boldsymbol{x};\phi)}{\partial \phi}\right)^2 p(\boldsymbol{x};\phi)\, d^\nu x\right]^{-1} \tag{3.23}$$

which is the so-called Cramer-Rao lower bound. The term in square brackets can be easily shown to be equal to the Fisher information

$$I_\nu(\phi) = -\left\langle \frac{\partial^2 \log p(\boldsymbol{x};\phi)}{\partial \phi^2} \right\rangle \tag{3.24}$$

where the subscript ν reminds that the Fisher information is computed over the distribution $p(\boldsymbol{x};\phi)$ of a sample of size ν. Thus, the Cramer-Rao lower bound can be written compactly as

$$\mathrm{Var}(\hat{\phi})\Big|_{\phi} \geq I_\nu(\phi)^{-1} \tag{3.25}$$

Hence, the greater the Fisher information the better the parameter can be estimated, and thus $I_\nu(\phi)$ can be understood as the sensitivity of the PDF on the parameter or the distinguishability of the parameter. Mathematically, the Fisher information is additive for independent realisations, i.e., if $\boldsymbol{x}$ consists in independent and identically distributed (iid) realisations, then $I_\nu(\phi) = \nu\, I(\phi)$, with $I(\phi)$ the Fisher information of a single realisation. Therefore, the CRLB tells us that for iid samples the optimal variance can scale with the size ν at least as $1/\nu$, i.e.

$$\mathrm{Var}(\hat{\phi})\Big|_{\phi} \geq \frac{1}{\nu\, I(\phi)}\,. \tag{3.26}$$

An estimator that saturates the CRLB is said to be an *efficient* estimator, but one has to keep in mind that there not always exists such estimator. However, from the saturability of the Cauchy-Schwarz inequality, a condition for the existence of an efficient estimator can be formulated

$$\frac{\partial \log p(\boldsymbol{x};\phi)}{\partial\phi} = \frac{1}{C(\phi)}\left(\hat{\phi}(\boldsymbol{x}) - \phi\right) \tag{3.27}$$

where the possible dependence of the constant C on ϕ and the dependence of the estimator on the sample $\boldsymbol{x}$ has been written explicitly. By taken the derivative with respect to the parameter ϕ and averaging with respect to $\boldsymbol{x}$, $C(\phi)$ reads

$$C(\phi) = -\left\langle \frac{\partial^2 \log p(\boldsymbol{x};\phi)}{\partial\phi^2}\right\rangle^{-1} = I(\phi)^{-1}, \tag{3.28}$$

where the unbiased and regularity conditions has been used. With this, it can be stated that if the PDF fulfils the condition

$$\frac{\partial \log p(\boldsymbol{x};\phi)}{\partial\phi} = I(\phi)\ \left(g(\boldsymbol{x}) - \phi\right) \tag{3.29}$$

for some function $g(\boldsymbol{x})$ with only dependence on the sample $\boldsymbol{x}$, then there exists an unbiased estimator that saturates the CRLB and it is equal to $g(\boldsymbol{x})$.

3.3 Quantum state discrimination

Before entering in the main topic of this thesis, the quantum parameter estimation, I discuss the quantum state discrimination (QSD) problem. I use it to introduce quantum optimisation problems and to show how the choice of different figures of merit lead to very different solution for the same initial situation.

As stated in the previous chapter, two non-orthogonal quantum states can not be distinguished perfectly. However, one can ask for the measurement that optimally discriminate between them. But what does "optimally" means in this context? Let me look at the following hypothetical situation: consider a chamber where is stored a a treasure. Inside the chamber there is also a big deadly explosive with a countdown timer. An adventurer is inside the chamber. He look at the bomb and discover two wires: one is yellow with red stripes and the other is blue with also red stripes. He is sure that cutting the one that connects directly with the explosive he can deactivate the bomb, and luckily he is prudent and always carries a wire cutter. He has to find out which is the cable to cut, so he keep searching on the bomb. He find a very little gap through which he can see the explosive and just a little piece of the wire. In fact, he sees no stripes in the wire and only one color. If that color is blue or yellow, the adventurer has certainty about which wire to cut, but if it is red it gives no information. Let me now consider two different situations: the door is closed or it is opened. Clearly, the adventurer has two goals: save his life and get the treasure, and this yield to different strategies depending on the door. If the door is closed he is forced to cut one of the wires, so if he see the wire piece red the only remaining option is to choose randomly one of the wires and cut it in the last second. But if the door is opened, another option arise: to choose no wire and run away. In this sense we have added to the decision a third outcome. In this case the measurement performed by the adventurer is classical and hence it can not be tuned in any way. Nevertheless, quantum mechanics allows for a bigger variety of measurements, and a change in the figure of merit (like the one presented above) leads two different optimal measurement, as we are going to see for quantum state discrimination.

Formally, the quantum state discrimination problems are defined like this: we are given a quantum system with the only information that it is in one state chosen from a set $\{\rho_k\}_{k=1}^n$ with some prior probability $\eta_k \geq 0$, with $\sum_k \eta_k = 1$. Our aim is to design a measurement to figure out in which of these possible states, or hypotheses, the system is. As in the fictional example above, different figures of merit yield different optimal measurement.

This is the case of two paradigmatic scenarios for QSD: the Minimum-Error (ME) and the Unambiguous (UA) cases. In the former, the measurement has to necessarily deliver a guess. In this scenario the probability of getting the correct answer has to be maximised, or what is the same, to minimise the probability of obtaining an erroneous outcome. On the contrary, in the UA discrimination scenario it is not allowed to give an erroneous answer, but the measurement can yield to an inconclusive outcome, i.e. an outcome with any hypothesis assigned and meaning "I don't know". Since ME and UA are extreme situations (fixing the inconclusive outcome probability or the erroneous outcome probability to its extremal minimum value 0, respectively), the quantum state discrimination allows for a generalisation, where some weaker bounding constraints are imposed to the probabilities.

Now I am to describe firstly the two extreme cases of ME and UA quantum state discrimination. Then I introduce the intermediate situation in which some error-margins are imposed and present an unpublished explicit computation for the case of a set of symmetric states.

3.3.1 Minimum-error quantum state discrimination

In the minimum-error (ME) scenario the measurement assigns to each outcome one of the possible hypothesis. For this reason, the number of possible outcomes is the same as the number of hypotheses. This does not mean that there is not any hypothesis with a null probability of obtaining it. The measurement will be represented by a POVM $\mathcal{P}$ with n elements, i.e. a set of n positive semi-denifite operators $\{\Pi_k\}$, each one associated with an outcome ω_k. The guessing rule is the trivial one that maps each outcome ω_k to the hypothesis ρ_k with the same label. From the Born rule, the probability of obtaining the outcome ω_j when the system is initially in the state ρ_k is $p(\omega_j|\rho_k) = \mathrm{tr}\,(\Pi_j\,\rho_k)$. With this, the *success* probability p_{succ} of getting a correct answer and the *error* probability p_{err} of getting an erroneous one can be written as

$$p_{succ} = \sum_{k=1}^{n} \eta_k\, p(\omega_k|\rho_k) = \sum_{k=1}^{n} \eta_k\, \mathrm{tr}\,(\Pi_k\,\rho_k) \tag{3.30}$$

$$p_{err} = \sum_{k=1}^{n}\sum_{j\neq k} \eta_j\, p(\omega_k|\rho_j) = \sum_{k=1}^{n}\sum_{j\neq k} \eta_j\, \mathrm{tr}\,(\Pi_k\,\rho_j) \tag{3.31}$$

Obviously $p_{succ} + p_{err} = 1$, what can be easily check using the normalisation constraint for the density matrices, the resolution of the identity of the POVM and the fact that $\sum_k \eta_k = 1$. Thus, the ME discrimination problems can be expressed as the following semidefinite programming:

$$\begin{aligned}
\text{minimise} \quad & \textstyle\sum_{k=1}^{n} \sum_{j\neq k} \eta_j \, \mathrm{tr}\, (\Pi_k \, \rho_j) \\
\text{subject to} \quad & \Pi_k \geq 0 \quad \forall k \\
& \textstyle\sum_k \Pi_k = \mathbb{1}
\end{aligned} \tag{3.32}$$

In general, this is not an easy problem to solve analytically, and it is very useful to use any existing symmetry in the hypothesis ensemble $\{\rho_k, \eta_k\}$ to simplify the problem.

As an example consider a simple case with only two hypothesis $\{\rho_1, \rho_2\}$ with prior probabilities $\{\eta_1, \eta_2\}$, which can be solved easily. In this simple case the probability of error can be written as

$$\begin{aligned}
p_{err} &= \eta_1 \mathrm{tr}\, [\Pi_2 \rho_1] + \eta_2 \mathrm{tr}\, [\Pi_1 \rho_2] = \\
&= \eta_1 \mathrm{tr}\, [\rho_1] + \mathrm{tr}\, [\Pi_1 (\eta_2 \rho_2 - \eta_1 \rho_1)] = \\
&\equiv \eta_1 + \mathrm{tr}\, [\Pi_1 \Gamma]
\end{aligned} \tag{3.33}$$

where the identity resolution $\Pi_1 + \Pi_2 = \mathbb{1}$ has been used. The operator $\Gamma = \eta_2 \rho_2 - \eta_1 \rho_1$ is the so-called *Helstrom matrix* [Helstrom, 1976]. In order to find an explicit form for the POVM elements, notice first that Γ can have positive, negative and null eigenvalues. Let $\Gamma = \sum_k \gamma_k \, |\psi_k\rangle\langle\psi_k|$ be the spectral decomposition of the Helstrom matrix. It can be split as a difference between two positive matrices in the following way

$$\Gamma = \sum_{k:\, \gamma_k > 0} |\gamma_k| \, |\psi_k\rangle\langle\psi_k| - \sum_{k:\, \gamma_k < 0} |\gamma_k| \, |\psi_k\rangle\langle\psi_k| \equiv \Gamma_+ - \Gamma_- \tag{3.34}$$

Plugging this into Eq. (3.33) yields

$$p_{err} = \eta_1 + \mathrm{tr}\, [\Pi_1 \Gamma_+] - \mathrm{tr}\, [\Pi_1 \Gamma_-] \tag{3.35}$$

Since Π_1 as well as $\Gamma_\pm$ are positive semi-definite, the two last terms fulfil $\mathrm{tr}\, [\Pi_1 \Gamma_\pm] \geq 0$. It follows that the optimal POVM $\mathcal{P}^* = \{\Pi_1^*, \Pi_2^*\}$, that is the one that minimise p_{err}, must be the one that maximise $\mathrm{tr}\, [\Pi_1 \Gamma_-]$ and for which $\mathrm{tr}\, [\Pi_1^* \Gamma_+] = 0$ holds. It follows that the POVM element Π_1 has to be the projector over the subspace spanned by the eigenvectors with negative eigenvalues, i.e.

$$\Pi_1^* = \sum_{k:\, \gamma_k < 0} |\psi_k\rangle\langle\psi_k| \tag{3.36}$$

In an equivalent way one can see that the other element Π_2 has to be the projector over the positive eigenspace. In order to fulfil the identity resolution, it remains to add the projector over the null eigenspace, but it has no effect in the probability of error. Hence it can be distributed indifferently between the two POVM elements. With this, the probability of error can be written in terms of the eigenvalues as

$$p_{err} = \eta_1 - \sum_{k:\gamma_k<0} |\gamma_k| = \frac{1}{2}\left(1 - \sum_k |\gamma_k|\right) \tag{3.37}$$

where it have been used that $\mathrm{tr}\,\Gamma = \eta_2 - \eta_1 = \sum_{k|\gamma_k\geq 0}\gamma_k + \sum_{k|\gamma_k<0}\gamma_k$ and $\eta_1 + \eta_2 = 1$. A compact way of writing this error probability is

$$p_{err} = \frac{1}{2}\left(1 - ||\eta_2\rho_2 - \eta_1\rho_1||_1\right) \tag{3.38}$$

where we have used the *trace norm* operation defined as $||\Gamma||_1 = \mathrm{tr}\,|\Gamma| = \sum_k |\gamma_k|$.

For the case of pure-state hypothesis, that is $\rho_k = |\psi_k\rangle\langle\psi_k|$, the probability of error can be written as

$$p_{err} = \frac{1}{2}\left(1 - \sqrt{1 - 4\eta_1\eta_2|\langle\psi_1|\psi_2\rangle|^2}\right) \tag{3.39}$$

Notice that if the state are orthogonal, i.e. $\langle\psi_1|\psi_2\rangle = 0$, the probability of error vanish, meaning that the state are completely distinguishable. In the other extreme, if the two hypothesis are the same, i.e. $\langle\psi_1|\psi_2\rangle = 1$, the optimal POVM is the one that always outcome the most probable hypothesis, and hence the probability of error is equal to the prior probability of the less likely hypothesis.

As a second example, let me consider the case of N symmetric pure states with equal prior probabilities $\eta_k = 1/N$. In this case, the set of hypothesis $\{\rho_k\}_{k=1}^N$ is described as

$$\rho_k = |\psi_k\rangle\langle\psi_k| = U^k\ |\Psi\rangle\langle\Psi|\ (U^\dagger)^k \tag{3.40}$$

where $|\Psi\rangle$ is a fiducial state and U is a unitary operation for which $U^N = \mathbb{1}$ holds. In general, a multi-hypotheses problem like that is hard to solve. However, in this case the strong symmetry of the hypothesis allows to find an analytical solution for the optimal POVM $\{E_k\}$ [Ban *et al.*, 1997]:

$$E_k = \Omega^{-1/2}\ |\psi_k\rangle\langle\psi_k|\ \Omega^{-1/2} \tag{3.41}$$

$$\Omega = \sum_{k=1}^{n} \rho_k \tag{3.42}$$

which discriminate the states defined in Eq (3.40) with a probability of error given by

$$p_{err} = 1 - |\ \langle\Psi|\,\rho^{-1/2}\,|\Psi\rangle\ |^2 \tag{3.43}$$

It is worth noting that, since the hermitian operator ρ commutes by construction with the unitary U, the POVM elements have the same symmetry property than the hypothesis states, i.e.

$$E_k = U^k \left(\rho^{-1/2} \left|\Psi\right\rangle\!\left\langle\Psi\right| \rho^{-1/2}\right) (U^\dagger)^k \tag{3.44}$$

This POVM is the so-called *square-root measurement*. It is a paradigmatic example of how the symmetries of the problem can allow to find a solution. Moreover, it is a general feature of problems that any symmetry of the hypothesis translates to a similar symmetry on the POVM elements. For instance, this will be the case in Bayesian estimation protocols (see Section 3.4.1) in which the circular symmetry of the parameter allows to define the covariant POVMs, which is a relevant result of this approach.

3.3.2 Unambiguous state discrimination

In the minimum-error scenario of state discrimination there is no absolutely certain outcome. This is so because whatever state is the system prepared in, there is always some probability, although small, for any of the possible measurement outcomes. As in the example of the introduction of the section, it may be some situation in which the only way to gain certainty in our outcomes is to allow the possibility of sometimes abstain of giving an answer. In 1987 Ivanovic [1987] presented an scheme in which this is the case. He propose a sequence of measurements over different copies of the state in a way that some sequence of outcomes is only possible for one of the hypothesis, but at the expense that some other possible outcome sequences that give no conclusive information about the system state. Following this scheme one ends up with a protocol that unambiguously discriminate between them. A year later Dieks [1988] unified the sequential protocol in a single POVM, that was proved to be optimal in Peres [1988]. This results was then generalised in Jaeger and Shimony [1995] for arbitrary prior probabilities of the hypothesis.

As stated in the last paragraph, the aim of unambiguous state discrimination (USD) is to find a measurement with a null probability of wrong answers, i.e. $p(\omega_j|\rho_k) \propto \delta_{j,k}$. Let Π_k (with $k = 1, 2, \ldots, n$) be a POVM element associated with the hypothesis state ρ_k with the same label. Applying the Born rule, the above condition translates to

$$\mathrm{tr}\left(\Pi_k \, \rho_j\right) = 0 \quad \forall k \neq j \tag{3.45}$$

This expression constraints the measurement, but also the hypothesis ensemble has to satisfy some conditions. Let me analyse the implications of USD condition in Eq (3.45). The POVM must fulfil that its element associated with an hypothesis must be orthogonal to the support of the density matrices of all the other hypotheses. This immediately discard the possibility of full-rank hypotheses for which no such operators exists. Furthermore, for pure states, this turn on to the condition that the hypotheses should form a linearly independent set, and hence, in particular, the number of hypotheses can not be greater than its dimension.

Let me focus now in the case of two qubits $\{\rho_1, \rho_2\}$ with prior probabilities $\{\eta_1, \eta_2 = 1 - \eta_1\}$. In this case, if the states are mixed, the POVM elements have to be the null operator in order to fulfil the condition. Therefore, for mixed qubits states there is no POVM that unambiguously discriminates between them. For pure states $\rho_k = |\psi_k\rangle\langle\psi_k|$ it can be explicitly solved. The condition in Eq. (3.45) implies that the POVM elements have to take the form

$$\begin{aligned} \Pi_1 &= \mu_1 \left|\psi_2^\perp\right\rangle\!\left\langle\psi_2^\perp\right| \\ \Pi_2 &= \mu_2 \left|\psi_1^\perp\right\rangle\!\left\langle\psi_1^\perp\right| \end{aligned} \tag{3.46}$$

where $\left|\psi_k^\perp\right\rangle$ is the vector orthogonal to $|\psi_k\rangle$ and μ_1 and μ_2 are constants. Therefore, if Π_1 clicks we tell that the system is in the state $|\psi_1\rangle$ for exclusion. This two detection operators do not sum up to the identity, except for the orthogonal case $\xi = |\langle\psi_1|\psi_2\rangle| = 0$. For his reason, one must allow a third POVM element $\Pi_0 = \mathbb{1} - \Pi_1 - \Pi_2$, which corresponds to an inconclusive result that gives no information about the state to be discriminated. In other words, there will be an abstention probability

$$\begin{aligned} Q &= \eta_1 \operatorname{tr}(\Pi_0\, \rho_1) + \eta_2 \operatorname{tr}(\Pi_0\, \rho_2) = \\ &= 1 - (1-\xi^2)(\eta_1\, \mu_1 + \eta_2\, \mu_2) \end{aligned} \tag{3.47}$$

that the experimentalist fails to give a conclusive answer. From this point, it only remains to find the constants μ_1 and μ_2 that minimize such abstention probability Q while keeping the POVM elements positive. The positivity condition for the detection operators can be written as $\mu_k \geq 0$, and for the abstention operator

$$1 - \mu_1 - \mu_2 \quad + (1-\xi^2) \quad \mu_1\, \mu_2 \geq 0 \tag{3.48}$$

These conditions together with Eq. (3.47) defines an optimisation problem and, since the objective function is linear, the solution has to be on the boundary; in fact, it has to lay in the boundary defined by Eq. (3.48). By putting it (as an equality) in Eq. (3.47) and minimising one gets

$$\mu_1 = \frac{1}{1-\xi^2}\left(1 - \sqrt{\frac{1-\eta_1}{\eta_1}}\,\xi\right) \tag{3.49}$$

A similar expression can be found for μ_2 by changing $\mu_1 \leftrightarrow \mu_2$. In order to keep them positive, not all values for the prior probability $\eta_1 = 1 - \eta_2 = \eta$ are possible. Only in the range

$$\frac{1}{1+\xi^2} \geq \eta \geq \frac{\xi^2}{1+\xi^2} \tag{3.50}$$

the solution in Eq. (3.49) is valid. Below this bound, the solution is $\mu_1 = 0$ and $\mu_2 = 1$, that means that the POVM effectively has only two elements. Therefore the measurement outputs only the hypothesis 2 with absolute certainty or an inconclusive answer. Above the bound the situation is the other way, i.e. the measurement only outputs the hypothesis 1 or abstains. Summarising, the inconclusive-output probability is

$$Q = \begin{cases} \eta + (1-\eta)\xi^2 & \text{if} \quad \eta \leq \frac{\xi^2}{1+\xi^2} \\ 2\sqrt{\eta(1-\eta)}\;\xi & \text{if} \quad \frac{\xi^2}{1+\xi^2} \leq \eta \leq \frac{1}{1+\xi^2} \\ 1-(1-\xi^2)\,\eta & \text{if} \quad \eta \geq \frac{1}{1+\xi^2} \end{cases} \tag{3.51}$$

and the probability of give a correct answer is $p_{succ} = 1 - Q$.

Further we have seen that the condition (Eq. (3.45)) limits the possible sets for which USD protocol can be performed. In 2008 Croke *et al.* [2008] proposed a relaxation of this condition: instead of imposing the condition of no erroneous outcomes, the protocol must maximise the *confidence* the experimentalist have in the obtained outcome. They define such confidence as the posterior probability of the true state being the hypothesis ρ_j conditioned to being obtaining the outcome with the same label j, i.e.

$$C_j = p(\rho_j|\omega_j) = \frac{\eta_j\, p(\omega_j|\rho_j)}{p(\omega_j)} \tag{3.52}$$

Notice that for the cases in which USD is possible, the maximum confidence for all the hypothesis will turn to be 1, meaning that the experimentalist is absolutely confident that the obtained outcome coincides with the true state of the system.

3.3.3 Quantum state discrimination with error margin

In state discrimination protocols there are three possible events (or measurement outcomes) that are to give a correct or erroneous answer or decline to giving any at all. This is summarised by the relation between the probabilities of such events $p_{succ} + p_{err} + Q = 1$. In both above presented protocols, minimum-error and unambiguous state discrimination, one of these probabilities is set to zero, $Q = 0$ and $p_{err} = 0$ respectively. In this sense, one might call these protocols to be extremal. However, notice that by allowing some inconclusive outcome in ME protocol the confidence in the conclusive ones will increase. Likewise, in the USD including the possibility of some erroneous outcomes leads to an enhancement in terms of success probability. Therefore, a general scheme can be defined by fixing either Q (discrimination with abstention) or p_{err} (discrimination with error margin) and maximising the probability of obtaining a correct outcome.

This general scheme was firstly introduced in **?**. In their work, the authors work out the case of discriminating two pure hypothesis. They find a lower bound for the probability of error given a fixed abstention rate Q. The author also analysed the results considering that the POVM acts with two steps: a first one that probabilistically transforms the input states into a more distinguishable pair or give an inconclusive outcome, and as a second step a projective measurement that corresponds to the minimum error protocol.

A different approach for analysing error-margin discrimination was introduced in the works Hayashi *et al.* [2008] and Sugimoto *et al.* [2009]. There the authors impose the probability of error to be less than an *error margin*, i.e. $p_{err} \leq m$, and then minimise the abstention probability. They named it the *weak error-margin condition*, in contrast to the *strong error-margin condition*, that they also introduce. The latter impose an upper bound to the probabilities of correct detection for each hypothesis, i.e. $p(\rho_j|\omega_{k\neq j}) \leq m$. However, in Sugimoto *et al.* [2009] they show that the two conditions are equivalent, that is, the strong condition leads to the same result (equal p_{succ} and equal POVM) than the weak one but with a tighter error margin.

There are different variations and approaches to the error-margins quantum state discrimination. For example, in Sentís *et al.* [2013a] the authors present a theoretical machine that discriminates between two *unknown* quantum states subject to a given error margin. Instead of having classical information a priori, the machine have access to program systems that are prepared in the possible hypothesis states that the target system could be. In this work the problem is tackled with the weak and the strong conditions, and they also find a relation between them.

Discrimination of symmetric quantum states with a fixed rate of inconclusive outcomes

In 2012, Bagan et al. introduced a method [Bagan *et al.*, 2012] to transform a general discrimination problem into the standard minimum-error scheme. Now I am to show as an example the case of N symmetric pure hypothesis following their discrimination with a fixed rate of inconclusive outcomes (FRIO) scheme.

Consider that an experimentalist received a system without knowing in which state it is. However, he has the promise that it is in one state within the set $\{\rho_n \in \mathcal{H}^d\}_{n=1}^N$ with equal prior probability $\eta_n = 1/N$. The hypothesis ρ_j are symmetric in the sense that fulfils the relation $\rho_n = U^{n-m}\,\rho_m\,U^{-(n-m)}$, where U is a unitary operation that can be written as

$$U = \sum_{k=1}^{d} \mathrm{e}^{i\,2\pi\,k/N}\,|u_k\rangle\!\langle u_k| \tag{3.53}$$

which fulfils that $U^N = \mathbb{1}$. The states $\{|u_k\rangle\}$ form the (orthonormal) eigenbasis of U. In this basis, the symmetric pure hypothesis can be written as

$$\rho_n = |\psi_n\rangle\!\langle\psi_n| \quad \text{with} \quad |\psi_n\rangle = \sum_{k=1}^{d} c_k\,\mathrm{e}^{i\,2\pi\,k\,n/N}\,|u_k\rangle \tag{3.54}$$

where by normalisation $\sum_{k=1}^d |c_k|^2 = 1$. Note that, without lost of generality, I can take $d \le N$. Our aim is to find a POVM $\mathcal{P} = \{\Pi_n\}_{n=0}^N$, where the elements Π_n with $n \ge 1$ identify the hypothesis with the same label and Π_0 represents the inconclusive outcome.

First of all, I show that the symmetry between the hypotheses allows me to take the POVM with a similar symmetry. Given a POVM $\widetilde{\mathcal{P}} = \{\widetilde{\Pi}_n\}_{n=0}^N$, one can construct another POVM $\mathcal{P} = \{\Pi_m\}_{m=0}^N$ such that

$$\Pi_m = U^m\,\Omega\,U^{-m} \quad \text{with} \quad \Omega = \frac{1}{N}\sum_{n=1}^{N} U^{-n}\widetilde{\Pi}_n U^n\,; \tag{3.55}$$

$$\Pi_0 = \frac{1}{N}\sum_{n=1}^{N} U^n\widetilde{\Pi}_0 U^{-n}\,. \tag{3.56}$$

It is easy to check that the set of operators defined in Eqs. (3.55) and (3.56) form indeed a POVM: the positivity of Π_m comes directly from the positivity of $\widetilde{\Pi}_m$ and the unitarity of U. Only remains to check that they sum up to the identity:

$$\begin{aligned}
\sum_{m=0}^{N}\Pi_m &= \Pi_0 + \sum_{m=1}^{N} U^m\,\Omega\,U^{-m} = \\
&= \Pi_0 + \frac{1}{N}\sum_{m,n=1}^{N} U^{m-n}\widetilde{\Pi}_n U^{-m+n} = \\
&= \frac{1}{N}\sum_{l=1}^{N} U^l\widetilde{\Pi}_0 U^{-l} + \frac{1}{N}\sum_{l,n=1}^{N} U^l\widetilde{\Pi}_n U^{-l} = \\
&= \frac{1}{N}\sum_{l=1}^{N} U^l\left(\sum_{n=0}^{N}\widetilde{\Pi}_n\right)U^{-l} = \mathbb{1}
\end{aligned} \tag{3.57}$$

where between the second and the third line I used that $U^N = \mathbb{1}$ and therefore $l = m - n$ take, modulo N, all the integer values between 1 and N. In the last equality I also used the fact that $\widetilde{\mathcal{P}}$ is a POVM ensures the resolution of the identity. Now consider the success probabilities p_{succ} and $\tilde{p}_{succ}$ obtained by the POVM's $\mathcal{P}$ and $\widetilde{\mathcal{P}}$ respectively.

$$\begin{aligned} p_{succ} &= \frac{1}{N}\sum_{m=1}^{N} \mathrm{tr}\,\left(\Pi_m\,\rho_m\right) = \\ &= \frac{1}{N}\sum_{n,m=1}^{N}\frac{1}{N}\mathrm{tr}\,\left(U^{m-n}\,\widetilde{\Pi}_n\,U^{-m+n}\,\rho_m\right) = \\ &= \frac{1}{N}\sum_{m=1}^{N}\frac{1}{N}\sum_{n=1}^{N}\mathrm{tr}\,\left(\widetilde{\Pi}_n\,\rho_n\right) = \\ &= \frac{1}{N}\sum_{m=1}^{N}\tilde{p}_{succ} = \tilde{p}_{succ} \end{aligned} \tag{3.58}$$

In the same way one can see that the two POVM's produce also the same abstention probability Q. Therefore, given a general POVM one can construct another one with the symmetries $\Pi_m = U^{m-n}\,\Pi_n\,U^{-m+n}$ for $n, m = 1, \ldots, N$ and $\Pi_0 = U\,\Pi_0\,U^\dagger$ that performs equally in discriminating the equiprobable hypothesis given by Eq. (3.54), i.e. the POVM can be choosen directly with this symmetry. Doing it, the set of variables reduces to the operators Ω and Π_0 and the probabilities are simplified to

$$p_{succ} = \mathrm{tr}\,(\Omega\,\rho_N) \tag{3.59}$$

$$Q = \mathrm{tr}\,(\Pi_0\,\rho_N) \tag{3.60}$$

With this at hand, one can see that the confidence of the outcomes, described in Eq. (3.52) of the previous section, takes a very simple expression

$$C_n = \frac{\eta_n\,\mathrm{tr}\,(\Pi_n\,\rho_n)}{\sum_{m=1}^{N}\eta_m\,\mathrm{tr}\,(\Pi_n\,\rho_m)} = \frac{p_{succ}}{p_{succ}+p_{err}} = \frac{p_{succ}}{1-Q} \tag{3.61}$$

Furthermore, since from the symmetry we have $[\Pi_0, U] = 0$, in the eigenbasis of U the abstention operator has a diagonal form $\Pi_0 = \mathrm{diag}(f_1, f_2, \ldots, f_d)$, where hermiticity and POVM positivity imply that $0 \leq f_k \leq 1$ has to hold for all $k = 1, \ldots, N$.

Now, let me consider the abstention rate to have a fixed value Q, so the problem reduces to find the maximum possible value for the success probability given it. From the resolution of the identity of the POVM one have

$$\sum_{n=1}^{N} U^n\,\Omega\,U^{-n} = \mathbb{1} - \Pi_0 \equiv \bar{\Pi}_0 \tag{3.62}$$

Multiplying from both sides by $\bar{\Pi}_0^{-\frac{1}{2}}$ yields

$$\sum_{n=1}^{N} U^n \,\bar{\Pi}_0^{-\frac{1}{2}} \Omega \,\bar{\Pi}_0^{-\frac{1}{2}} \,U^{-n} = \sum_{n=1}^{N} U^n \,\tilde{\Omega}\, U^{-n} = \mathbb{1} \tag{3.63}$$

where we have implicitly defined $\tilde{\Omega} = \bar{\Pi}_0^{-\frac{1}{2}} \Omega \,\bar{\Pi}_0^{-\frac{1}{2}}$. Notice that $\bar{\Pi}_0^{-\frac{1}{2}}$ only exist if all $f_k \neq 1$. We will see later that the optimal solution fulfils $f_k < 1$, except for the non-practical case $Q = 1$ in which the only POVM element that remain is $\Pi_0 = \mathbb{1}$. With this definition the success probability in Eq. (3.59) can be transformed in the following way

$$\begin{aligned} p_{succ} &= \operatorname{tr}\left(\bar{\Pi}_0^{\frac{1}{2}} \,\tilde{\Omega}\, \bar{\Pi}_0^{\frac{1}{2}} \,\rho_N\right) = \operatorname{tr}\left(\bar{\Pi}_0 \,\rho_N\right) \operatorname{tr}\left(\tilde{\Omega}\, \tilde{\rho}_N\right) = \\ &= (1-Q) \operatorname{tr}\left(\tilde{\Omega}\, \tilde{\rho}_N\right) \end{aligned} \tag{3.64}$$

where I have defined

$$\begin{aligned} \tilde{\rho}_N &= \frac{\bar{\Pi}_0^{\frac{1}{2}} \,\rho_N\, \bar{\Pi}_0^{\frac{1}{2}}}{\operatorname{tr}\left(\bar{\Pi}_0 \,\rho_N\right)} = \sum_{k,j=1}^{d} \sqrt{\frac{1-f_k}{1-Q}}\, c_k \sqrt{\frac{1-f_j}{1-Q}}\, c_j^* \,|u_k\rangle\langle u_j| = \\ &= \sum_{k,j=1}^{d} \tilde{c}_k \,\tilde{c}_j^* \,|u_k\rangle\langle u_j| \quad \text{with} \quad \tilde{c}_k \equiv \sqrt{\frac{1-f_k}{1-Q}}\, c_k \end{aligned} \tag{3.65}$$

By explicitly writing the abstention probability in terms of the state amplitudes c_k and the parameters f_k, i.e.

$$1 - Q = \operatorname{tr}\left(\bar{\Pi}_0 \,\rho_N\right) = \sum_{k=1}^{d} |c_k|^2 (1 - f_k), \tag{3.66}$$

it is easy to check that the transformed states are normalised. The success probability in Eq. (3.64), except for the factor $1 - Q$, defines a minimum-error discrimination problem with the hypothesis being the pure states in Eq. (3.65). It can be interpreted as if the system go first through a probabilistic channel (with a Kraus operator $\bar{\Pi}_0^{1/2}$) that transforms the input state $\rho_n \to \tilde{\rho}_n$ with some probability $1 - Q$ and then the measurement $\{U^n \,\tilde{\Omega}\, U^{-n}\}_{n=1}^{N}$ is performed. If the channel fails to transform the state the experimentalist does not perform the measurement and take as output an "I-don't-know" answer. Since, as we will see, the optimal $\widetilde{\Omega}$ does not depend in f_k, it can be optimised separately from Π_0. So, I will find first the optimal measurement to be applied over the transformed states $\tilde{\rho}_n$, and then I will find the channel that generate an ensemble of these states that maximise the performance of such measurement.

Let me write $\tilde{\Omega}$ in the eigenbasis of the unitary operator as

$$\tilde{\Omega} = \sum_{j,k=1}^{d} \omega_{jk} \, |u_j\rangle\langle u_k| \, . \tag{3.67}$$

Then, using the relation

$$\sum_{n=1}^{N} \mathrm{e}^{i\frac{2\pi n}{N}(j-k)} = N\delta_{jk}, \tag{3.68}$$

the resolution of the identity in Eq. (3.63) implies that $\omega_{jj} = 1/N$. Furthermore, the positivity of $\tilde{\Omega}$ yield the constraint $|\omega_{jk}| \leq \sqrt{\omega_{jj}\,\omega_{kk}} = 1/N$. It can be used to find an upper bound to the success probability in discriminating the transformed states, i.e.

$$\begin{aligned}
p_{succ} &= (1-Q) \sum_{j,k=1}^{d} \omega_{jk}\, \tilde{c}_j^* \, \tilde{c}_k \\
&= (1-Q) \left(\sum_{j=1}^{d} \frac{1}{N} |\tilde{c}_j|^2 + \sum_{j,k=1|j\neq k}^{d} \omega_{jk}\, \tilde{c}_j^* \, \tilde{c}_k \right) \\
&\leq (1-Q) \left(\sum_{j=1}^{d} \frac{1}{N} |\tilde{c}_k|^2 + \sum_{j,k=1|j\neq k}^{d} |\omega_{jk}|\, |\tilde{c}_j|\, |\tilde{c}_k| \right) \\
&\leq (1-Q) \left(\sum_{j=1}^{d} \frac{1}{N} |\tilde{c}_j|^2 + \sum_{j,k=1|j\neq k}^{d} \frac{1}{N} |\tilde{c}_j|\, |\tilde{c}_k| \right) \\
&= \frac{1-Q}{N} \left(\sum_{j=1}^{d} |\tilde{c}_j| \right)^2 = \frac{1}{N} \left(\sum_{j=1}^{d} \sqrt{1-f_j}\, |c_j| \right)^2
\end{aligned} \tag{3.69}$$

The bound is saturated for $|\omega_{jk}| = 1/N$ and when all the complex phases of the state amplitudes $\tilde{c}_k$ are compensated by the phases of the operator $\tilde{\Omega}$, i.e.

$$\omega_{jk} = \frac{1}{N} \frac{c_j\, c_k^*}{|c_j|\,|c_k|} \tag{3.70}$$

From now on, I will consider, without lost of generality, either the state amplitudes c_j and the coefficients of $\tilde{\Omega}$ to be real and positive. In other words, the second fraction in Eq. (3.70) will be equal to one, and therefore $\omega_{j,k} = 1/N$ for $j, k = 1, 2, \ldots, d$. Moreover, since the phases play no role in the success probability, I group together all the coefficients with the same modulus and reorder the indices in decreasing order, i.e.

$$c_1 > c_2 > \cdots > c_M \tag{3.71}$$

and let n_j be the multiplicity of the original coefficients with modulus c_j. Note that $\sum_{j=1}^{M} n_j = d$ and the state normalisation is now imposed by $\sum_{j=1}^{M} n_j\, c_j^2 = 1$. To simplify further the notation let me define $\xi_j = \sqrt{1-f_j}$. With all this, the remaining problem can be expressed as

$$\text{Maximise} \qquad p_{succ} = \frac{1}{N}\left(\sum_{j=1}^{d} n_j\, c_j\, \xi_j\right)^2 \tag{3.72}$$

$$\text{subject to} \qquad 1 - Q = \sum_{j=1}^{d} n_j\, c_j^2\, \xi_j^2 \tag{3.73}$$

$$0 \leq \xi_j \leq 1 \quad \forall j \tag{3.74}$$

The solution must fulfil the Karush-Kuhn-Tucker (KKT) conditions [Boyd and Vandenberghe, 2004]. In addition to the *primal feasibility* conditions in Eqs. (3.73) and (3.74), it has also to satisfy for all values $j = 1; \ldots, d$ the *stationarity* condition

$$\frac{\partial}{\partial \xi_j}\sum_{k=1}^{d} n_k\, \xi_k\, c_k - \frac{\lambda}{2}\frac{\partial}{\partial \xi_j}\left(\sum_{k=1}^{d} n_k\, \xi_k^2\, c_k^2 - 1 + Q\right) - \frac{\partial}{\partial \xi_j}\sum_{k=1}^{d} \mu_k\,(\xi_k - 1) = 0 \tag{3.75}$$

where λ and μ_k are the KKT multipliers. Moreover, the *dual feasibility* condition $\mu_j \geq 0$ and the *complementary slackness* condition $\mu_j\,(\xi_j - 1) = 0$ must also hold for all j.

Solving the stationary condition yields

$$\xi_j = \frac{c_j - \mu_j}{\lambda\, c_j^2} \tag{3.76}$$

To determine the values of the multipliers μ_j one should focus on the slackness condition. It implies that either ξ_j takes the boundary value 1 or its corresponding multiplier μ_j vanish. Taking into account the dual feasibility condition, one can ensure that if some ξ_k is in the boundary, all the other with an smaller corresponding c_j will also be in the boundary. Therefore, I can define a threshold m to express the solution for ξ_j as

$$\xi_j = \begin{cases} \frac{1}{\lambda\, c_j} & \text{if} \quad j \leq m \\ 1 & \text{if} \quad j > m \end{cases} \tag{3.77}$$

The threshold m will depend on the concrete value of the abstention rate Q. To see it let first determine the value of λ. Plugging Eq. (3.77) into Eq. (3.73) one gets

$$\begin{aligned} 1-Q &= \frac{1}{\lambda^2}\sum_{j=1}^{m} n_j + \sum_{j=m+1}^{M} n_j\, c_j^2 \\ \Rightarrow \lambda^2 &= \frac{d_m}{1-Q-\sum_{j=m+1}^{M} n_j\, c_j^2} \end{aligned} \tag{3.78}$$

with $d_m \equiv \sum_{j=1}^{m} n_j$. Plugging it into Eq. (3.77) we get the final solution for ξ_j, i.e.

$$\xi_j = \begin{cases} \frac{1}{c_j}\sqrt{\frac{1-Q-\sum_{k=m+1}^{M} n_k\, c_k^2}{d_m}} & \text{if} \quad j \le m \\ 1 & \text{if} \quad j > m \end{cases} \tag{3.79}$$

In order to find a relation between the threshold m and the rate of abstention Q recall that this solution has to fulfil the constraint in Eq. (3.74), i.e.

$$\frac{1}{c_j}\sqrt{\frac{1-Q-\sum_{k=m+1}^{M} n_k\, c_k^2}{d_m}} \le 1 \quad \forall j \le m \tag{3.80}$$

$$\Rightarrow Q \ge 1 - d_m\, c_m^2 - \sum_{k=m+1}^{M} n_k\, c_k^2 \equiv Q_m \tag{3.81}$$

where I have used the fact that the most restrictive inequality is the one involving the smallest coefficient c_m. Therefore, the threshold m is determined by the condition $Q_{m+1} > Q \ge Q_m$, with $Q_{M+1} = 1$. One should ask if the factor inside the square root in the solution (3.79) is positive in order to keep ξ_j real, i.e.

$$\begin{aligned} 1-Q-\sum_{k=m+1}^{d} n_k\, c_k^2 &> 1-Q_{m+1}-\sum_{k=m+1}^{d} n_k\, c_k^2 = && (3.82) \\ &= d_{m+1}\, c_{m+1}^2 - n_{m+1}\, c_{m+1}^2 = && (3.83) \\ &= d_m\, c_{m+1}^2 > 0 && (3.84) \end{aligned}$$

Let me remark that this expression is also valid for $m = M$ by taking $c_{M+1} = 0$. With this, an expression can be obtained for the success probability by plugging the solution (3.79) into Eq. (3.72), i.e.

$$p_{succ}(Q) = \frac{1}{N}\left[\sqrt{d_m\left(1-Q-\sum_{k=m+1}^{M} n_k\, c_k^2\right)} + \sum_{k=m+1}^{M} n_k\, c_k\right]^2 \tag{3.85}$$

for $Q_{m+1} \geq Q \geq Q_m$. Moreover, considering the interpretation of the protocol as a probabilistic channel followed by a measurement, we can analyse this result in terms of the transformed states $\tilde{\rho}_n$ defined in Eq. (3.65). Using the Eq. (3.79), one can see that the coefficients $\tilde{c}_k$ are transformed by the channel into

$$\tilde{c}_k = \frac{\xi_k}{1-Q} = \begin{cases} \sqrt{\frac{1}{d_m}}\sqrt{1-\frac{\sum_{j=m+1}^{M} n_j\, c_j^2}{1-Q}} & \text{if} \quad j \leq m \\ \frac{c_k}{1-Q} & \text{if} \quad j > m \end{cases} \tag{3.86}$$

Note that the largest coefficients ($j \leq m$) are modified to have the same value, independently of which they have initially. The other subset ($j > m$) is only modified to keep the state normalised. In the case $Q > Q_M = 1 - d\, c_M^2$, all the coefficients take the value $c_j = d^{-1/2}$, which correspond to the optimal states in a minimum-error discrimination with symmetric states, as it can be check with the Eq. (3.43). The probability of success in this case is $p_{succ} = (d/N)(1-Q)$. Note that if $d = N$, the corresponding error probability $p_{err} = 1 - p_{succ} - Q = 0$, as one can deduce from the fact that with this relation the states are linearly independent and hence unambiguous discrimination can be performed between them. Furthermore, in this range of Q the confidence in Eq. (3.61) is independent of Q, i.e.

$$C_n = \frac{p_{succ}}{1-Q} = \frac{d}{N} \tag{3.87}$$

Therefore, to increase the abstention rate beyond Q_M does not produces an enhancement in the confidence of the obtained outcomes.

One may wonder how to determine the abstention rate (or the error margin) that he must impose in a given quantum state discrimination problem. In the recent paper Combes and Ferrie [2015] the authors uses the Bayesian formalism of cost function to analyse in this way the case of two pure hypothesis. To do so, they define cost functions that assign different costs to success, erroneous and inconclusive outcomes, and they find the POVM that minimised it. They recover the results of the previous works on this topic but in terms of these costs. In particular, they find as extremal cases the minimum-error and unambiguous discrimination solutions presented above, but with the exception that in the case of equal cost for abstention and success the USD is not the only protocol that minimise this cost function, the always-abstain protocol provides the same averaged cost. This formalism can be applied in all state discrimination problems, the only need is a function $p_{err}(Q)$ for the minimum error probability given an abstention rate (or vice-versa) to plug in the averaged cost function and find the optimal triplet $\{p_{succ}, p_{err}, Q\}$ that minimise it.

3.4 Quantum estimation

The aim of quantum metrology is to estimate the value of a (or a set of) unknown parameter $\phi \in \Phi$ codified on N quantum systems *independently* [**?**].

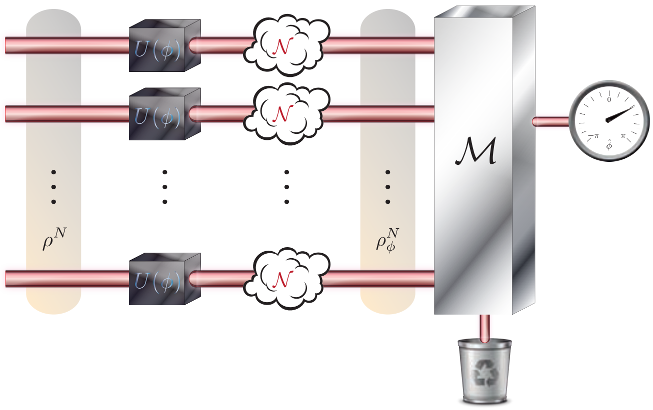


Figure 3.1. Pictorial representation of a probabilistic metrology protocol with n qubits. The probe state ρ^N, which needs not necessarily be a product of identical copies, undergoes an evolution $U_\phi^{\otimes N}$ controlled by the unknown parameter ϕ. Experimental noise $\mathcal{N}$ decoheres the system before a collective measurement on *all* qubits is performed. The measurement apparatus either returns an ultra-precise estimate $\hat{\phi}$ of the parameter or shows a failure signal. In the event of a failure, some information could be in principle scavenged (see last section in Results).

I call ρ^N the global input state and $\mathcal{H}^{\otimes N}$ the Hilbert space of the N systems. The parameter ϕ is imprinted in each system independently. I will restrict the discussion to those cases in which it is codified by a unitary transformation $U(\phi)$. During this process the systems might suffer the effect of some noise, represented by a quantum channel $\mathcal{N}$, that might act in a correlated or uncorrelated way among the N systems. The state in which the experimentalist has access is $\rho_\phi^N = \mathcal{N}\left[U(\phi)^{\otimes N}\,\rho^N\,(U(\phi)^\dagger)^{\otimes N}\right]$. Then he will apply a metrology protocol defined by a pair $(\mathcal{M}, \mathcal{G})$, where $\mathcal{M}$ is a measurement whose outcomes are associated to a value of the parameter by a guessing rule, or estimator, $\mathcal{G}$. The measurement is described by a POVM, i.e. a set of positive semi-defined operators $\{\Pi_x\}$ that fulfils $\sum_x \Pi_x = \mathbb{1}$, as presented in Chapter 2. The outcomes $x \in \mathrm{X}$ of this measurement can be treated as as a random variable that follows a probability distribution given by $p_{\mathcal{M}}(x\,|\,\phi) = \mathrm{tr}\,(\Pi_x\ \rho_\phi^N)$[1]. Here I write the sub-index $\mathcal{M}$ to remark explicitly the dependence on the chosen POVM of the probability distribution. The measurement can be repeated ν times (over different ensemble, since the information about the parameter is destroyed after each measurement) to get a sample $\boldsymbol{x} \in \mathrm{X}^\nu$. Then the estimator $\mathcal{G} : \mathrm{X}^\nu \to \hat{\Phi}$ is a function that associates a guessed value $\hat{\phi}_x \in \hat{\Phi}$ for the parameter to the gathered sample $\boldsymbol{x}$. The performance of the protocol is then evaluated by a figure of merit that compares the true value of the parameter with the estimated one.

An important remark should be done here: if the sample is required to be independently distributed it is the parameter ν which indicates its size, the number of systems N has nothing to do with it. This is so because the probability distribution $p_{\mathcal{M}}(x\,|\,\phi)$ is in general not factorisable on each subsystem, being this way only if the measurement is local in the subsystems and the system is in a product state $\rho_N = \bigotimes_{n=1}^N \rho_n$. But if it is so, the problem can be redefined in order to get an independently distributed sample of size ν and system size N. Hence, it what follows only the parameter ν will indicate the size of a independently distributed sample $\boldsymbol{x}$. The importance of this distinction becomes apparent in the scaling about these two parameter. Since the estimator $\mathcal{G}$ is purely classical, independent repetitions of the experiment leads to an improvement of the MSE of at most $1/\nu$. On the contrary, the scaling on the number of subsystem can beat the *Standard Quantum Limit* (SQL) of $1/N$ by using quantum tricks such as entanglement or state squeezing to get the so-called *Heisenberg Scaling* (HS) of $1/N^2$ [Giovannetti *et al.*, 2006]. The name of this quantum improvement on the precision comes from the Heisenberg uncertainty relation, which is the principle that dictates the best scaling achievable by quantum protocols.

In conclusion, the quantum parameter estimation splits into a quantum and a classical parts. In the former the optimisation consist in finding the measurement that leads to a probability distribution $p(\boldsymbol{x}|\phi)$ from which we can extract useful information about the parameter. Moreover, in some cases the input state can be not fixed and therefore one has to find also the one in which ϕ optimally extracted. Then, the classical part consists in finding the optimal estimator to extract the information from the PDF given in the quantum part. This second part is fully classical and the results presented in the section 3.2 apply.

---

[1] The POVM is assumed to be a collective measurement. With this assumption all possible scenarios are included, e.g. local measurement in which each subsystem is measured separately ($\Pi_x = \bigotimes_{n=1}^N \Pi_x^{(n)}$), or adaptive strategies in which the measurement on a subset of systems depends on the outcome of some other subset.

3.4.1 Bayesian approach

The classical subjective approach to parameter estimation can be generalised to its quantum version. In this case, the measurement must also be designed taking into account the prior knowledge about the parameter codified via a prior PDF $p(\phi)$. This is so because the classical Bayesian techniques presented in Section 3.2.1 will be applied over the posterior PDF $p(\phi|\boldsymbol{x}) \propto p_{\mathcal{M}}(\boldsymbol{x}|\phi)\, p(\phi)$. Nevertheless, the updating process that characterises the Bayesian protocols affect the successive measurements of the ν repetitions. Hence, in general the different sample points x_k will not be independent and the Eq. (3.2) will not apply. However, since the independent character of the sample is not required in this approach, one can consider a single-shot measurement that encompass all the repetitions by considering a global POVM that act over all the νN particles[2].

In contrast with the classical case presented in Section 3.2.1 where the performance of the protocol is evaluated with cost functions, I will use instead utility functions $f(\hat{\phi}, \phi)$. As stated in the this two formalisms are completely equivalent, with the only difference being the direction of optimisation of the figure of merit. Then, the quantity to be maximised is $f(\hat{\phi}, \phi)$ averaged over all the possible values of ϕ and $\hat{\phi}$, i.e.

$$F = \sum_x \int d\phi\, f(\hat{\phi}_x, \phi)\, \mathrm{tr}\, \left(\Pi_x\, \rho_\phi^N\right)\, p(\phi) \tag{3.88}$$

where the estimator is implicitly added in the guessed value $\hat{\phi}_x$ associated with each POVM outcome x.

One of the most used figure of merit is the *quantum state fidelity*, i.e. a measure of closeness of two quantum states ρ_1 and ρ_2 defined as

$$f(\rho_1, \rho_2) = \left[\mathrm{tr}\, \left(\sqrt{\sqrt{\rho_1}\, \rho_2 \sqrt{\rho_1}}\right)\right]^2 \tag{3.89}$$

which can take values in the interval $[0, 1]$. It fulfils the conditions for the case of circular parameter introduced in Section 3.2.1: symmetric, group invariant, periodic and it has the required monotonicity. For the case of classical distributions, i.e. $[\rho_1, \rho_2] = 0$ it reduces to the classical fidelity $f(\rho_1, \rho_2) = \sum_k \sqrt{p_k\, q_k}$, where p_k and q_k are the eigenvalues of ρ_1 and ρ_2 respectively. Moreover, for pure states $\rho_k = |\psi_k\rangle\langle\psi_k|$ the fidelity coincides with the overlap between them, i.e. $f(\rho_1, \rho_2) = |\langle\psi_1|\psi_2\rangle|^2$.

[2]The experimental limitations on the measured system size can be imposed as separability constraints over the POVM elements.

One of the main advantages of the single-shot formalism that Bayesian approach allows is the possibility of using the symmetries in the optimisation. For instance, if we deal with a circular parameter that is codified via a unitary transformation and we have no prior information, that is our prior PDF is uniform (i.e. $p(\phi) = 1/2\pi$), we can always chose a *covariant measurement*. Such measurement are a generalisation of the POVM presented in Eq. (3.55) for the symmetric states discrimination. Firstly notice that the circular symmetry of the parameter ϕ, and therefore the group invariance of $f(\hat{\phi}, \phi)$, allows to rewrite the Eq. (3.88) as

$$F = \sum_x \int d\phi \, f(\hat{\phi}_x + \theta, \phi + \theta) \, \mathrm{tr} \left(U_\theta \, \Pi_x \, U_\theta^\dagger \, \rho_{\phi+\theta}^N \right) \frac{1}{2\pi} \tag{3.90}$$

for any real value θ. Such independence in the introduced parameter θ ensures that averaging on it does not modify the total value of F, i.e.

$$F = \int \frac{d\theta}{2\pi} \sum_x \int \frac{d\phi}{2\pi} f(\hat{\phi}_x + \theta, \phi + \theta) \, \mathrm{tr} \left(U_\theta \, \Pi_x \, U_\theta^\dagger \, \rho_{\phi+\theta}^N \right) \tag{3.91}$$

Again, the cyclical symmetry allows me to make the change of variable $\phi \to \phi - \theta$ which yields

$$F = \sum_x \int \frac{d\theta}{2\pi} \int \frac{d\phi}{2\pi} f(\hat{\phi}_x + \theta, \phi) \, \mathrm{tr} \left(U_{\theta+\hat{\phi}_x} \, U_{\hat{\phi}_x}^\dagger \, \Pi_x \, U_{\hat{\phi}_x} \, U_{\theta+\hat{\phi}_x}^\dagger \, \rho_\phi^N \right) \tag{3.92}$$

where I have used the relation of the cyclical unitary transformation $U_g \, U_h = U_{g+h}$. Now, by absorbing in each term of the sum the guessed value $\hat{\phi}_x$ in the auxiliary variable θ, i.e. making the change of variable $\theta \to \hat{\phi} = \theta + \hat{\phi}_x$ I get the final expression

$$\begin{aligned} F &= \int \frac{d\hat{\phi}}{2\pi} \int \frac{d\phi}{2\pi} f(\hat{\phi}, \phi) \, \mathrm{tr} \left(U_{\hat{\phi}} \left[\sum_x U_{\hat{\phi}_x}^\dagger \, \Pi_x \, U_{\hat{\phi}_x} \right] U_{\hat{\phi}}^\dagger \, \rho_\phi^N \right) = \\ &= \int \frac{d\hat{\phi}}{2\pi} \int \frac{d\phi}{2\pi} f(\hat{\phi}, \phi) \, \mathrm{tr} \left(U_{\hat{\phi}} \, \Omega \, U_{\hat{\phi}}^\dagger \, \rho_\phi^N \right) \end{aligned} \tag{3.93}$$

I have implicitly defined the seed operator $\Omega \equiv \sum_x U_{\hat{\phi}_x}^\dagger \, \Pi_x \, U_{\hat{\phi}_x}$, which is by construction positive. It can be used to define a *covariant POVM* with infinite number of elements $\tilde{\Pi}_{\hat{\phi}} = U_{\hat{\phi}} \, \Omega \, U_{\hat{\phi}}^\dagger$. This elements are labelled directly by the estimated value $\hat{\phi}$, so the estimator has been absorbed in the POVM itself. Furthermore, using the cyclic property of the trace and the group invariance of the fidelity, one of the integral can be made straightforwardly yielding to the simplified expression

$$F = \int \frac{d\phi}{2\pi} f(0, \phi) \, \mathrm{tr} \left(\Omega \, \rho_\phi^N \right) \tag{3.94}$$

As an example consider the case in which we want to estimate a phase ϕ from a system composite of N indistinguishable two-dimensional subsystems. The global state is the invariant under permutations of the subsystems and hence it can be express in a block-diagonal form (see Section 2.2.4). For sake of simplicity, let me consider first that the state has only the projection on a single subspace, say the symmetric subspace $j = N/2$. The extension to the more general full-projection states is straightforward. With a suitable choice of axis the unitary operator that encodes the phase can be express as a diagonal matrix $U_\phi = \sum_{m=-N/2}^{N/2} \mathrm{e}^{i\,m\,\phi}\ |m\rangle\langle m|$, i.e.

$$\rho_\phi^N = \sum_{m,m'=-N/2}^{N/2} \rho_{m,m'}\ \mathrm{e}^{i(m-m')\phi}\ |m\rangle\langle m'| \tag{3.95}$$

where I have simplified the notation being $|m\rangle \equiv |j = N/2, m\rangle$. With this encoding the fidelity between two single-copy states is

$$f(\hat{\phi}, \phi) = \tfrac{1}{2}\left(1 + \cos(\hat{\phi} - \phi)\right). \tag{3.96}$$

Let me write the seed operator as $\Omega = \sum_{m,m'} \omega_{m,m'} |m\rangle\langle m'|$. The identity resolution implies that the diagonal element must be $\omega_{m,m} = 1$, and positivity constraint the off-diagonal element with the condition $\omega_{m,m'} \leq \sqrt{\omega_{m,m}\,\omega_{m',m'}} = 1$. Plugging all this into the Eq. (3.94)

$$\begin{aligned} F &= \frac{1}{2}\left(1 + \sum_{m,m'=-N/2}^{N/2} \omega_{m,m'}\rho_{m',m} \int \frac{d\phi}{2\pi}\ \mathrm{e}^{i(m-m')\phi} \cos\phi \right) = \\ &= \frac{1}{2}\left(1 + \sum_{m=-N/2}^{N/2-1} \omega_{m,m+1}\rho_{m+1,m}\right) = \\ &\leq \frac{1}{2}\left(1 + \sum_{m=-N/2}^{N/2-1} \rho_{m+1,m}\right) \equiv F^\star \end{aligned} \tag{3.97}$$

where I used the Kronecker delta definition $\delta_{m,m'} = \int \frac{d\phi}{2\pi}\ \mathrm{e}^{i(m-m')\phi}$. The inequality can be saturated by choosing the seed operator to be $\Omega = |\Psi\rangle\langle\Psi|$ with $|\Psi\rangle = (N{+}1)^{-1/2} \sum_m |m\rangle$. For the case of state with projection on all the subspaces, the optimal fidelity is the average of the optimal $F^\star$ in Eq. (3.97) over all the subspaces weighted with the probability and the multiplicity of each one, and the optimal seed operator is the direct sum of projectors like the one described above.

Let me now consider the case of a pure state for which $\rho_{m,m'} = c_m c_{m'}^*$, with $\sum_m |c_m|^2 = 1$. One might wonder which is the state that maximises $F^\star$.

By optimising the Eq. (3.94) subject to the normalisation constraint one find that the optimal state to encode a phase is (see **Berry2000** and Chapter 4)

$$|\psi\rangle = \sum_m \sqrt{\frac{1}{N+2}} \cos\left(\frac{m+1}{N+2}\pi\right) |m\rangle \tag{3.98}$$

That is, an state with a wide support and with the maximum centred. This do not coincide with the optimal state found with the Frequentist approach, as we will see in the next section.

3.4.2 Frequentist approach

As in the classical case, the quantum Frequentist approach to parameter estimation treat the unknown parameter as a fixed value. In this sense, if the input state and the measurement are specified, the CRLB can be applied over the resulting probability distribution $p(x;\phi) = \mathrm{tr}\,(\Pi_x\,\rho_\phi^N)$. However, a more general bound can be derived in which the measurement must not be fixed: the *Quantum Cramér-Rao* (QCR) lower bound [Helstrom, 1976; Holevo, 1982]. First I define the *quantum Fisher information*

$$I_Q(\rho_\phi^N) = \mathrm{tr}\left(\rho_\phi^N\,L\left[\rho_\phi^N\right]^2\right) \tag{3.99}$$

where the hermitian operator $L\,[\rho]$ is the *symmetric logarithmic derivative* (SLD) defined implicitly with the relation $\frac{\partial\rho_\phi}{\partial\phi} = \frac{1}{2}\left(\rho_\phi\,L\,[\rho_\phi] + L\,[\rho_\phi]\;\rho_\phi\right)$. Like in the classical case (Sec. 3.2.2), the quantum Fisher information can be understood as the sensitivity on the parameter of the state ρ_ϕ^N. In this sense, for any probability distribution $p(x;\phi)$ that can be get by measuring the state, the quantum Fisher information $I_Q\left(\rho_\phi^N\right) \geq I\,[p(x;\phi)]$. Therefore, the MSE of a quantum estimation protocol can be lower-bounded with a quantum version of Eq. (3.26), i.e.

$$\mathrm{Var}_\nu\left[\rho_\phi^N\right] \geq \frac{1}{\nu\,I_Q\left[\rho_\phi^N\right]} \tag{3.100}$$

The subscript ν refers to the number of independent repetitions required in the Frequentist approach. In the work Braunstein and Caves [1994] the authors show that a projective measurement in the eigenbasis of the SLD operator $L[\rho_\phi^N]$ produces a PDF that its classical information attains the $I_Q\left[\rho_\phi^N\right]$. Note that, since the SLD operator depends on the true value of the parameter, the optimal POVM obtained depends also on it. This is consistent with the locality of the Frequentist approach. In this sense, the QCR bound will be ensured to be attainable in the same situations as the classical Cràmer-Rao bound: not only in the local scheme but also in the limit of infinitely many independent repetitions $\nu \to \infty$. Moreover, as such conditions implies a high previous information and an infinite number of resources, this bound is said to be the ultimate bound of quantum parameter information.

For the phase estimation pure case discussed in the Bayesian approach section, the Frequentist approach leads to a very different result. Instead of a state with a sinusoidal profile in Eq. (3.98), the optimal state for encoding a phase within this approach has only support on the extremes, i.e.

$$|\psi\rangle = \frac{1}{\sqrt{2}}\left(|m=-N/2\rangle + |m=N/2\rangle\right) \tag{3.101}$$

It is the so-called $NOON$-state[3]. The reason because the two approaches gives such different optimal states is that they refer to different situations: the Bayesian approach consider a global estimation with no independent repetitions, whereas the Frequentist approach requires a very large previous information and an infinite amount of resources.

3.5 Probabilistic quantum parameter estimation

The aim of this Section is to introduce some generalities about the probabilistic estimation protocols that will be presented in the subsequent chapters. In them a new ingredient is added: the possibility that the experimentalist is allowed to decide whether to provide a guess or abstain to doing so, or in other words, some of the outcomes are associated to a "decline output" in pursuance of an enhancement in performance of the protocol. Of course, this decision cannot be based on the actual state of the system (which is unknown by definition) but rather on the result of a measurement. This relaxation of the original setting is very useful because it enables the experimentalist to post-select her measurement outcomes in order to provide a more accurate guess. That is, the possibility of abstaining enables her to discard instances where the measurement outcome turns out not to be informative enough. I will find that abstention can provide an important advantage, specially in noisy scenarios. Post-selection is a widely used tool in quantum information, particularly in experimental scenarios, where one has special demands or constrains. A form of abstention has been already explored in state discrimination, as presented in Section 3.3.3.

[3]The name NOON-state comes from the expression that this state takes in a basis labelled by the occupation number of each mode $|0\rangle$ and $|1\rangle$. In this basis the state in Eq. (3.101) is $|\psi\rangle = \frac{1}{\sqrt{2}}\left(|N,0\rangle + |0,N\rangle\right)$

In this work I will use the term probabilistic strategy to refer to a process that probabilistically ends in two possible cases: performs the expected task (e.g. outputs an estimate, prepares an state...) or decline. The task itself can be stochastic, but here the word "probabilistic" refers to the fact that it is not sure to obtain an outcome. Hence, it is characterized by giving the figure of merit that describes how well you perform the task (when it is performed) and supplement it, of course, with the deliver/decline probability. It is important to remark that this figure of merit takes only into account a single-shoot of the measurement, since no average over different repetitions is performed. Therefore, this scheme can be only be tackle with a Bayesian approach, in which no independent repetitions of the measurement are required.

To incorporate this abstention possibility, the formalism introduced at the beginning of Section3.4 has to be slightly modified: the guessing rule $\mathcal{G}$ will apply now only on a subset $\mathrm{D} \subseteq \mathrm{X}$ of outcomes of the measurement $\mathcal{M}$ and the outcomes outside this subset will produce a "decline" output. As in the case of probabilistic state discrimination, the measurement is represented by a POVM with a set of operators Π_x each one associated with a conclusive outcome $x \in \mathrm{D}$, plus another operator Π_0 that encompasses the remaining outcomes that yield inconclusive answers. It is useful also in this setting to split the protocol into a two-step process: a first step consisting in a probabilistic filter that modifies the state of the system and a second one that corresponds to the measurement itself. To do so, I define the operators $\bar{\Pi}_x = \bar{\Pi}_0^{-1/2}\Pi_x\bar{\Pi}_0^{-1/2}$ with $\bar{\Pi}_0 = \mathbb{1} - \Pi_0 \geq 0$. These new operator are positive and sum up to the identity by construction. Therefore they form a new POVM that have no abstention element. The probability of getting an outcome χ can be rewritten in terms of this new POVM as

$$p(x \mid \phi) = \mathrm{tr}\,(\Pi_x\rho_\phi) = \mathrm{tr}\,(\bar{\Pi}_x\bar{\rho}_\phi)\,\mathrm{tr}\,(\bar{\Pi}_0\rho_\phi) \quad \text{with} \quad \bar{\rho}_\phi \equiv \frac{\bar{\Pi}_0^{1/2}\rho_\phi\bar{\Pi}_0^{1/2}}{\mathrm{tr}\,(\bar{\Pi}_0\rho_\phi)}. \tag{3.102}$$

The operators $\bar{\Pi}_0^{1/2}$ can be viewed as the Krauss operator that characterizing a probabilistic filter that transforms the state of the system into $\bar{\rho}_\phi$ before the measurement is performed on it. Hence, the second term in the last expression corresponds to the probability that the probabilistic filter success in transforming it, i.e. $p(succ|\phi) = \mathrm{tr}\,(\bar{\Pi}_0\rho_\phi)$. Then, the averaged probability that the filter performs successfully is

$$p_{succ} = \sum_{x\in\mathrm{D}} \int \frac{d\phi}{2\pi}\,\mathrm{tr}\,(\Pi_x\rho_\phi) = \int \frac{d\phi}{2\pi}\,\mathrm{tr}\,(\bar{\Pi}_0\rho_\phi), \tag{3.103}$$

and its counterpart, the probability of abstention

$$Q = 1 - p_{succ} = \int \frac{d\phi}{2\pi}\,\mathrm{tr}\,(\Pi_0\,\rho_\phi). \tag{3.104}$$

With all this, we can define a version of Eq. (3.88) for probabilistic parameter estimation protocols

$$F = \sum_{x\in \mathrm{D}} \int \frac{d\phi}{2\pi} \, \frac{\mathrm{tr}\,(\Pi_x \, \rho_\phi^N)}{p_{succ}} f(\hat{\phi}_x, \phi) \, p(\phi). \tag{3.105}$$

The terms $\mathrm{tr}\,(\Pi_x \rho_\phi)/p_{succ}$ are the probabilities of getting the outcome x with the conditions that the actual state is ρ_ϕ and we get not an inconclusive outcome, and using the Eq. (3.103) one can check that they sum up to unity. This new notation allows us to use the results for the deterministic protocols of estimation and it will only remain to optimize over the probabilistic filter, or what is the same, to find the optimal $\bar{\rho}_0$. Moreover, the covariant POVM presented in Section 3.4.1 has its version for probabilistic settings. From Eq. (3.104) one can use the circular symmetry of the parameter to show that the abstention element of the POVM can be choosen group-invariant, i.e.

$$\begin{aligned}
Q &= \int \frac{d\phi}{2\pi} \, \mathrm{tr}\,(\Pi_0 \, U_\theta \, \rho_{\phi-\theta} \, U_\theta^\dagger) = \\
&= \int \frac{d\phi}{2\pi} \int \frac{d\theta}{2\pi} \, \mathrm{tr}\,(U_\theta^\dagger \, \Pi_0 \, U_\theta \, \rho_{\phi-\theta}) = \\
&= \int \frac{d\phi}{2\pi} \, \mathrm{tr} \left[\left(\int \frac{d\theta}{2\pi} U_\theta^\dagger \, \Pi_0 \, U_\theta \right) \rho_\phi \right] = \\
&\equiv \int \frac{d\phi}{2\pi} \, \mathrm{tr} \left[\tilde{\Pi}_0 \, \rho_\phi \right]
\end{aligned} \tag{3.106}$$

where by definition $\left[\tilde{\Pi}_0, U_\phi\right] = 0$. Now, following similar steps as in the deterministic case, it is easy to check that the set comprised of the continuous-labelled operators $\tilde{\Pi}_{\hat{\phi}} = U_{\hat{\phi}} \, \Omega \, U_{\hat{\phi}}^\dagger$ with $\Omega \equiv \sum_{x\in \mathrm{D}} U_{\hat{\phi}_x}^\dagger \, \Pi_x \, U_{\hat{\phi}_x}$, plus the invariant operator $\tilde{\Pi}_0$ form a POVM that produces the same fidelity of Eq. (3.105). Again, the use of this symmetric POVM simplifies the expressions for the fidelity, and in this case also for the abstention probability to

$$F = \int \frac{d\phi}{2\pi} \, f(0,\phi) \, \frac{\mathrm{tr}\left(\Omega \, \rho_\phi^N\right)}{1-Q} \tag{3.107}$$

$$Q = \mathrm{tr}\left(\tilde{\Pi}_0 \, \rho_{\phi=0}^N\right) \tag{3.108}$$

For examples in the use of this formalism see the next chapters, where different cases will be treated in detail, with different initial states and different parameter encodings.

CHAPTER 4

Probabilistic estimation of pure states

In this chapter we address the problem of quantum parameter estimation with pure states of N qubits (or Rydberg atomic states of total angular momentum $N/2$). More precisely, we deal with an infinite covariant family of such states, parametrised by some continuous variables, and we aim to estimate the values of these variables for a given sample state by performing suitable measurements on it. Specifically we study the problems of phase and spatial directions estimation, which are assumed to be encoded in a given N-spin state. We also study the reference frame case, i.e, the estimation of three mutually orthogonal directions. The later will be referred to as frame estimation for brevity.

In this Chapter we present a general technique to obtain the asymptotic form of pure state parameter estimation problems, with or without abstention, that is interesting on its own. The main idea is that the components of the encoding state can be viewed as a discretization of some continuous function $\varphi(t)$ on the unit interval $[0,1]$, and likewise, the problem of maximizing the fidelity over those components can be viewed as a discretization of a constrained variational problem. The solution gives the asymptotic expression of the accuracy. This solution can be worked out analytically for many physically relevant settings.

For finite N, we have also formulated the optimisation problem as a semidefinite programming problem (SDP), which allows us to obtain numerical results in a very efficient way.

4.1 Definition of the problems

For phase estimation the covariant family we are referring to is the set of states of the form $\{|\Psi(\theta)\rangle = U(\theta)|\Psi_0\rangle\}_{\theta\in[0,2\pi)}$, where $U(\theta)$ stands for the unitary transformation $U(\theta)|j\rangle = \mathrm{e}^{i\theta j}|j\rangle$, $|\Psi_0\rangle$ is a fiducial state, which in the eigenbasis of $U(\theta)$ can be written as $|\Psi_0\rangle = \sum_{j=0}^{n} c_j|j\rangle \in (\mathbb{C}^2)^{\otimes n}$, and the number of qubits is $N = n$. The components c_j are given arbitrary coefficients, subject to the normalization condition $\sum_{j=0}^{n}|c_j|^2 = 1$. For direction estimation I consider instead $\{|\Psi(\boldsymbol{n})\rangle = U_{\boldsymbol{n}}|\Psi_0\rangle\}_{\boldsymbol{n}\in\mathbb{S}^2}$, where $U_{\boldsymbol{n}}$ stands for the unitary representation of the rotation that takes $\boldsymbol{z}$ (the unit vector in the z-axis; likewise, $\boldsymbol{x}$ and $\boldsymbol{y}$ stand for the other two unit vectors) into $\boldsymbol{n}$. The fiducial state is now given by $|\Psi_0\rangle = \sum_{j=0}^{n} c_j|j,0\rangle$, which may be thought of as pointing along the z-axis (in the sense that it is invariant under rotations about that axis), $N = 2n$, and we use the standard notation $|j,m\rangle$ for the total angular momentum eigenstates. Therefore, we will use the representation presented in Section 2.2.4. The choice $m = 0$ is both for simplicity and also because the optimal state for direction encoding is known to have null total magnetic number [Bagan *et al.*, 2000], however the method can be extended to any m. More general states, i.e., those that are not eigenstates of J_z, do not fit into this pure state framework, since for the sake of direction estimation the subset $\{\mathrm{e}^{-i\gamma J_z}|\Psi_0\rangle\}_{\gamma\in[0,2\pi)}$ that encodes $\boldsymbol{z}$ is equivalent to the mixed state $\rho_0 = \int_0^{2\pi}(d\gamma/2\pi)\mathrm{e}^{-i\gamma J_z}|\Psi_0\rangle\langle\Psi_0|\mathrm{e}^{i\gamma J_z}$, and this case will be treated in the following Chapter.

For frame estimation, the relevant family of states is

$$\{|\Psi(g)\rangle = U(g)|\Psi_0\rangle\}_{g\in\mathbb{S}^3}\,, \tag{4.1}$$

where g stands for the three Euler angles: $g = (\alpha,\beta,\gamma)$. They specify the rotation that takes the axes x, y and z into those of the Cartesian frame we wish to estimate, with unit vectors $(\boldsymbol{n}^1,\boldsymbol{n}^2,\boldsymbol{n}^3)$. It can be shown that optimality requires a fiducial state of the form $\sum_{j=0}^{n} c_j\left(\sum_{m=-j}^{j}|j,m,\alpha_m\rangle\right)/\sqrt{2j+1}$, where $N = 2n$ and the third quantum number in the ket, α_m, labels the degeneracy of the representation of angular momentum j (See Section 2.2.4). Except for the representation of highest angular momentum, $j = n$, for each $j < n$ we have (maximally) entangled the magnetic number m with the degeneracy number α_m. This entanglement with 'ancillary' degrees of freedom that are invariant under the action of the group is responsible for an important enhancement in the estimation precision. We note in passing that this degeneracy is known to be useless for single direction estimation, and thus we dropped the corresponding label there. Indeed, following the symmetry argument used at the end of the previous paragraph, any entanglement between magnetic number and degeneracy labels would in effect turn into an incoherent sum on subspaces of different m values, which is clearly suboptimal.

From a formal point of view, it will be seen that the optimization of the frame estimation protocol for this family of states is equivalent to that of phases for large N. Thus, we find it more interesting to ignore the degeneracy of the representations and consider instead the family generated by the fiducial state $|\Psi_0\rangle = \sum_{j=0}^{n} c_j|j,j\rangle$. States of this form could be produced if, e.g., a hydrogen atom in a Rydberg state of total angular momentum up to n is used instead of N spins Peres and Scudo [2001]. In this scenario, the optimal encoding state for a Cartesian frame is known to belong to this family, but it does not lead to a Heisenberg scaling precision. Also in this case, the method we will introduce can be applied to more general pure states.

For all these estimation problems, a finite acceptance rate $\bar{Q} = 1 - Q$ suffices to lower the coefficient of the leading order in the asymptotic expansion of the average error in inverse powers of N. If an exponentially vanishing acceptance rate is affordable, the leading order in this expansion becomes $1/N^2$, thus attaining the Heisenberg limit, except for frame estimation with Rydberg states. It will be shown that the effect of abstention can be understood in terms of a probabilistic map from the original family to a better one (closer to optimal), $\{\tilde{\Psi}(\theta)\}_{\theta\in[0,2\pi)}$ (or $\{\tilde{\Psi}(\boldsymbol{n})\}_{\boldsymbol{n}\in\mathbb{S}^2}$, etc.), which fails with probability Q.

4.2 General framework

The problems of phase, direction and frame estimation described above can be treated in a unified framework by writing $U(\theta) = U(g)$, $|\Psi(\theta)\rangle = |\Psi(g)\rangle$, where $g \in \mathbb{S}^1$; and $U_{\boldsymbol{n}} = U(g)$, $|\Psi(\boldsymbol{n})\rangle = |\Psi(g)\rangle$, where $g \in \mathbb{S}^2$. Since the magnetic number is fixed to zero (j) for direction (frame) estimation, we also drop this quantum number and write $|j,0\rangle \equiv |j\rangle$ ($|j,j\rangle \equiv |j\rangle$). Then, for the three problems we have a family of states $\{|\Psi(g)\rangle = U(g)|\Psi_0\rangle\}_{g\in\mathbb{S}^d}$, where $d = 1$, 2, 3 for phase, direction, and frame estimation, respectively. As already mentioned above, in direction (frame) estimation the fiducial state $|\Psi_0\rangle$ can be thought of as encoding the unit vector $\boldsymbol{z}$ [the cartesian frame $(\boldsymbol{x},\boldsymbol{y},\boldsymbol{z})$]. Similarly, in phase estimation, $|\Psi_0\rangle$ can be interpreted as encoding the reference unit vector $\boldsymbol{x}$ (to which we assign a zero phase), and $U(g)$ as a rotation of (Euler) angle $\alpha = \theta$ around the z-axis [1]. Hence, in this unified framework, we can define a cost function in terms of the (quadratic) error per axis is: i.e. $e_1(g, g_\chi) = |\boldsymbol{n} - \boldsymbol{n}_\chi|^2$ for phase and direction estimation,

[1]Note that from this point of view the the label $j = 0, \ldots, n$ would correspond to the magnetic quantum number taking values $m = -n/2, \ldots, n/2$.

and the total error $e_3(g, g_\chi) = \sum_{a=1}^{3} |\boldsymbol{n}^a - \boldsymbol{n}_\chi^a|^2$ for frame estimation. In these expressions, the subscript χ specifies that the estimate is based on the outcome χ of a generalized measurement that will be introduced below. These errors are related to the 'relative rotation' $U^\dagger(g_\chi)U(g) = U(g_\chi^{-1}g)$ through

$$e_1(g, g_\chi) = 2 - 2\langle 1, 0|U(g_\chi^{-1}g)|1, 0\rangle, \tag{4.2}$$

$$e_3(g, g_\chi) = 6 - 2\sum_{m=-1}^{1} \langle 1, m|U(g_\chi^{-1}g)|1, m\rangle, \tag{4.3}$$

where we recognize the sum in (4.3) as the character of $U(g_\chi^{-1}g)$ in the $j = 1$ representation. Note that $0 \leq e_1(g, g_\chi) \leq 4$, and $0 \leq e_3(g, g_\chi) \leq 8$ (we can at most get two axes completely wrong since we assume right-handed Cartesian frames). As a figure of merit, the fidelity $f(g, g_\chi) = (1+\boldsymbol{n}\cdot\boldsymbol{n}_\chi)/2 = 1 - e_1(g, g_\chi)/4$ is most commonly used in phase and direction estimation. One has $0 \leq f(g, g_\chi) \leq 1$, where 1 corresponds to perfect estimation. For frame estimation one can also define a fidelity with the same range of values as $f(g, g_\chi) = 1 - e_3(g, g_\chi)/8$. These fidelities are also trivial functions of the relative rotation $U(g_\chi^{-1}g)$ in the $j = 1$ representation through Eqs. (4.2) and (4.3).

The generalized measurements we are interested in are characterized mathematically by a positive operator valued measure (POVM) $\Pi = \{\Pi_\chi\}_{\chi\in C} \cup \{\Pi_0\}$, where Π_0 and each Π_χ are non-negative operators that add up to the identity, i.e., $\Pi_0 + \sum_{\chi\in C} \Pi_\chi = \mathbb{1}$, C is the set of conclusive outcomes (from which an estimate is proposed) and Π_0 outputs 'abstention'. The probability of such abstention taking place is

$$Q = \int_{\mathbb{S}^d} d^d g \, \langle\Psi(g)|\Pi_0|\Psi(g)\rangle, \tag{4.4}$$

and $\bar{Q} = 1 - Q$ is the acceptance probability. In this notation, the average fidelity in Eq. (3.105) is rewritten as

$$F(Q) = \frac{1}{\bar{Q}} \sum_\chi \int_{\mathbb{S}^d} d^d g \, f(g, g_\chi) \, \langle\Psi(g)|\Pi_\chi|\Psi(g)\rangle. \tag{4.5}$$

Recall that the first factor $\bar{Q}^{-1}$ ensures the normalisation of the weights in the average. In Eqs. (4.4) and (4.5), $d^d g$ stands for the (normalized) 'volume' elements $dg = d\alpha/(2\pi)$ (for $d = 1$), $d^2 g = \sin\beta \, d\alpha \, d\beta/(4\pi)$, and $d^3 g = \sin\beta \, d\alpha \, d\beta \, d\gamma/(8\pi^2)$. They are invariant measures on $\mathbb{S}^d$, i.e., $d^d(gg') = d^d g$.

At this point, we summon the interpretation of the role of abstention in this optimization problem that we presented in the Section 3.5: each initial state $|\Psi(g)\rangle$ is transformed into a new $|\tilde{\Psi}(g)\rangle$ that encodes the unknown parameter(s) g in a more efficient way. In this way, estimation with abstention can be reduced to a standard estimation problem (without abstention) by simply introducing the new POVM $\tilde{\Pi}$, with elements given by

$$\tilde{\Pi}_\chi \equiv (\mathbb{1}-\Pi_0)^{-1/2}\,\Pi_\chi\,(\mathbb{1}-\Pi_0)^{-1/2}\,, \tag{4.6}$$

and the new family of (normalized) states

$$\left\{|\tilde{\Psi}(g)\rangle \equiv \frac{(\mathbb{1}-\Pi_0)^{1/2}}{\bar{Q}^{1/2}}|\Psi(g)\rangle\right\}_{g\in\mathbb{S}^d}. \tag{4.7}$$

as it has been introduced in the paragraph above Eq. (3.102). With these two definitions we can write the fidelity as

$$F(\Pi_0) = \sum_\chi \int_{\mathbb{S}^d} d^dg\, f(g,g_\chi)\,\langle\tilde{\Psi}(g)|\tilde{\Pi}_\chi|\tilde{\Psi}(g)\rangle, \tag{4.8}$$

where it is emphasized that this expression depends on the choice of Π_0.

The map improves the estimation precision by effectively increasing the distinguishability between the signal states, therefore it can only be implemented in a probabilistic fashion (it succeeds with probability $\bar{Q}$). This stochastic map is fully specified by the optimal choice of Π_0:

$$F(Q) = \max_{\Pi_0\,:\,\text{Eq. (4.4)}} F(\Pi_0). \tag{4.9}$$

Although this may seem a difficult optimization problem, the covariance of the family of states produces a huge simplification, as explained in Section 3.5. That is to say, already from Eqs. (4.4) and (4.5) one can easily see that the optimal POVM can be chosen to be covariant under the set of unitaries $\{U(g)\}_{g\in\mathbb{S}^d}$. In particular this means that Π_0 can be taken invariant under the corresponding unitary group. For $d=1$ this is just the group $U(1)$. For $d=2$ ($d=3$) the integral over the 2-sphere (3-sphere) can be turned into (is) a $SU(2)$ group integral. Thus, Shur's lemma can be applied to all the cases, which results in Π_0 being proportional to the identity on each irreducible block: $\Pi_0=\sum_j f_j|j\rangle\langle j|$ (phase estimation), $\Pi_0=\sum_j f_j\sum_m |j,m\rangle\langle j,m| \equiv \sum_j f_j \mathbb{1}_j$ (direction and frame estimation). Hence, the maximization in Eq. (4.9) is over $\{f_j\,:\,0\le f_j\le 1\}_{j=0}^n$. Note that the transformed set of states $\{|\tilde{\Psi}(g)\rangle\}_{g\in\mathbb{S}^d}$ is also a covariant family, just as the original one. The corresponding reference state is

$$|\tilde{\Psi}_0\rangle = \sum_{j=0}^n \frac{c_j\sqrt{\bar{f}_j}}{\sqrt{\bar{Q}}}|j\rangle = \sum_{j=0}^n \xi_j|j\rangle \equiv |\xi\rangle\,, \tag{4.10}$$

where $\bar{f}_j \equiv 1 - f_j$. From Eq. (4.4), and using Shur's lemma, we find

$$Q = \sum_{j=0}^{n} |c_j|^2 f_j = 1 - \sum_{j=0}^{n} |c_j|^2 \bar{f}_j. \tag{4.11}$$

Thus, $|\tilde{\Psi}_0\rangle = |\xi\rangle$ is a normalized state, as it should, i.e., $\sum_j |\xi_j|^2 = 1$. Since the transformed states are still covariant, we can choose the POVM $\tilde{\Pi}$ to be the well known continuous and covariant POVM for each of the problems at hand [Holevo, 1982; Bagan *et al.*, 2000]: $\{\tilde{\Pi}_g = U(g)|\Phi_d\rangle\langle\Phi_d|U^\dagger(g)\}_{g\in\mathbb{S}^d}$, where the unnormalized state $|\Phi_d\rangle$ is given by

$$|\Phi_1\rangle = \sum_{j=0}^{n} |j\rangle; \quad |\Phi_{2,3}\rangle = \sum_{j=0}^{n} \sqrt{2j+1}|j\rangle. \tag{4.12}$$

Note that g plays the role of χ, i.e., g specifies the different outcomes of the measurement. Hereafter, it is assumed that the states have non-negative coefficients $c_j \geq 0$ (and hence $\xi_j \geq 0$). This is a valid assumption since any phases present in the coefficients c_j (or ξ_j) can be absorbed by the above POVM's. This result makes the calculation of the fidelity $F(\Pi_0)$ straightforward:

$$F(\Pi_0) = \frac{1}{2} + \frac{1}{2} \langle\xi|\, \mathsf{M}\, |\xi\rangle\,, \tag{4.13}$$

where in the canonical basis, $\{|j\rangle\}_{j=0}^{n}$, M is a real matrix of tridiagonal form

$$\mathsf{M} = \begin{pmatrix} h_0^d & a_1^d & & & & 0 \\ a_1^d & \ddots & \ddots & & & \\ & \ddots & h_{n-2}^d & a_{n-1}^d & & \\ & & a_{n-1}^d & h_{n-1}^d & a_n^d & \\ 0 & & & a_n^d & h_n^d \end{pmatrix}, \tag{4.14}$$

with

$$a_j^1 = \frac{1}{2}; \quad a_j^2 = \frac{j}{\sqrt{4j^2-1}}, \quad a_j^3 = \frac{1}{2}\sqrt{\frac{2j+1}{2j+3}},$$
$$h_j^1 = h_j^2 = 0, \qquad h_j^3 = -\frac{1}{2(j+1)}, \tag{4.15}$$

where we recall that the superscripts 1, 2 and 3 refer to phase, direction and frame estimation, respectively.

At this point one can easily check the statement in the introduction that frame estimation with the family generated by the fiducial state

$$|\Psi_0\rangle = \sum_{j=0}^{n} c_j \,(\sum_{m=-j}^{j} |j,m,\alpha_m\rangle)/\sqrt{2j+1} \tag{4.16}$$

is formally equivalent to phase estimation for large n. For this family, the diagonal entries of the matrix M are zero with the exception of $h_0 = -1/2$ and $h_n = -1/(2n+2)$, whereas the off-diagonal ones are $a_j = 1/2$, for $0 \leq j \leq n-1$ and $a_n = 1/(2\sqrt{2n+1})$. Thus, except for four entries, M is the same for phase and frame estimation. For very large n, this finite differences have no effect at leading order and the asymptotic result obtained for phases also hold for frames when the degeneracy of the representations is used in the encoding.

Here, we have given the explicit form of M for the particular fiducial states under study . However, it is worth noting that for general states the matrix M will always have a tridiagonal structure and hence the methods that we use readily apply. As shown in Bagan *et al.* [2000, 2001], this structure is a generic feature that stems from the fact that the fidelity $f(g, g_\chi)$ is a linear function of $\langle 1, m|U(g_\chi^{-1}g)|1, m'\rangle$ ($j = 1$ representation). Its appearance in the integrant of (4.8) enforces selection rules that prevent the presence of other off-diagonal elements in M.

The maximization over $\{f_j\}$ of (4.13) can be turned into a maximization over the transformed states $|\xi\rangle$, namely:

$$\Delta \equiv \max_{|\xi\rangle} \langle\xi|\, \mathsf{M}\, |\xi\rangle \,, \tag{4.17}$$

subject to the constraints

$$\langle\xi|\xi\rangle = \sum_{j=0}^{n} \xi_j^2 = 1, \tag{4.18}$$

$$\xi_j = \sqrt{\frac{\bar{f}_j}{\bar{Q}}}c_j \leq \frac{c_j}{\sqrt{\bar{Q}}} \equiv \lambda c_j, \quad \lambda \geq 1. \tag{4.19}$$

Then, the maximum fidelity for a given rate of abstention Q is $F(Q) = (1+\Delta)/2$.

For large enough abstention rates (i.e. large enough values of λ) the constraint (4.19) has no effect (provided all components c_j are different from zero) and Δ becomes the maximum eigenvalue of the matrix M. In this case, $F(Q \to 1) = F^*$ is the maximum fidelity that can be achieved by optimizing the components of the fiducial state; these are given by the corresponding eigenvector $|\xi^*\rangle$ of M. The resulting fiducial state thus generates the optimal signal states $|\Psi(g)\rangle$. From (4.19) it is straightforward to obtain the critical acceptance rate

$$\bar{Q}^* = \min_j \frac{c_j^2}{\xi_j^{*2}}. \tag{4.20}$$

That is, for abstention rates such that $Q \geq Q^* = 1 - \bar{Q}^*$ the fidelity attains its absolute maximum value F^* (and higher rates cannot improve the estimation quality). In the other extreme, when no abstention is allowed ($Q = 0$), the solution is determined by the constraints $\xi_j = c_j$ (no maximization is possible), and $\Delta = \langle c|\mathsf{M}|c\rangle$.

For intermediate values of $Q \in (0, Q^*)$ the problem becomes more tricky. For moderate values of n one can use standard non-linear optimization packages to solve the above constrained convex optimization problem, Eqs. (4.17)–(4.19). This can also be easily cast as a semidefinite programming (SDP) problem. The SDP approach is efficient and, furthermore, provides rigorous bounds on the precision of the solution. One simply linearizes these equations by introducing a SDP (positive operator) variable B to play the role of $|\xi\rangle\langle\xi|$. The SDP form of Eqs. (4.17) and (4.19) is then

$$\Delta \equiv \max_{\mathsf{B}} \operatorname{tr}(\mathsf{MB}), \tag{4.21}$$

subject to the constraints

$$\begin{aligned} &\operatorname{tr}\mathsf{B} = 1, \quad \mathsf{B} \geq 0, \\ &\mathsf{B}_{jj} \leq \frac{|c_j|^2}{\bar{Q}} \equiv \lambda^2 |c_j|^2, \quad \lambda \geq 1. \end{aligned} \tag{4.22}$$

One can easily prove that the optimal B for this problem must necessarily have rank one: since all the entries of M are non-negative, $\operatorname{tr}(\mathsf{MB})$ increases with increasing values of the off-diagonal entries $\mathsf{B}_{i,i+1}$. Their maximum value consistent with positivity is given by rank one matrices. Therefore, the optimal B is of the form $|\xi\rangle\langle\xi|$ and the SDP solution provides in turn a solution of Eqs. (4.17)–(4.19).

However, the main focus of this Chapter is on the regime of asymptotically large n and, in particular, on presenting an approach that enables obtaining analytical expressions in this regime, thus complementing the SDP analysis. we will first introduce and discuss in some detail the approach for phase estimation. The generalization to direction and frame estimation will be discussed afterwards.

4.3 Asymptotic regime: phase estimation

Here we consider the problem of phase estimation, for which $\langle\xi|\,\mathsf{M}\,|\xi\rangle$ can be cast as

$$\langle\xi|\,\mathsf{M}\,|\xi\rangle = \sum_{j=0}^{n-1} \xi_j \xi_{j+1}. \tag{4.23}$$

This expression can be easily rewritten as

$$\langle \xi | \, \mathsf{M} \, | \xi \rangle = 1 - \frac{1}{2} \left[\sum_{j=0}^{n-1} (\xi_{j+1} - \xi_j)^2 + \xi_0^2 + \xi_n^2 \right], \tag{4.24}$$

where the first term (unity) results from using the normalization condition (4.18). Instead of maximizing this expression, we will equivalently minimize $S \equiv 1 - \langle \xi | \, \mathsf{M} \, | \xi \rangle$. A slight difficulty arises here because of the inequality constraints in (4.19). To deal with them we need to use the so called Karush-Kuhn-Tucker (KKT) conditions (see e.g., Boyd and Vandenberghe [2004] and Section 3.3.3), which are a generalization of the Lagrange method. We first have to introduce a multiplier for each constraint: $b^2/2$; s_j, $j = 0, \ldots, n$; much in the same way as the Lagrange method requires. Hence, we will find the local minima of

$$\begin{aligned} S &= \frac{1}{2} \left[\sum_{j=0}^{n-1} (\xi_{j+1} - \xi_j)^2 + \xi_0^2 + \xi_n^2 \right] \\ &- \frac{b^2}{2} \left(\sum_{j=0}^{n} \xi_j^2 - 1 \right) + \sum_{j=0}^{n} s_j \left(\xi_j - \lambda c_j \right). \end{aligned} \tag{4.25}$$

Besides the constraints specified in (4.19), which are referred to as *primal feasibility* conditions, we also need to impose the so called *dual feasibility* conditions,

$$s_j \geq 0, \quad j = 0, 1, \ldots, n, \tag{4.26}$$

and, finally,

$$s_j \left(\xi_j - \lambda c_j \right) = 0, \quad j = 0, 1, \ldots, n, \tag{4.27}$$

known as *complementary slackness* conditions.

Rather than attempting to solve this system of conditions for arbitrary n, which appears to be a difficult task, we will take n to be asymptotically large and reframe the minimization above as a variational problem for a continuous function $\varphi(t)$ in the unit interval $[0, 1]$. To do so, we proceed as follows: we first note that as n goes to infinity j/n approaches a continuous real variable t. So, we define

$$0 \leq t \equiv \frac{j}{n} \leq 1, \quad j = 0, 1, \ldots n, \tag{4.28}$$

and assume $\{\xi_j\}$ and $\{c_j\}$ are a discretization of some continuous functions, $\varphi(t)$ and $\psi(t)$ respectively, so that

$$\xi_j = \frac{\varphi(t)}{\sqrt{n}}, \quad c_j = \frac{\psi(t)}{\sqrt{n}} \tag{4.29}$$

[note in passing that $\varphi(t) \geq 0$ and $\psi(t) \geq 0$]. The normalization condition for $\{\xi_j\}$ and $\{c_j\}$ holds if we impose

$$\int_0^1 dt\, \varphi^2(t) = 1, \quad \int_0^1 dt\, \psi^2(t) = 1. \tag{4.30}$$

From (4.29), we have $\xi_{j+1} - \xi_j \simeq n^{-3/2}[d\varphi(t)/dt]$, and Eq. (4.25) can be viewed as a discretized version of the functional $S[\varphi]$, defined by

$$\begin{aligned} S[\varphi] = & \frac{\varphi^2(0) + \varphi^2(1)}{2n} \\ & + \frac{1}{n^2}\int_0^1 dt \left[\frac{1}{2}\left(\frac{d\varphi}{dt}\right)^2 - \frac{\omega^2}{2}\left(\varphi^2 - 1\right) + \sigma(\varphi - \lambda\psi) \right], \end{aligned} \tag{4.31}$$

where ω is a positive constant (the properly scaled Lagrange multiplier: $\omega = nb$) and $\sigma(t)$ is a function that interpolates the set of multipliers $\{s_j\}$, i.e.,

$$s_j = n^{-5/2}\sigma(t). \tag{4.32}$$

With this, Eq. (4.26) becomes $\sigma(t) \geq 0$. Similarly, the primal feasibility conditions in (4.19) and the slackness condition (4.27) become

$$\varphi(t) - \lambda\psi(t) \quad \leq \quad 0, \tag{4.33}$$

$$\sigma(t)[\varphi(t) - \lambda\psi(t)] \quad = \quad 0. \tag{4.34}$$

Note that by imposing the boundary conditions $\varphi(0) = 0$ and $\varphi(1) = 0$, the functional $S[\varphi]$ becomes $O(n^{-2})$.

More interestingly, the minimization of $S[\varphi]$ defines a mechanical problem, of which the second line in Eq. (4.31) is the 'action' and the corresponding integrant the 'Lagrangian':

$$L = \frac{1}{2}\left(\frac{d\varphi}{dt}\right)^2 - \frac{\omega^2}{2}\varphi^2 + \sigma\varphi. \tag{4.35}$$

It describes a driven harmonic oscillator with angular frequency ω, whose 'equation of motion' is

$$\frac{d^2\varphi}{dt^2} + \omega^2\varphi = \sigma. \tag{4.36}$$

To solve this problem, we first note that the slackness conditions imply that either $\varphi(t) = \lambda\psi(t)$, in which case t is in the so called *coincidence set* $\mathscr{C}$, or $\sigma(t) = 0$. In the second case, $t \in \mathscr{C}^c$ ($\mathscr{C}^c$ stands for the complement of $\mathscr{C}$), the primal feasibility condition is $\varphi(t) < \lambda\psi(t)$, and Eq. (4.36) becomes homogeneous (the equation of motion of a free harmonic oscillator). It has the familiar solution

$$\varphi(t) = A\sin\omega t + B\cos\omega t, \tag{4.37}$$

where A, B and ω are constants to be determined. In the coincidence set $\mathscr{C}$, σ is determined by (4.36), where we make the substitution $\varphi(t) = \lambda\psi(t)$ (recall that ψ is a given function, as the components c_j are themselves given). If we restrict ourselves to fiducial states $|\Psi_0\rangle$ whose components c_j are such that $\psi(t)$, defined through Eq. (4.29), is continuous in the whole unit interval, one can show that the solution $\varphi(t)$ and its first derivative must be also continuous there [except in points of $\mathscr{C}$ where $\psi(t)$ itself is not differentiable]. Most of the physically relevant cases are of this type; some of them are considered in the examples below. By taking into account the boundary conditions, as well as the continuity of $\varphi(t)$ and its derivative in the boundaries of $\mathscr{C}$, one can determine the arbitrary constants that arise in solving the equation of motion.

Before presenting examples of this approach, we note that the minimum value of S can be expressed in terms of the Lagrange multiplier (function) ω (σ), and the given function ψ, as

$$S_{\min} = \frac{1}{n^2}\left(\frac{\omega^2}{2} - \frac{\lambda}{2}\int_0^1 dt\,\sigma\psi\right). \tag{4.38}$$

To prove this, we just have to integrate by parts (4.31) and use the equation of motion (4.36) and the boundary conditions $\varphi(0) = \varphi(1) = 0$. Note that the integral is effectively over the coincidence set $\mathscr{C}$, where the expression for $\sigma(t)$ is given by: $\sigma = \lambda(d^2\psi/dt^2 + \omega^2\psi)$, as discussed above.

4.3.1 Large abstention ($\boldsymbol{\lambda \gg 1}$)

For values of the abstention rate very close to one (large λ), and provided $c_j > 0$ for all j, the quantities λc_j are also very large and $\mathscr{C} = \emptyset$. In this case $\sigma \equiv 0$ in $[0,1]$, Eq. (4.36) becomes homogeneous and we are dealing with a regular Sturm-Liouville eigenvalue problem. The solution is

$$\varphi(t) = A\sin\omega t; \quad \omega = \pi m, \quad m = 1, 2, \ldots, \tag{4.39}$$

where the boundary conditions $\varphi(0) = \varphi(1) = 0$ have been taken into account to discard the independent $\cos\omega t$ solution. Since we must have $\varphi(t) \geq 0$ in the whole unit interval, we find that $m = 1$ (which gives the minimum eigenvalue of d^2/dt^2 for the given boundary conditions). The constant A is fixed by normalization and takes the value $A = \sqrt{2}$, thus

$$\varphi(t) = \sqrt{2} \sin \pi t, \tag{4.40}$$

namely $\xi_j \simeq \sqrt{2/n} \sin(\pi j/n)$. The minimum value of S is

$$S^* = \frac{\pi^2}{2n^2}. \tag{4.41}$$

This leads to an asymptotic maximum fidelity of

$$F^* = 1 - \frac{\pi^2}{4N^2}, \tag{4.42}$$

which coincides with the known fidelity results for optimal phase encoding Bagan *et al.* [2005]; Fiurášek [2006].

4.3.2 $|\mathbf{\Psi_0}\rangle$ proportional to the POVM seed state $|\mathbf{\Phi_1}\rangle$

The example we consider here is very simple from a computational point of view and yet illustrates that even a tiny rate of abstention can drastically improve the asymptotic fidelity F of parameter estimation. More precisely, we will show that any finite amount of abstention enables changing the shot noise limit scaling N^{-1} of $1 - F$ for large N into the Heisenberg limit scaling: N^{-2}. The elements of the family are equal superposition of all 'Fock' states $|j\rangle$, i.e. $c_j = 1/\sqrt{n+1}$. Despite of having such a large support, in the standard approach, $Q = 0$ ($\lambda = 1$), the phase estimation fidelity these states provide does not exceed the shot noise limit: $1 - F = 1/(2N + 2)$. This can be exactly computed for any N with ease from (4.23). Of course it also agrees with the analytic asymptotic results: using Eq. (4.29) we obtain $\varphi(t) = \psi(t) = 1$, for $t \in [0, 1]$, and the $1/n$ ($= 1/N$) boundary term in the action (4.38) is dominant.

Let us now address the more interesting case of $Q > 0$ ($\lambda > 1$). Here we can freely impose $\varphi(0) = \varphi(1) = 0$ and get rid of the shot-noise type term $1/n$. In a sufficiently small neighbourhood of $t = 0$, i.e., for $0 \le t < \alpha$, where α is likewise small, we have $\varphi(t) - \lambda < 0$, and the complementary slackness condition (4.34) implies $\sigma(t) = 0$ there. If α is the maximum value of t less that $1/2$ for which this condition holds, it must be a boundary point of the coincidence set $\mathscr{C}$. Then, for $t \ge \alpha$ the solution is given by the rescaled input state $\varphi(t) = \lambda\psi(t) = \lambda$. Thus,

$$\varphi(t) = \begin{cases} A \sin \omega t, & 0 \le t < \alpha; \\ \lambda, & \alpha < t \le 1/2, \end{cases} \tag{4.43}$$

where the constants α, ω and A are to be determined. Continuity of $\varphi(t)$ and its derivative at $t=\alpha$ yields

$$A\sin\omega\alpha=\lambda, \qquad A\omega\cos\omega\alpha=0. \tag{4.44}$$

We are left with the following possibilities for ω and A:

$$\omega\alpha=(2m+1)\frac{\pi}{2},\;\; A=(-1)^m\lambda;\;\; m=0,1,2,\ldots. \tag{4.45}$$

The positivity condition $\varphi(t)\geq 0$ requires $m=0$, and normalization, Eq. (4.30),

$$\alpha=1-\frac{1}{\lambda^2}=Q. \tag{4.46}$$

Note that since $\alpha\leq 1/2$ we have $Q^*=1/2$. Combining these results we obtain

$$\omega=\frac{\pi}{2Q}. \tag{4.47}$$

Extending the solution to the entire unit interval by applying the obvious symmetry of the problem, namely $\varphi(t)=\varphi(1-t)$, one has for $0<Q\leq Q^*$ $(1<\lambda\leq\sqrt{2})$

$$\varphi(t)=\begin{cases}\bar{Q}^{-\frac{1}{2}}\sin\dfrac{\pi t}{2Q}, & 0\;\leq t<\;Q;\\[2ex] \bar{Q}^{-\frac{1}{2}}, & Q\;\leq t\leq\;\bar{Q};\\[2ex] \bar{Q}^{-\frac{1}{2}}\sin\dfrac{\pi(1-t)}{2Q}, & \bar{Q}\;<t\leq\;1.\end{cases} \tag{4.48}$$

Note that $\mathscr{C}=[Q,\bar{Q}]$ and $\sigma(t)=\omega^2\lambda$ for $t\in\mathscr{C}$ $[\sigma(t)=0$ for $t\in\mathscr{C}^c]$. Therefore, Eq. (4.38) gives

$$S_{\min}=\frac{\pi^2}{8Q\bar{Q}\,n^2}, \qquad 0<Q\leq Q^*, \tag{4.49}$$

from which

$$F=1-\frac{\pi^2}{16Q\bar{Q}N^2}, \qquad 0<Q\leq Q^*=1/2. \tag{4.50}$$

For $1/2<Q\leq 1$ the solution is (4.40) and the fidelity in (4.42). Note that even the slightest abstention rate unlocks the encoding power of the phase states and drastically changes the estimation precision from the original N^{-1} to N^{-2}.

The above results are illustrated in Fig. 4.1, where we represent the optimal solution for a 17% abstention rate. Notice how the slackness conditions

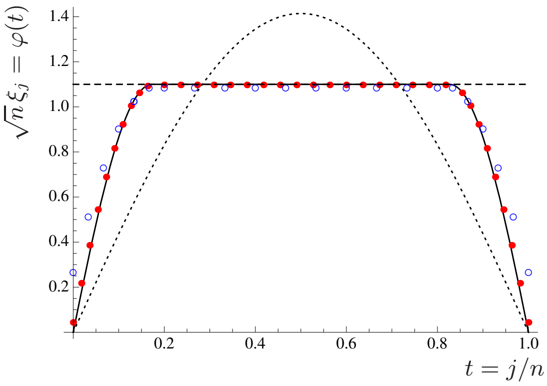


Figure 4.1. Profile of the components ξ_j of the transformed fiducial state $|\tilde{\Psi}_0\rangle$ in the asymptotic limit. The solid line is the limiting function $\varphi(t)$ for $Q = 0.17$ ($\lambda = 1.1$). The points represent the actual components, as obtained by numerical optimization for $n = 30$ (empty circles) and $n = 220$ (filled circles). The dotted line is the solution for unrestricted abstention, Eq. (4.40), whereas the dashed horizontal line represents the components $\psi(t) = 1 = \sqrt{n}c_j$ of the initial fiducial state $|\Psi_0\rangle$ scaled by λ.

apply in the different regions: the straight part of φ (corresponding to $t \in \mathscr{C}$) is just $\lambda\psi = \lambda$, while the sinusoidal curves in the extremes (corresponding to the unconstrained region $\mathscr{C}^c$) smoothly match the straight line at the boundary. The agreement between the numerical points and the analytic continuum limit is also quite evident.

4.3.3 Multiple copies on the equator

Let us now focus on phase estimation with a signal of the form

$$|\Psi(g)\rangle = \left(\frac{|0\rangle + e^{i\theta}\,|1\rangle}{\sqrt{2}} \right)^{\otimes n}, \tag{4.51}$$

that is, with $N = n$ copies of states lying on the equator of the Bloch sphere. For these the coefficients c_j read

$$c_j = 2^{-\frac{n}{2}} \sqrt{\binom{n}{j}}. \tag{4.52}$$

The maximum fidelity that can be attained with this signal without abstention is well known to be $1 - F = 1/(4N) = 1/(4n)$ for large n Bagan *et al.*

[2005]; Fiurášek [2006]. To compute the effect of abstention we proceed along the lines of the previous section. In the asymptotic limit Eq. (4.29) leads to

$$\psi(t) = \left[\frac{n}{2\pi t(1-t)}\right]^{1/4} \exp\left\{-\frac{n}{2}[\log 2 - H(t)]\right\}, \tag{4.53}$$

where $H(t) = -t\log t - (1-t)\log(1-t)$ is the Shannon entropy, and we have used Stirling's approximation. Note that $\log 2 - H(t)$ is the (binary) relative entropy $H(t\,\|\,1/2)$ between a Bernoulli distribution with success probability $p = t$ and the flat one ($p = 1/2$). As in the previous case, the problem is invariant under $t \to 1-t$, which suggest using the variable $\tau = t - 1/2$, $\tau \in [-1/2, 1/2]$, instead of t. Hence, the solution must be an even function of τ. In the region $|\tau| \lesssim n^{-1/2}$ [i.e., around the peak of the distribution (4.53)], we can use the Gaussian approximation

$$\psi(\tau) \approx \left(\frac{2n}{\pi}\right)^{1/4} \mathrm{e}^{-n\tau^2}, \tag{4.54}$$

where we slightly abuse notation here and in the rest of the section and use $\psi(\tau)$ to denote $\psi(t(\tau))$. At the tails ($|\tau| > n^{-1/2}$), $\psi(\tau)$ falls off with an exponential rate given by $H(1/2+\tau\,\|\,1/2)$.

Since the solution of the minimization must be an even function of τ, it must have the form

$$\varphi(\tau) = \begin{cases} A\cos\omega\tau, & 0 \le |\tau| \le \alpha, \\ \lambda\psi(\tau), & \alpha < |\tau| \le 1/2, \end{cases} \tag{4.55}$$

The continuity of both $\varphi(\tau)$ and $\varphi'(\tau)$ at the boundary of $\mathscr{C}$, i.e., at the point $\tau = \alpha$ read:

$$A\cos(\Omega) = \lambda\psi(\alpha), \tag{4.56}$$

$$-\Omega A\sin(\Omega) = \alpha\lambda\psi'(\alpha), \tag{4.57}$$

where we have defined $\Omega \equiv \omega\alpha$. Combining these equations we obtain

$$\Omega\tan\Omega = -\alpha\frac{\psi'(\alpha)}{\psi(\alpha)}, \tag{4.58}$$

$$A^2 = \lambda^2\left\{\psi^2(\alpha) + \frac{\alpha^2}{\Omega^2}\left[\psi'(\alpha)\right]^2\right\}. \tag{4.59}$$

The normalization condition (4.30) turns out to be

$$A^2\frac{\alpha(2\Omega+\sin 2\Omega)}{2\Omega} + 2\lambda^2\int_\alpha^{1/2}\psi^2(\tau)\,d\tau = 1. \tag{4.60}$$

Eqs. (4.56) through (4.60) cannot be solved analytically, but we can find asymptotic solutions by focusing on some specific regimes. The first we will consider arises when the boundary points $\pm\alpha$ scale as $n^{-1/2}$, so that $\mathscr{C}$ stretches to the region around the peak of $\psi(\tau)$. In this case, $\varphi(\tau)=\lambda\psi(\tau)$ gives the dominant contribution to $S_{\min}$ and, as one intuitively expects, $S_{\min}\sim n^{-1}$. The two pieces of φ in Eq. (4.55) can be matched for arbitrary values of λ and the abstention rate can be finite (is not required to scale with n). The second regime arises when α is fixed. In this situation, for sufficiently large n, the coincidence set $\mathscr{C}$ lies on the tails of $\psi(\tau)$. Matching the two pieces of φ requires that λ scales exponentially with n, which means that the acceptance rate $\bar{Q}$ must vanish also exponentially. In return, the piece of φ in the first line of Eq. (4.55) has a wide (non vanishing) domain, $[-\alpha,\alpha]$, and $S_{\min}\sim n^{-2}$ $(1-F\sim N^{-2})$, thus attaining the Heisenberg limit. Let us now consider the two regimes in more detail.

$1/n$ regime

we write $\alpha=a/\sqrt{n}$, where a is fixed. Using the Gaussian approximation in Eq. (4.54), Eqs. (4.58) through (4.60) become

$$a^2 = \frac{\Omega\tan\Omega}{2}, \tag{4.61}$$

$$A^2 = \left(\frac{2n}{\pi}\right)^{1/2}\frac{\lambda^2\mathrm{e}^{-2a^2}(4a^4+\Omega^2)}{\Omega^2}, \tag{4.62}$$

$$1 = A^2\frac{a(2\Omega+\sin 2\Omega)}{2\sqrt{n}\,\Omega}+\lambda^2\left[1-\mathrm{Erf}(\sqrt{2}\,a)\right], \tag{4.63}$$

where $\mathrm{Erf}(x)$ is the error function. Eq. (4.63) is correct up to exponentially vanishing contributions, which can be neglected here. In deriving this equation we also used that $\mathrm{Erf}(\sqrt{n/2})\to 1$ for large n. Substituting Eq. (4.62) in Eq. (4.63) we obtain

$$\frac{1}{\lambda^2}=\mathrm{Erfc}(\sqrt{2}a)+\frac{a(4a^4+\Omega^2)(2\Omega+\sin 2\Omega)}{\sqrt{2\pi}\Omega^3}\mathrm{e}^{-2a^2}, \tag{4.64}$$

where Erfc is the complementary error function, defined as $\mathrm{Erfc}(x)=1-\mathrm{Erf}(x)$. Finally, with the help of the Gaussian approximation (4.54), we compute the minimum action from Eq. (4.38) and obtain

$$S_{\min}=\frac{\omega^2}{2n^2}-\frac{\lambda^2}{2n^2}\left[(\omega^2-n)\mathrm{Erfc}(\sqrt{2}a)+\frac{4na}{\sqrt{2\pi}}\mathrm{e}^{-2a^2}\right]. \tag{4.65}$$

Eqs. (4.61) and (4.64), along with $\omega = \Omega\sqrt{n}/a$ and $Q = 1 - 1/\lambda^2$, enable writing all variables in terms of the single parameter Ω. By further substituting in Eq. (4.65) we obtain the curve $(Q, S_{\min})$ in parametric form:

$$Q = \mathrm{Erf}\left(\sqrt{\Omega \tan\Omega}\right) - \left(\Omega \sec^2\Omega + \tan\Omega\right)\sqrt{\frac{\tan\Omega}{\pi\,\Omega}}\,\mathrm{e}^{-\Omega\tan\Omega}, \tag{4.66}$$

$$S_{\min} = \frac{1}{2n}\left[1 + \frac{\tan^2\Omega - \Omega\,(2\Omega - \tan\Omega)\sec^2\Omega}{2\,\Omega^2\sec^2\Omega + \sqrt{\pi\,\Omega\tan\Omega}\;\mathrm{Erfc}\left(\sqrt{\Omega\tan\Omega}\right)\,\mathrm{e}^{\Omega\tan\Omega}}\right]^{-1}. \tag{4.67}$$

Note that, as announced above, $1 - F$ goes as $1/N$.

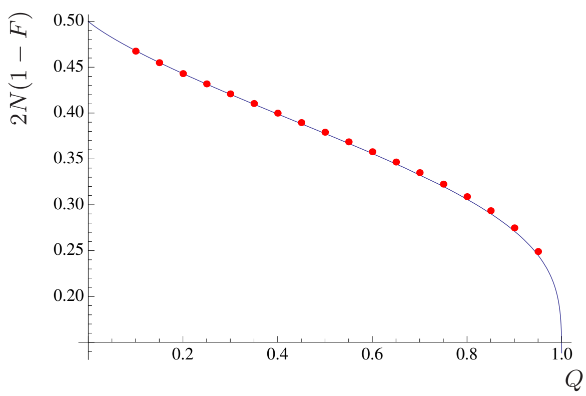


Figure 4.2. Plot of $nS_{\min} = 2N(1-F)$ vs Q (solid line) for an asymptotically large number, $N = n$, of parallel spins on the equator of the Bloch sphere. The dots have been obtained by numerical optimization with $n = 100$.

In Fig. 4.2 we plot $nS_{\min} = 2N(1-F)$ as a function of Q, using Eqs. (4.66) and (4.67). The plot shows a strong dependence on Q. Hence, e.g., allowing about 90% of abstention, has the same effect as doubling the number of copies in the standard approach (without abstention). Note also that for $Q \to 0$ we recover the well known result $2N(1-F) = 1/2$. The profile of the transformed fiducial state $|\tilde{\Psi}_0\rangle$ is shown in Fig. 4.3, where $\varphi(\tau)$ and $\lambda\psi(\tau)$ are plotted as a function of $t = j/n$ for two different values of n (recall that $\tau = t - 1/2$).

$1/n^2$ regime

Here we assume that α is fixed (does not scale with n). As n goes to infinity, the boundaries of the coincidence set, $\tau = \pm\alpha$, lie on the tails of $\psi(\tau)$, where the Gaussian approximation is not valid, and Eq. (4.53) must be used

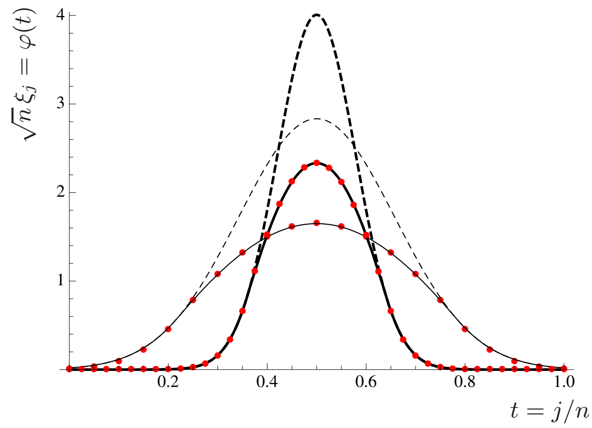


Figure 4.3. Profile of the transformed fiducial state $|\tilde{\Psi}_0\rangle$ for $Q = 0.56$ ($\lambda = 1.5$). The thin (thick) lines correspond to $n = 20$ ($n = 80$). The circles are obtained by numerical optimization. The dashed lines represent the constraint $\lambda\psi(t-1/2)$, where $\psi(\tau)$ is given in Eq, (4.54).

instead. Eqs. (4.58) and (4.59) now become

$$\Omega \tan\Omega = n\alpha \operatorname{arctanh} 2\alpha - \frac{2\alpha^2}{1-4\alpha^2}, \tag{4.68}$$

$$A^2 = n^{5/2}\sqrt{\frac{2}{\pi}}\frac{\alpha^2\lambda^2 \operatorname{arctanh}^2 2\alpha}{\Omega^2\sqrt{1-4\alpha^2}} \times \exp\{-n[\log 2 - H(\alpha+\tfrac{1}{2})]\}\,. \tag{4.69}$$

The first equation can be solved for Ω as an asymptotic series in powers of $1/n$:

$$\Omega = \frac{\pi}{2} + O(n^{-1}), \tag{4.70}$$

which implies $\omega \simeq \pi/(2\alpha)$. To evaluate the integral in Eq. (4.60), we expand the exponent $-n[\log 2 - H(1/2+\tau)]$ around $\tau = \alpha$, so that

$$\int_\alpha^{\frac{1}{2}} \psi^2(\tau)d\tau \approx \psi^2(\alpha)\int_\alpha^{\frac{1}{2}} \mathrm{e}^{-2n(\tau-\alpha)\operatorname{arctanh} 2\alpha} d\tau \approx \frac{1}{\sqrt{2\pi n}}\frac{\exp\{-n[\log 2 - H(\alpha+\frac{1}{2})]\}}{\sqrt{1-4\alpha^2}\operatorname{arctanh} 2\alpha}. \tag{4.71}$$

we note that, although this contribution falls off exponentially exactly as A^2, it can be neglected in evaluating Eq. (4.60) since its prefactor is $O(n^{-1/2})$, as compared to that of A^2, which is $O(n^{5/2})$. Taking this into account and substituting $\Omega \approx \pi/2$ and Eq. (4.69) into Eq. (4.60), we have

$$A = \frac{1}{\sqrt{\alpha}}, \tag{4.72}$$

$$\bar{Q} = \frac{1}{\lambda^2} \approx \left(\frac{2n}{\pi}\right)^{5/2} \frac{\alpha^3 \operatorname{arctanh}^2 2\alpha}{\sqrt{1-4\alpha^2}} \times \exp\{-n[\log 2 - H(\alpha+\tfrac{1}{2})]\}, \tag{4.73}$$

and the critical acceptance rate is $\bar{Q}^* = 2^{-n}$ (corresponding to $\alpha \to 1/2$).

The minimum action can be computed form Eq. (4.38) using the same approximation as in Eq. (4.71). we obtain

$$n^2 S_{\min} = \frac{\pi^2}{8\alpha^2} - \frac{n^{\frac{3}{2}}}{\sqrt{2\pi}} \frac{\lambda^2 \operatorname{arctanh} 2\alpha}{\sqrt{1-4\alpha^2}} \exp\{-n[\log 2 - H(\alpha+\tfrac{1}{2})]\}. \tag{4.74}$$

Note that the exponential factor in the second line of this equation is cancelled by λ^2, given in Eq. (4.73), and only the product of the pre-factors, of order n^{-1}, remains. Thus the second line can be safely neglected in the asymptotic limit and we have

$$F = 1 - \frac{\pi^2}{16N^2\alpha^2} + O(N^{-3}), \quad 0 < \alpha \leq 1/2, \tag{4.75}$$

with an abstention rate given by Eq. (4.73). The maximum fidelity is attained by the largest value of $\alpha = 1/2$, for which $F = F^*$ as it should be. In summary, high abstention rate (exponentially small acceptance rate) enables a drastic change in the scaling with the number of copies of the estimation precision. With such rates, one can attain $1 - F \sim 1/N^2$, i.e., achieve the Heisenberg limit.

4.4 Direction estimation

Proceeding along the same lines as in Sec. 4.3, we can write $\langle\xi|\mathsf{M}|\xi\rangle$ [recall Eq. (4.13)] as

$$\langle\xi|\mathsf{M}|\xi\rangle = \sum_{j=1}^{n} \frac{2j}{\sqrt{4j^2-1}} \xi_j \xi_{j-1}, \tag{4.76}$$

and $S = 1 - \langle\xi|\mathsf{M}|\xi\rangle$ becomes now

$$S = \frac{1}{2}\left[\sum_{j=1}^{n} j\left(\frac{\xi_j}{\sqrt{j+\frac{1}{2}}} - \frac{\xi_{j-1}}{\sqrt{j-\frac{1}{2}}}\right)^2 + \frac{(n+1)\,\xi_n^2}{n+\frac{1}{2}}\right], \tag{4.77}$$

where we have used the normalisation constraint in Eq. (4.18). Introducing Lagrange multipliers according to KKT, and assuming $N = 2n$ asymptotically large, we obtain the equivalent variational problem of minimizing the action

$$\begin{aligned} S = {} & \frac{\varphi^2(1)}{2n} \\ & + \frac{1}{n^2}\int_0^1 dt \left\{ \frac{t}{2}\left[\frac{d}{dt}\left(\frac{\varphi}{\sqrt{t}}\right)\right]^2 - \frac{\omega^2}{2}\left(\varphi^2 - 1\right) + \sigma(\varphi - \lambda\psi)\right\}, \end{aligned} \tag{4.78}$$

where the primal feasibility condition (4.33) and the slackness condition (4.34) still apply. For $\lambda = 1$ no transformation of the state is possible, therefore the first, order n^{-1}, term in (4.78) is fixed by the boundary value of the initial state $\psi(1)$. For $\lambda > 1$ we can impose $\varphi(1) = 0$, hence opening the door to order n^{-2} scaling (i.e., to attaining the Heisenberg limit).

The evolution equation corresponding to the second line in Eq. (4.78) is more conveniently expressed in terms of $\tilde{\varphi}(t) = \varphi(t)/\sqrt{t}$. It reads

$$t^2\frac{d^2\tilde{\varphi}}{dt^2} + t\frac{d\tilde{\varphi}}{dt} + \omega^2 t^2\,\tilde{\varphi} = t^{3/2}\,\sigma. \tag{4.79}$$

The minimum value of the action can be written as in Eq. (4.38), where we recall that $\sigma(t)$ can be only different from zero in the coincidence set $\mathscr{C}$. Now, $\sigma(t)$ is given by Eq. (4.79) with $\tilde{\varphi}(t) = \lambda\psi(t)/\sqrt{t}$.

4.4.1 Large abstention ($\boldsymbol{\lambda \gg 1}$)

For abstention rates close to unity, and provided $c_j > 0$ for all j, one has $\mathscr{C} = \emptyset$, so $\sigma(t) \equiv 0$. Eq. (4.79) becomes homogeneous and its solution is

$$\varphi(t) = A\sqrt{t}\,J_0(\omega t) + B\sqrt{t}\,Y_0(\omega t), \tag{4.80}$$

where J_0 and Y_0 are Bessel functions of first and second kind respectively, and A, B and ω are constants fixed by requiring $\varphi(1) = 0$ (otherwise S is order $1/n$) and the convergence of the integral in Eq. (4.78). The latter implies $B = 0$. The former condition and the positivity of $\varphi(t)$ fixes ω to be the first zero of J_0, which we call γ_1. Hence, $\omega = \gamma_1 \approx 2.405$. Imposing normalization we finally fix A, and the solution is

$$\varphi(t) = \frac{\sqrt{2t}}{J_1(\gamma_1)} J_0(\gamma_1 t). \tag{4.81}$$

Using Eq. (4.38), we obtain $S^* = \gamma_1^2/2n^2$, and the maximum fidelity is

$$F^* = 1 - \frac{\gamma_1^2}{N^2}, \tag{4.82}$$

in agreement with Bagan *et al.* [2000]. The abstention rate required to achieve the Heisenberg limit strongly depends on the initial family of states, as will be shown in the following two examples.

4.4.2 $|\boldsymbol{\Psi_0}\rangle$ proportional to the POVM seed state $|\boldsymbol{\Phi_2}\rangle$

In analogy with Sec. 4.3.2, in this example we choose the fiducial state $|\Psi_0\rangle$ to be proportional to the POVM seed $|\Phi_2\rangle$ in Eq. (4.12). This leads to $\psi(t) = \sqrt{2t}$, and the solution has the form

$$\varphi(t) = \begin{cases} \lambda\sqrt{2t}, & 0 \le t \le \alpha, \\ A\sqrt{t}\, J_0(\omega\, t) + B\sqrt{t}\, Y_0(\omega\, t), & \alpha < t \le 1. \end{cases} \tag{4.83}$$

Then, $\sigma(t) = \lambda\omega^2\sqrt{2t}$, if $t \in \mathscr{C} = [0, \alpha]$ (and it vanishes otherwise). Substituting in Eq. (4.38), the minimum action can be written as

$$S_{\min} = \frac{\omega^2}{2n^2}(1 - \alpha^2\lambda^2). \tag{4.84}$$

Continuity of $\varphi(t)$ and its first derivative at $t = \alpha$, imply

$$A = -\frac{\pi\alpha\lambda\omega}{\sqrt{2}} Y_1(\omega\alpha), \quad B = \frac{\pi\alpha\lambda\omega}{\sqrt{2}} J_1(\omega\alpha), \tag{4.85}$$

and the boundary condition $\varphi(1) = 0$ requires,

$$J_1(\omega\alpha) Y_0(\omega) - Y_1(\omega\alpha) J_0(\omega) = 0. \tag{4.86}$$

we will not attempt to find the exact analytical solution of this transcendental equation, but rather, consider two particular regions of α (the boundary of the coincidence set $\mathscr{C}$) where approximate solutions can be easily derived. They are given by $\alpha \gtrsim 0$ and $\alpha \lesssim 1$. That will suffice to capture the main features of $S_{\min}$ (see Figure 4.4). Note that small α corresponds to large λ, since the coincidence set $\mathscr{C} = [0,\alpha]$ is a small region and thus $\varphi(t)$ cannot differ much from the unconstrained solution that leads to F^*. On the other hand, $\alpha \lesssim 1$ must correspond to small abstention.

If $\alpha \gtrsim 0$, we substitute the ansatz $\omega = \gamma_1 + a\alpha + b\alpha^2 + \ldots$ in (4.86). After some algebra, we obtain

$$\omega = \gamma_1 \left[1 + \frac{\alpha^2}{2J_1^2(\gamma_1)} + O(\alpha^4 \log \alpha)\right], \tag{4.87}$$

where we have made use of the relation

$$J_1(z)Y_0(z) - Y_1(z)J_0(z) = \frac{2}{\pi z}, \quad \text{for all } z; \tag{4.88}$$

in particular, $Y_0(\gamma_1) = 2J_1^{-1}(\gamma_1)/(\pi\gamma_1)$.

If $\alpha \lesssim 1$, Eq. (4.86) can only hold for very large ω and $\alpha\omega \approx \omega$, as is apparent from Eq. (4.88), and we can replace the Bessel functions for their well known asymptotic approximations

$$\begin{aligned} J_k(z) &\approx \sqrt{\frac{2}{\pi z}} \cos\left(z - \frac{k\pi}{2} - \frac{\pi}{4}\right), \\ Y_k(z) &\approx \sqrt{\frac{2}{\pi z}} \sin\left(z - \frac{k\pi}{2} - \frac{\pi}{4}\right). \end{aligned} \tag{4.89}$$

With this, Eq. (4.86) becomes

$$\frac{2\cos\omega(1-\alpha)}{\pi\omega\sqrt{\alpha}} = 0, \tag{4.90}$$

from which

$$\omega = \frac{\pi}{2(1-\alpha)}. \tag{4.91}$$

we next impose the normalisation condition to find the relationship between λ and α. For $\alpha \gtrsim 0$, we find

$$\lambda^2 = \frac{1}{J_1^2(\gamma_1)} + O(\alpha^2 \log \alpha). \tag{4.92}$$

Taking the limit $\alpha \to 0$ we find the critical value of λ: $\lambda^* = 1/J_1(\gamma_1)$; and the critical rate of abstention:

$$Q^* = 1 - J_1^2(\gamma_1) \approx 0.73. \tag{4.93}$$

Substituting Eq. (4.87) and (4.92) in Eq. (4.84) we readily see that the various contributions to order α^2 cancel, and

$$S_{\min} = \frac{1}{n^2}\left[\frac{\gamma_1^2}{2} + O(\alpha^4 \log\alpha)\right]. \tag{4.94}$$

One can check that, as expected, $S_{\min}$ (and thus the fidelity) is flat in the region $\alpha \gtrsim 0$ ($Q \lesssim Q^*$); i.e., $S_{\min}$ is a smooth function of Q at $Q = Q^*$. Indeed, Eq. (4.92) implies $\alpha^2 = o(\lambda^{*2} - \lambda^2) = o(Q^* - Q)$, and

$$n^2 S_{\min} = \gamma_1^2/2 + o[(Q^* - Q)^2 \log(Q^* - Q)] \tag{4.95}$$

The correction can be computed explicitly with some effort. we find that $S_{\min}$ increases up to 3.5% for $Q \approx 0.6$, at which point the approximation breaks down.

For $\alpha \lesssim 1$, we find

$$\lambda^2 \approx \frac{1}{\alpha}. \tag{4.96}$$

Combining all these results, we find

$$n^2 S_{\min} \approx \begin{cases} \dfrac{\pi^2}{8Q}, & Q \gtrsim 0 \\ \dfrac{\gamma_1^2}{2}, & Q \lesssim Q^*, \end{cases} \tag{4.97}$$

where we insist that this expression is a very good approximation down to relatively small values of Q, as can be seen in Fig.4.4. In this figure we plot Eq. (4.97) for each regime (lines), along with some numerical results (points). The plot shows a very good agreement for most of the values of the abstention rate Q. One can see that the flat region extends to values of Q fairly smaller than Q^*. Note again, that any nonzero amount of abstention enables the estimation accuracy to change behaviour from $1/N$ to $1/N^2$, thus attaining the Heisenberg limit.

4.4.3 Antiparallel spins

As for the case of phase estimation, here we focus on signals consisting in product states of $N = 2n$ spins. The simplest possibility is, of course, identical copies. However, this case is of no relevance to direction estimation with abstention, since the seed state $|\Psi_0\rangle$ has only a single component in

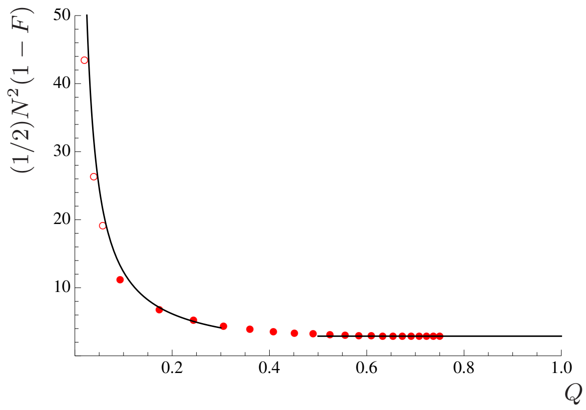


Figure 4.4. Plot of $n^2 S_{\min}[= (1/2)N^2(1-F)]$ versus Q. The solid lines are the analytical expressions in (4.97), whereas the circles are numerical results. In order to approach the asymptotic limit, higher values of n are needed for smaller Q. Accordingly, two different values of n have been used; $n = 50$ (filled circles) and $n = 120$ (empty circles).

the symmetric subspace of $j = n$, i.e., $c_j = 0$, if $0 \leq j < n$, and abstention can only change the components by a multiplicative factor, as shown in Eq. (4.10). Thus $\xi_j = 0$, if $0 \leq j < n$ and $\xi_n = c_n$. Instead, we consider a seed state consisting of $2n$ antiparallel spins; n of them pointing along the positive z-axis and the other n pointing along the opposite direction,

$$|\Psi_0\rangle = \Bigg|\overbrace{\uparrow\uparrow\ldots\uparrow}^{n}\overbrace{\downarrow\downarrow\ldots\downarrow}^{n}\Bigg\rangle = \sum_{j=0}^{n} c_j\,|j,0\rangle\,. \tag{4.98}$$

Such state has zero magnetic number, $m = 0$ and non-vanishing components c_j given by

$$c_j = \langle \tfrac{n}{2}, \tfrac{n}{2}; \tfrac{n}{2}, -\tfrac{n}{2}|j,0\rangle = n!\sqrt{\frac{2j+1}{(n-j)!(n+j+1)!}}, \tag{4.99}$$

where $\langle j,m;j',m'|J,M\rangle$ are the standard Clebsch-Gordan coefficients. The 'continuous version' of these components is given by ($t = j/n$)

$$\psi(t)=\sqrt{\frac{2nt}{(1+t)\sqrt{1-t^2}}}\,\exp\left\{-n\left[\log 2-H\left(\frac{1-t}{2}\right)\right]\right\}, \tag{4.100}$$

which has a peak at $t = 0$. The solution to the minimisation problem in Eq. (4.78) has the form

$$\varphi(t) = \begin{cases} A\sqrt{t}\,J_0(\omega t), & 0 \leq t \leq \alpha; \\ \lambda\,\psi(t), & \alpha < t \leq 1. \end{cases} \tag{4.101}$$

Following the same lines as in Sec. 4.3.3, we consider two scalings of the boundary point $t = \alpha$: one where it goes to zero as $1/\sqrt{n}$, and a second one, where α is fixed. These will lead to two regimes, where $1 - F$ vanishes respectively as N^{-1} and N^{-2}.

1/n regime

In this regime we set $\alpha = a/\sqrt{n}$. As in the phase case, we can use the 'Gaussian approximation' for (4.100):

$$\psi(t) = \sqrt{2\,n\,t}\,\mathrm{e}^{-n\,t^2/2}. \tag{4.102}$$

Note that $\int_0^1 \psi^2(t)\,dt = 1$, up to contributions that vanish exponentially with n. The following expressions follow from the conditions of continuity of the solution and its derivative as well as normalisation:

$$a^2 = \Omega\,\frac{J_1(\Omega)}{J_0(\Omega)}, \tag{4.103}$$

$$A = \frac{\sqrt{2n}\,\lambda\,\mathrm{e}^{-a^2/2}}{J_0(\Omega)}, \tag{4.104}$$

$$\bar{Q} = \frac{1}{\lambda^2} = \left(1 + a^2 + \frac{a^6}{\Omega^2}\right)\mathrm{e}^{-a^2}, \tag{4.105}$$

where we have defined $\Omega \equiv \omega\alpha = \omega a/\sqrt{n}$. The minimum action $S_{\min}$ is given by

$$S_{\min} = \frac{\Omega^2}{2n}\,\frac{1 - a^2 + a^4 + \Omega^2}{a^6 + (1 + a^2)\Omega^2}, \tag{4.106}$$

where we have neglected exponentially vanishing terms. This expression, together with Eqs. (4.103) and (4.105) defines the curve $(Q, S_{\min})$ in terms of the free parameter $\Omega \in [0, \gamma_1)$. The corresponding plot is shown in Fig. 4.5. We see that for moderate values of the abstention rate one can substantially improve the estimation precision. E.g., a rate of abstention of 95% has the same effect as doubling the number of spins in the standard approach (without abstention). Note, however, that with finite acceptance rate we cannot beat the shot noise limit.

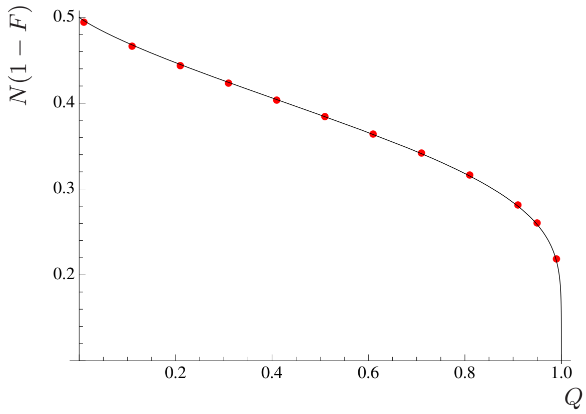


Figure 4.5. Plot of $nS_{\min}[= N(1-F)]$ vs. Q (solid line) for a signal state consisting of an asymptotically large number, $N=2n$, of antiparallel spins with null total magnetic number. The dots have been obtained by numerical optimization with $n=100$.

$1/n^2$ regime

Here we take α to be fixed. From Eq. (4.101), continuity of $\varphi(t)$ and $\varphi'(t)$ at $t=\alpha$ yield

$$A\sqrt{\alpha}\, J_0(\Omega) = \lambda\, \psi(\alpha), \tag{4.107}$$

$$\frac{A}{2\sqrt{\alpha}}\left[J_0(\Omega) - 2\Omega J_1(\Omega)\right] = \lambda\, \psi'(\alpha), \tag{4.108}$$

where, as before, $\Omega = \alpha\omega$. It follows that

$$\frac{J_0(\Omega)-2\Omega J_1(\Omega)}{2\alpha\, J_0(\Omega)} = \frac{\psi'(\alpha)}{\psi(\alpha)} = -n\ \mathrm{arctanh}\,\alpha + O(n^0), \tag{4.109}$$

where Eq. (4.100) has been used. This equation can be solved for Ω as a series in inverse powers of n, obtaining

$$\Omega = \gamma_1 + O(n^{-1}), \tag{4.110}$$

where we recall that γ_1 stands for the first zero of the function $J_0(z)$. Substituting this result into Eq. (4.108) we obtain

$$A = \frac{\sqrt{2}\, n^{\frac{3}{2}} \alpha\lambda\ \mathrm{arctanh}\,\alpha}{(1-\alpha)^{\frac{1}{4}}(1+\alpha)^{\frac{3}{4}}\gamma_1 J_1(\gamma_1)} \tag{4.111}$$

$$\times\ \exp\left\{-n\left[\log 2 - H\left(\frac{1-\alpha}{2}\right)\right]\right\}. \tag{4.112}$$

Neglecting the contribution from the coincidence set, by the same arguments as in the paragraph after Eq. (4.71), the normalization condition is

$$A = \frac{\sqrt{2}}{\alpha J_1(\gamma_1)}. \tag{4.113}$$

Combining the last two equations, we find

$$\bar{Q} = \frac{1}{\lambda^2} \sim n^3 \exp\left\{-2n\left[\log 2 - H\left(\frac{1-\alpha}{2}\right)\right]\right\}. \tag{4.114}$$

As for phase estimation, the acceptance rate $\bar{Q}$ falls off exponentially. The minimum action $S_{\min}$ can be computed from Eq. (4.38) along the same lines as in the analogous phase estimation example. This leads to

$$F = 1 - \frac{\gamma_1^2}{N^2\alpha^2}. \tag{4.115}$$

As in Sec. 4.3.3, abstention enables exceeding the shot noise limit. Note that for $\alpha = 1$, we have $F = F^*$, Eq. (4.82), as expected.

4.5 Frame estimation

As anticipated in the introduction, if the encoding system consists of N qubits one can make use of the multiplicities of the different irreducible representations (i.e. the degeneracy of the j quantum number, also presented in Section 2.2.4) to provide a very efficient encoding of the orientation of a Cartesian frame, or equivalently, of the rotation group parameters g. States of the form $|\Psi_0\rangle = \sum_{j=0}^{n} c_j \left(\sum_{m=-j}^{j} |j, m, \alpha_m\rangle\right)/\sqrt{2j+1}$ exploit optimally these ancillary degrees of freedom and lead to a matrix M that is (almost) equal to that corresponding to phase estimation. Hence, most of the expressions and conclusions derived in Section 4.3 also hold in this case, but one must recall that $N = 2n$ for frames (whereas $N = n$ for phases), i.e. one must perform the change $N \to N/2$ in the formulae of that section to obtain the corresponding formulae for frames. In particular, Eq. (4.42) becomes $F^* = 1 - \pi^2/N^2$ for frame estimation, in agreement with Bagan *et al.* [2004]; Eq. (4.75) becomes $F = 1 - \pi^2/(4N^2\alpha^2) + \ldots$, and so on. Note in particular that direction estimation does not provide an optimal strategy for frame estimation, namely, the optimal frame fidelity cannot be attained by splitting the N qubits in three groups, encoding each orthogonal direction in one of them, and performing three independent direction estimations.

In my final example we move away from the N-qubit encoding towards a scenario where the degeneracy of the angular momentum representations cannot be used to improve the frame estimation accuracy, as is the case of, e.g., an atom in a Rydberg state. In this scenario, we have

$$\begin{aligned}\langle\xi|\mathsf{M}|\xi\rangle &= 1-\frac{1}{2}\sum_{j=1}^{n}\left(\frac{1}{j+1}+\frac{1}{j+\frac{1}{2}}\right)\xi_j^2-\xi_0^2-\frac{\xi_n^2}{2} \\ &- \frac{1}{2}\sum_{j=0}^{n-1}\left(j+\frac{1}{2}\right)\left(\frac{\xi_{j+1}}{\sqrt{j+1+\frac{1}{2}}}-\frac{\xi_j}{\sqrt{j+\frac{1}{2}}}\right)^2 .\end{aligned} \tag{4.116}$$

In the asymptotic limit, the continuous version of this expression is cast as $\langle\xi|\mathsf{M}|\xi\rangle = 1 - S$, with the action

$$\begin{aligned}S &= \frac{1}{n^2}\int_0^1 dt\left\{\frac{t}{2}\left[\frac{d}{dt}\left(\frac{\varphi}{\sqrt{t}}\right)\right]^2\right. \\ &+ \left. 2n\frac{\varphi^2}{2t}-\frac{\omega^2}{2}\left(\varphi^2-1\right)+\sigma(\varphi-\lambda\psi)\right\},\end{aligned} \tag{4.117}$$

which includes the constraints (4.18) and (4.19) and the corresponding Lagrange multipliers ω and σ, and where we have set $\varphi(0) = \varphi(1) = 0$. This action and that for direction estimation, Eq. (4.78), look much the same but for the term proportional to n. This apparently minor difference leads however to very different asymptotic behaviors. The equation of motion that follows from (4.117) turns out to be

$$\frac{d^2}{dt^2}\varphi + \left(\omega^2 - \frac{2n}{t} + \frac{1}{4t^2}\right)\varphi = \sigma. \tag{4.118}$$

Since n is assumed to be asymptotically large, the term proportional to n in (4.117) forces $\varphi(t)$ to peak at $t \approx 1$ in order to minimize the action. Therefore, the last term in (4.118) can be safely neglected. The minimum value of S can be written in terms of the Lagrange multipliers and $\psi(t)$ as in Eq. (4.38), with $\sigma = \lambda[\psi'' + (\omega^2\psi - 2n/t)\psi]$ for $t \in \mathscr{C}$, and $\sigma = 0$ otherwise.

4.5.1 Large abstention

Once again, for abstention rates close to one, and provided $c_j \neq 0$ for all j, Eq. (4.118) becomes homogeneous, i.e., $\sigma = 0$, and, along with the boundary conditions $\varphi(0) = \varphi(1) = 0$, defines an eigenvalue problem. Its solution can be given in terms of Whittaker functions, but unfortunately is rather involved. It proved much simpler to formulate and solve a less demanding eigenvalue problem with the same large n asymptotic behavior, as we explain next.

Since $\varphi(t)$ is peaked at $t \approx 1$, we can Taylor expand the term $2n/t$ in Eq. (4.118) around this point. The leading and sub-leading contributions to $S_{\min}$ come from the first two terms in this expansion. That is, from the linear approximation: $2n/t \approx 2n + 2n(1-t)$. Within this approximation the equation of motion becomes

$$\frac{d^2}{dt^2}\varphi + 2nt\varphi + (\omega^2 - 4n)\varphi = 0, \qquad \varphi(1) = 0, \tag{4.119}$$

and we relax the boundary condition $\varphi(0) = 0$ by requiring only $\varphi(t)$ to vanish as $t \to -\infty$. This may seem unnatural at first, but it will become immediately apparent that the solution to this well-posed Sturm-Liouville eigenvalue problem vanishes exponentially with n if $t \leq 0$ (in particular $\varphi(0) \to 0$ exponentially as $n \to \infty$), which is enough to ensure that the resulting asymptotic expansion of $S_{\min}$ in inverse powers of n will be correct. Such solution is:

$$\varphi(t) = C \, \mathrm{Ai}\left[\frac{4n - \omega^2 - 2nt}{(2n)^{2/3}}\right], \tag{4.120}$$

where Ai is the Airy function and the constant C is fixed by normalization. Imposing the second boundary condition, $\varphi(1) = 0$, we have (for the smallest eigenvalue) $\omega^2 = 2n - \gamma_1 (2n)^{2/3}$, where in this section γ_1 stands for the first zero of $\mathrm{Ai}(x)$, whose value is $\gamma_1 \approx -2.33811$. Using (4.38), we obtain the minimum action

$$S^* = \frac{1}{n} - \frac{\gamma_1}{2^{1/3} n^{4/3}} + O(n^{-5/2}), \tag{4.121}$$

from which (recall that here $N = 2n$)

$$F^* = 1 - \frac{1}{N} + \frac{\gamma_1}{N^{4/3}} + O(N^{-5/3}). \tag{4.122}$$

For the average of the error e_3 with which we estimate the three axes of the Cartesian frame, we obtain $\langle e_3 \rangle = 8/N - 8\gamma_1/N^{4/3} + O(N^{-5/3})$. These results are in complete agreement with those in Bagan *et al.* [2001].

The asymptotic series we have obtained turns out to be in powers of $N^{-1/3}$. To obtain accurate values of F^* for moderately large N, the next term in (4.122), of order $N^{-5/3}$, might be important. Using the presented approach the calculation of this term is straightforward. One simply needs to include in (4.119) the next term in the Taylor expansion of $2n/t$, i.e., $2n(1-t)^2$, and use perturbation theory to obtain the correction $\delta\omega^2 = 2n \int_{-\infty}^{1} (1-t)^2 \varphi^2(t) \, dt$. The corresponding correction to S^* can then be computed via Eq. (4.38). The result is $\delta S^* = 2^{7/3}\gamma_1^2/(15 n^{5/3})$. From this, the correction to the fidelity turns out to be $\delta F^* = 8\gamma_1^2/(15 N^{5/3})$.

4.5.2 Limited Abstention

As in the previous examples, if the rate of abstention is fixed to a value strictly less than one the resulting precision very much depends on the given signal state, namely, on the shape of c_j (or ψ). In order to give a concrete expression for the fidelity, here we will assume that, maybe because of some energy limitations, the probability amplitudes c_j of exciting a state (e.g., of a Rydberg atom) with angular momentum j is a decreasing function. Let us further assume as a first approximation, and also for simplicity, that this decrease is linear: $c_j \propto n - j$, which implies $\psi(t) = \sqrt{3}(1-t)$. This simple example will allow us to illustrate the most characteristic features of frame estimation enhanced by abstention.

If no abstention is allowed (standard estimation), one can show that the averaged error [i.e., $8(1-F)$] vanishes as $(1/N)\log N$ as N increases, much slower than using the optimal signal states. we will show that even a tiny amount of abstention is enough to turn this scaling into $1/N$. Moreover, the coefficient in this scaling law can be reduced down to almost the minimum value in (4.122) with a finite amount of abstention.

For $0 < Q < 1$ ($\lambda > 1$) and large n, the very same argument used for large abstention shows that $\varphi(t)$ will be peaked away from $t = 0$, at some value close to the boundary of the coincidence set. we can thus Taylor expand the term $2n/t$ in (4.118) around $t = \alpha$ to sub-leading order. The differential equation becomes

$$\frac{d^2}{dt^2}\varphi + \frac{2n}{\alpha^2} t\,\varphi + \left(\omega^2 - \frac{4n}{\alpha}\right)\varphi = \sigma, \tag{4.123}$$

whose solution in $\mathscr{C}^c$ (where $\sigma = 0$) is

$$\varphi(t) = C\ \mathrm{Ai}\left[\frac{4\alpha n - \alpha^2\omega^2 - 2nt}{(2\alpha n)^{2/3}}\right]. \tag{4.124}$$

Here, we have used the weaker boundary condition $\lim_{t\to-\infty}\varphi(t) = 0$, and C is determined in terms of the remaining free parameters α and ω by imposing continuity at the boundary of the coincidence set: $\varphi(\alpha) = \lambda\psi(\alpha)$. This combined with continuity of the first derivative implies $\varphi(\alpha)/\varphi'(\alpha) = \psi(\alpha)/\psi'(\alpha)$, thus

$$\frac{\alpha^{2/3}\mathrm{Ai}\left[\frac{\alpha(2n-\alpha\omega^2)}{(2\alpha n)^{2/3}}\right]}{(2n)^{1/3}\mathrm{Ai}'\left[\frac{\alpha(2n-\alpha\omega^2)}{(2\alpha n)^{2/3}}\right]} = 1 - \alpha. \tag{4.125}$$

By inspection, we see that in order for this expression to make sense for asymptotically large n, the Lagrange multiplier ω must be of the form

$$\omega^2 = \frac{2n}{\alpha} - (2\alpha n)^{2/3}\frac{\gamma_1'}{\alpha^2} + \epsilon(n)\,n^{1/3} \equiv \omega_0^2 + O(n^{1/3}), \tag{4.126}$$

with $\epsilon(n) = o(n^0)$ and γ_1' being the first zero of the Ai$'$ function ($\gamma_1' \approx -1.0188$). To compute $\epsilon(n)$, we assume it has an asymptotic series expansion in inverse powers of $n^{1/3}$ and plug it into (4.125). we then obtain the coefficients of the resulting series recursively. At leading order we have $\epsilon(n) = -(2\alpha)^{1/3}/[\alpha(1-\alpha)\gamma_1']$. There is however an additional order $n^{1/3}$ contribution to ω^2 coming from the next (quadratic) order in the Taylor expansion of $2n/t$ in (4.118). It can be computed using perturbation theory. Namely, as $\delta\omega^2 = (2n/\alpha^3)\int_{-\infty}^{\alpha} dt\,(\alpha-t)^2\varphi^2$ [in this expression φ is assumed to be normalized to one in $(-\infty,\alpha)$]. Combining the two order $n^{1/3}$ contributions one has

$$\omega^2 = \omega_0^2 + \frac{(8\gamma_1'^3 - 3) - 4\alpha(2\gamma_1'^3 + 3)}{15\alpha^2(1-\alpha)\gamma_1'}(2\alpha n)^{1/3} + O(n^0), \tag{4.127}$$

where ω_0^2 is defined in (4.126). This equation gives ω^2 as an explicit function of α.

The rate of abstention (equivalently, λ) can also be expressed as a function of α by imposing normalization to the solution of (4.123) in the whole interval $(-\infty, 1]$, i.e., $\int_{-\infty}^{1} dt\,\varphi^2(t) = 1$. One has

$$\begin{aligned}\bar{Q} = \frac{1}{\lambda^2} = {} & (1-\alpha)^3 \\ & -3\alpha(1-\alpha)^2\left[\left(1 - \frac{\alpha\omega^2}{2n}\right) - \frac{\alpha}{2n(1-\alpha)^2}\right].\end{aligned} \tag{4.128}$$

Using (4.38) once again, we obtain

$$\begin{aligned} S_{\min} &= -\frac{3\lambda^2}{n}\left(\log\alpha + 2 - 2\alpha - \frac{1-\alpha^2}{2}\right) \\ &+ \frac{\omega^2}{2n^2}\left[1 - \lambda^2(1-\alpha)^3\right].\end{aligned} \tag{4.129}$$

Eqs. (4.128) and (4.129), define the curve $(Q, S_{\min})$ in terms of α, which we view as a free parameter that takes values in the range $0 < \alpha < 1$. This curve, which is accurate up to order $n^{-4/3}$, is plotted in Fig. 4.6 (dashed line) for $n = 20$. In the same figure, we also plot the asymptotic (leading) contribution alone (solid line) and some numerical optimization results for $n = 20$ (empty blue circles) and $n = 90$ (filled red circles). we see that $n = 20$ is still not quite in the asymptotic regime, and that the sub-leading corrections play a significant role, improving the agreement to almost perfect for central values of Q.

At leading order, $S_{\min}$ can be easily written as an explicitly function of Q, since only the leading term in (4.128) contributes and we have $\alpha = 1 - \bar{Q}^{1/3}$.

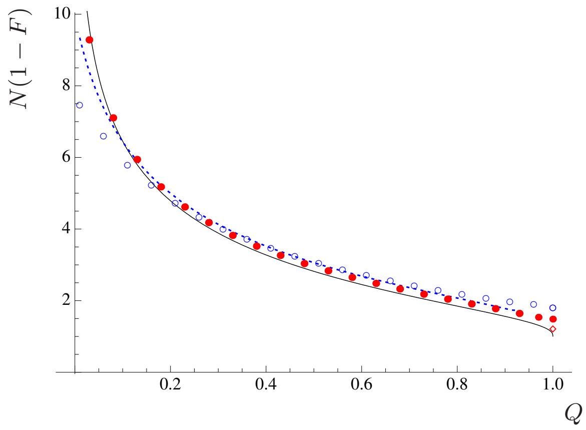


Figure 4.6. Plot of $nS_{\min}[= N(1-F)]$ vs. Q. The solid black line is the leading asymptotic expression in Eq. (4.130). The dashed line is the curve $(Q, nS_{\min})$ given by Eqs. (4.128) and (4.129) for $n = 20$. The blue empty (red filled) circles are numerical results for $n = 20$ ($n = 90$). The empty diamond is also a numerical result for $Q = 1$ and $n = 1000$.

Substituting this in the first line of (4.129), we obtain

$$S_{\min} = -\frac{3}{2n\bar{Q}}\left[2\log(1-\bar{Q}^{1/3}) + 2\bar{Q}^{1/3} + \bar{Q}^{2/3}\right]. \tag{4.130}$$

Interestingly, the corrections to this result can be shown to be of order $n^{-5/3}$, whereas the implicit form given by (4.128) and (4.129) has non-zero contributions of order $n^{-4/3}$. In the limit $Q \to 1$, Eq. (4.130) yields the leading order in (4.121), but the slope of $S_{\min}(Q)$ becomes vertical at $Q = 1$ (see solid line in Fig. 4.6). At this point our asymptotic approximation breaks down —as can be seen by noticing that the higher order terms, e.g., Eq. (4.127), diverge as some negative power of $1 - \alpha \approx \bar{Q}^{1/3}$— and the numerical results approach the leading asymptotic curve very slowly. This is apparent from Fig. 4.6, where an extra point (empty diamond), corresponding to a numerical result for $n = 1000$, has been added to further emphasize this behavior.

At the other end, for $Q \to 0$, Eq. (4.130) diverges. That should not come as a surprise, since, as mentioned above, for zero abstention the error scales as $(1/n)\log n$. This also explains why the agreement with the numerical results (circles) in Fig. 4.6 worsens as Q becomes very small.

4.6 Conclusions

We have studied the effect of abstention, or post-selection, in parameter estimation with pure states. We have seen that the use of probabilistic protocols can enhance the precision of the estimation, even turning the SQL to the Heisenberg limit in the scaling of the system size. The problem can be rephrased as a two step process in which a probabilistic filter transforms the state into a new one that yields a higher estimation fidelity. However, if the initial state is already optimal, the abstention protocol has no effect and no enhancement is possible. In this sense, probabilistic protocols with pure states helps the experimentalist if she has some limitations in the initial preparation of the system.

In the following two chapters we will address the estimation protocols in the presence of noise in the system or in the measurements. We focus on direction and phase estimation. In the latter we will see that the filter has a more effective action than compensate the initial limitations and can reduce the effect of such noise.

CHAPTER 5

Probabilistic direction estimation of mixed states

We have seen that probabilistic strategies for parameter estimation without noise can compensate the limitations of the experimentalist on the state preparation step. In this Chapter we analyse probabilistic protocols of direction estimation in the presence of noise. In particular we study the exemplary case of multiple copies of a mixed state. In this case the global state system can be expressed in a block-diagonal form by writing it in an angular momentum-like basis, as presented in Section 2.2.4. This decomposition is one of the cornerstones to finding the optimal probabilistic measurement.

The Chapter is organised as follows. In the next section we consider estimation without abstention. More precisely, we obtain the protocol that gives the best estimate of the state of a qubit based upon non-ideal measurements on N independent and identically prepared systems. In Section 5.2, estimation with abstention is introduced, and the optimal protocol for a fixed value of the abstention rate Q is obtained. We study the asymptotic regime of large N and derive the corresponding maximum fidelity and probability of abstention. As an example, we also consider an scenario where abstention gives a drastic improvement. This is the case when noise increases with N in such a way that the fidelity of the estimation approaches a finite value less than one as N becomes large.

5.1 No abstention

Let us consider N copies of a completely unknown pure qubit state $|\boldsymbol{n}\rangle$ (throughout the paper $\boldsymbol{n}$ will denote a unit Bloch vector) that we wish to estimate by performing a realistic, and therefore noisy, quantum measurement. We model it as an ideal measurement preceded by the single-qubit depolarizing channel acting on every copy:

$$\mathcal{E}(\rho) = (1-\eta)\rho + \frac{\eta}{3}(\sigma_x\rho\sigma_x + \sigma_y\rho\sigma_y + \sigma_z\rho\sigma_z), \tag{5.1}$$

where with probability $1-\eta$ no error occurs, while with probability η the state is affected by either a bit-flip, a phase-flip, or both. This error probability η is assumed to be known by the experimentalist, therefore, for the purpose of analyzing the effects of noise in the estimation process, we will transfer its effect to the input states and optimize the estimation protocol over ideal measurements. Hence, we will consider input states of the form

$$\rho(\boldsymbol{n}) = r\,|\boldsymbol{n}\rangle\langle\boldsymbol{n}| + (1-r)\frac{\mathbb{1}}{2} = \frac{\mathbb{1} + r\,\boldsymbol{n}\cdot\boldsymbol{\sigma}}{2}, \tag{5.2}$$

with $r = 1-(4/3)\eta$. In words, we will assume that the input states either do not change with probability r or they become completely randomized with probability $1-r = (4/3)\eta$. The original problem is thus equivalent to the estimation of a pure state $|\boldsymbol{n}\rangle$ (or of a uniformly distributed Boch vector $\boldsymbol{n}$) based upon the outcomes of an appropriate ideal measurement on N copies of the mixed state $\rho(\boldsymbol{n})$ in Eq. (5.2), i.e., on the state $\rho(\boldsymbol{n})^{\otimes N} = \tau(\boldsymbol{n})$.

For each measurement outcome χ an estimate $|\boldsymbol{n}_\chi\rangle$ is provided according to some guessing rule $\chi \to |\boldsymbol{n}_\chi\rangle$. We choose to quantify the quality of the estimate by means of the squared overlap

$$f(\boldsymbol{n}, \boldsymbol{n}_\chi) = |\,\langle\boldsymbol{n}|\,\boldsymbol{n}_\chi\rangle|^2, \tag{5.3}$$

also known as the fidelity. The overall quality of the estimation protocol is then given by the average fidelity

$$F = \sum_\chi \int dn f(\boldsymbol{n}, \boldsymbol{n}_\chi)\, p(\chi|\boldsymbol{n}), \tag{5.4}$$

where $dn = \sin\theta d\theta\, d\phi/(4\pi)$ is the uniform probability distribution on the two-sphere and $p(\chi|\boldsymbol{n})$ is the conditional probability of obtaining the outcome χ if the input state is $\tau(\boldsymbol{n})$. This probability is given by the Born rule $p(\chi|\boldsymbol{n}) = \mathrm{tr}\,[\Pi_\chi\tau(\boldsymbol{n})]$, where $\Pi_\chi \geq 0$ are the elements of a POVM characterising the measurement. Recall that they satisfy the completeness relation

$\sum_\chi \Pi_\chi = \mathbb{1}$, where $\mathbb{1}$ denotes the identity operator in the space spanned by the input states $\{\tau(\boldsymbol{n})\}$. The index χ may be discrete, continuous or both.

For pure states, $r = 1$, the maximum fidelity is well-known Massar and Popescu [1995]:

$$F = \frac{N+1}{N+2} = 1 - \frac{1}{N} + \mathscr{O}(N^{-2}). \tag{5.5}$$

It is also known that the (continuous) covariant POVM

$$\Pi(\boldsymbol{s}) = (2J+1)U(\boldsymbol{s})\,|J\ J\rangle\!\langle J\ J|\,U^\dagger(\boldsymbol{s}) \tag{5.6}$$

(with the obvious guessing rule $\Pi(\boldsymbol{s}) \to |\boldsymbol{s}\rangle$) is optimal. In (5.6), we use the standard notation, where $\{|jm\rangle\}_{m=-j}^{j}$ is the eigenbasis of the total angular momentum operators J^2 and J_z. We denote by $U(\boldsymbol{s}) = [u(\boldsymbol{s})]^{\otimes N}$, $u(\boldsymbol{s}) \in \mathrm{SU}(2)$, (the unitary representation of) the rotation that maps the unit (Bloch) vector $\hat{z}$ into $\boldsymbol{s}$ [thus $u(\boldsymbol{s})|\frac{1}{2}\frac{1}{2}\rangle = |\boldsymbol{s}\rangle$], and we have also introduced the definition $J \equiv N/2$. Note that the POVM $\{\Pi(\boldsymbol{s})\}$ acts on the symmetric subspace of largest total angular momentum J, of dimention $2J+1 = N+1$. In terms of J, (5.5) can also be written as

$$F = \frac{1}{2}\left(1 + \frac{J}{J+1}\right) \equiv \frac{1}{2}\left(1 + \Delta_J\right). \tag{5.7}$$

Mixed states span a much larger Hilbert space and the computation becomes more involved. It greatly simplifies in the total angular momentum basis, where the input state $\tau(\boldsymbol{n})$ is block-diagonal, as it is shown in (2.2.4). The general expression 2.19 for the global state can be recast as

$$\tau(\boldsymbol{n}) = \sum_{j=j_{\min}}^{J} \sum_{\alpha=1}^{n_j} p_{j\alpha}\tau_{j\alpha}(\boldsymbol{n}), \tag{5.8}$$

where $\tau_{j\alpha}(\boldsymbol{n})$ is the normalized mixed state

$$\tau_{j\alpha}(\boldsymbol{n}) = \frac{1}{Z_j} \sum_{m=-j}^{j} R^m\, U(\boldsymbol{n})\,|jm;\alpha\rangle\,\langle jm;\alpha|\,U^\dagger(\boldsymbol{n}), \tag{5.9}$$

with the definitions:

$$Z_j = \sum_{m=-j}^{j} R^m = \frac{R^{j+1} - R^{-j}}{R-1}, \quad R = \frac{1+r}{1-r} > 1. \tag{5.10}$$

Recall that the additional index α, where $\alpha = 1, 2, \ldots, n_j$, labels the various occurrences of the irreducible representation of total angular momentum j.

The multiplicity n_j in (2.15) can be rewritten as

$$\begin{aligned} n_j &= \binom{2J}{J-j} - \binom{2J}{J-j-1} \\ &= \binom{2J}{J-j} \frac{2j+1}{J+j+1}. \end{aligned} \tag{5.11}$$

In the sum (5.8)), j runs from $j_{\min} = 0$ ($j_{\min} = 1/2$) for N even (odd) to the maximum total angular momentum J, in contrast to the pure state case where only the maximum value J appears. The numbers $p_{j\alpha} > 0$ are the probabilities that the state $\tau(\boldsymbol{n})$ has quantum numbers j and α, i.e., $p_{j\alpha} = \operatorname{tr}[\mathbb{1}_{j\alpha}\tau(\boldsymbol{n})]$, where $\mathbb{1}_{j\alpha} = \sum_{m=-j}^{j} |jm;\alpha\rangle\langle jm;\alpha|$ is the projector onto the corresponding eigenspace. The projector onto the whole subspace of total angular momentum j is then

$$\mathbb{1}_j = \bigoplus_{\alpha=1}^{n_j} \mathbb{1}_{j\alpha}. \tag{5.12}$$

Since the input state is permutation invariant (under the interchange of the individual qubits) representations with the same j are just mere repetitions of the same representation, they contribute a multiplicative factor of n_j to the fidelity through the marginal probability $p_j = \sum_\alpha p_{j\alpha}$, which reads

$$\begin{aligned} p_j &= \left(\frac{1-r^2}{4}\right)^J n_j Z_j = \\ &= \left(\frac{1-r^2}{4}\right)^J \binom{2J}{J-j} \frac{2j+1}{J+j+1} \frac{R^{j+1}-R^{-j}}{R-1}. \end{aligned} \tag{5.13}$$

One can easily check that $\sum_j p_j = 1$, as it should be.

Because of the block diagonal form of the input states, an obvious optimal measurement consists of a direct sum of covariant POVMs,

$$\Pi(\boldsymbol{s}) = \bigoplus_{j=j_{\min}}^{J} \bigoplus_{\alpha=1}^{n_j} \Pi_{j\alpha}(\boldsymbol{s}), \tag{5.14}$$

where each of them is a straightforward generalization of Eq.(5.6):

$$\Pi_{j\alpha}(\boldsymbol{s}) = (2j+1)\, U(\boldsymbol{s})\, |j\ j;\alpha\rangle\langle j\ j;\alpha|\, U^\dagger(\boldsymbol{s}). \tag{5.15}$$

One can easily check that the completeness condition $\int ds\, \Pi(\boldsymbol{s}) = \mathbb{1}$ holds. The total fidelity then is

$$F = \frac{1}{2}\left(1 + \sum_{j=j_{\min}}^{J} p_j \Delta_j\right), \tag{5.16}$$

where Holevo [1982]

$$\Delta_j = \frac{\langle J_z \rangle_j}{j+1} = \frac{\mathrm{tr}\,[J_z\, \tau_j(\hat{z})]}{j+1}, \tag{5.17}$$

with $\tau_j(\hat{z})$ being any one of the normalized states defined in Eq. (5.9), (say, the one with $\alpha = 1$). A straightforward calculation gives

$$\langle J_z \rangle_j = \frac{1}{Z_j} \sum_{m=-j}^{j} m R^m = j - \frac{1}{R-1} + \frac{2j+1}{R^{2j+1}-1}. \tag{5.18}$$

Notice that for pure states, one has $R \to \infty$, and in turn $\langle J_z \rangle_J \to J$, in agreement with Eq. (5.7).

As will be shown in the next section, for asymptotically large N the probability p_j peaks at a value of $j \simeq rJ$, which gives the dominant and subdominant contributions to the sum in (5.16). Up to order $1/N$, and discarding exponentially vanishing contributions [e.g., $\sim R^{-rJ}$], the asymptotic fidelity turns out to be

$$F = 1 - \frac{1}{Nr}\frac{r+1}{2r} + \cdots\ . \tag{5.19}$$

This result is interesting on its own and, to the best of our knowledge, has not been presented before. Note that for pure states ($r = 1$) Eq. (5.19) agrees with the asymptotic expression of the fidelity in Eq. (5.5).

5.2 Abstention

In this section we focus on estimation protocols where the experimentalist is allowed not to produce an answer, or abstain, if the outcome of the measurement she performed cannot provide a good enough estimate of the unknown state. Obviously, F cannot decrease by excluding these abstentions from the average. In noisy scenarios, such as that considered here, F actually increases, as will be shown below. Our aim is to quantify this gain and find the optimal protocol. In our approach, the probability of abstention, Q, is kept fixed, rather than unrestricted, since usually in practical situations one cannot afford discarding an unlimited amount of resources/state preparations.

5.2.1 General framework

We denote by Π_0 the abstention operator, which in addition to the operators $\{\Pi_\chi\}$ form the POVM representing the measurement. Thus, the completeness relation reads

$$\sum_{\chi} \Pi_{\chi} + \Pi_0 = \mathbb{1}. \tag{5.20}$$

The probability of abstention (abstention rate) and that of producing an estimate (acceptance rate) are then given respectively by

$$Q = \int dn \, \mathrm{tr}\, [\Pi_0 \tau(\boldsymbol{n})] \text{ and } \bar{Q} = 1 - Q, \tag{5.21}$$

and the mean fidelity defined in (5.4) becomes now

$$F(Q) = \frac{1}{\bar{Q}} \sum_{\chi} \int dn \, f(\boldsymbol{n}, \boldsymbol{n}_{\chi}) \mathrm{tr}\, [\Pi_{\chi} \tau(\boldsymbol{n})] \,, \tag{5.22}$$

where notice that the sum does not include the Π_0 operator and $\bar{Q}$ takes into account the abstentions excluded from the average.

We next note that for any unitary transformation U of the type defined after Eq. (5.6), the operators $\{U \Pi_{\chi} U^{\dagger}, U \Pi_0 U^{\dagger}\}$ give the same value of Q and $F(Q)$ as the original set $\{\Pi_{\chi}, \Pi_0\}$, provided we change the guessing rule as $\boldsymbol{n}_{\chi} \to \mathscr{R}_U \boldsymbol{n}_{\chi}$, where $\mathscr{R}_U$ is the SO(3) rotation whose unitary representation is U. Therefore, one can easily prove that Π_0 (the set $\{\Pi_{\chi}\}$) can always be chosen to be SU(2) invariant (covariant) by simply averaging over U. In other words, with no loss of generality the POVM elements that provide a guess $|\boldsymbol{s}\rangle$ can be chosen as

$$\tilde{\Pi}(\boldsymbol{s}) = U(\boldsymbol{s}) \, \Pi \, U^{\dagger}(\boldsymbol{s}), \tag{5.23}$$

where $\Pi \geq 0$ is the so called seed of the POVM (in particular, note that $\tilde{\Pi}(\hat{z}) = \Pi$). The abstention operator then reads

$$\Pi_0 = \mathbb{1} - \int ds \, \tilde{\Pi}(\boldsymbol{s}), \tag{5.24}$$

which is manifestly rotationally invariant (as claimed above). It is thus proportional to the identity on each invariant subspace

$$\Pi_0 = \bigoplus_{j=j_{\min}}^{J} \bigoplus_{\alpha=1}^{n_j} a_{j\alpha} \mathbb{1}_{j\alpha} = \bigoplus_{j=j_{\min}}^{J} a_j \mathbb{1}_j, \tag{5.25}$$

where a_j are coefficients that satisfy the condition $0 \leq a_j \leq 1$ and $\mathbb{1}_j$ is defined in Eq. (5.12). Here we have used the permutation invariance of the input state to fix, without loss of generality, $a_{j\alpha} = a_j$ for all α). We can also choose $\tilde{\Pi}(\boldsymbol{s})$ to have the block-diagonal form of the input state $\tau(\boldsymbol{n})$, namely,

$$\tilde{\Pi}(\boldsymbol{s}) = \bigoplus_{j=j_{\min}}^{J} \bigoplus_{\alpha=1}^{n_j} \tilde{\Pi}_{j\alpha}(\boldsymbol{s}). \tag{5.26}$$

For given $\{a_j\}$, the optimality of $\Pi_{j\alpha}(\boldsymbol{s})$, defined in Eq. (5.15), clearly ensures that

$$\tilde{\Pi}_{j\alpha}(\boldsymbol{s}) = (1 - a_j)\Pi_{j\alpha}(\boldsymbol{s}), \tag{5.27}$$

are also optimal for estimation with abstention. Recalling that the label α is unsubstantial, aside from the multiplicative factor n_j, we have from (5.21) that the abstention probability is simply

$$Q = \sum_{j=j_{\min}}^{J} p_j a_j, \tag{5.28}$$

where p_j is given in Eq. (5.13). The coefficients a_j can be understood as the probabilities of abstention conditioned to the input state having total angular momentum j, i.e., $a_j = p(\text{abstention}|j)$. Similarly, for a given j, the probability of producing an estimate, or accepting, is $\bar{a}_j = 1 - a_j = p(\text{acceptance}|j)$.

From Eq. (5.22) we obtain

$$F(Q) = \frac{1}{2}\left(1 + \sum_{j=j_{\min}}^{J} p_j \tilde{\Delta}_j\right), \tag{5.29}$$

where

$$\tilde{\Delta}_j = \frac{1 - a_j}{1 - Q}\Delta_j = \frac{\bar{a}_j}{\bar{Q}}\Delta_j, \tag{5.30}$$

and the quantity Δ_j is given in Eqs. (5.17) and (5.18). Thus, we are only left with the free parameters a_j, which have to be optimized in order to maximize $F(Q)$, subject to the constraints $0 \le a_j \le 1$ and (5.28). Somehow expected, one can show that Δ_j is a monotonically increasing function of j, i.e., $\Delta_{j-1} < \Delta_j$, therefore the largest contribution to the fidelity is given by Δ_J. This corresponds to $\bar{a}_J = 1$ and $\bar{a}_j = 0$, $j < J$. Hence, for unrestricted probability of abstention, the optimal protocol discards any contribution with $j < J$. This protocol, however, would provide an estimate with a probability that decreases exponentially with N, for $r < 1$, as $p_J \simeq (1/r)[(1 + r)/2]^{N+1}$. Notice that in a noiseless scenario, $r = 1$, there is only the contribution $j = J$, which is already the optimal one and therefore abstention is of no use in such case.

Clearly, for finite Q there can be contributions from other total angular momentum eigenspaces ($j < J$) compatible with Eq. (5.28). Recalling the monotonicity of Δ_j, and by convexity, it is obvious from Eqs. (5.29) and (5.30) that there must exist an angular momentum threshold j^* such that $\bar{a}_j = 0$ ($\bar{a}_j = 1$), if $j < j^*$ ($j > j^*$). The value j^* is determined through Eq. (5.28) to be

$$j^* = \max\left\{j \text{ such that } Q - \textstyle\sum_{j'=j_{\min}}^{j-1} p_{j'} \ge 0\right\}. \tag{5.31}$$

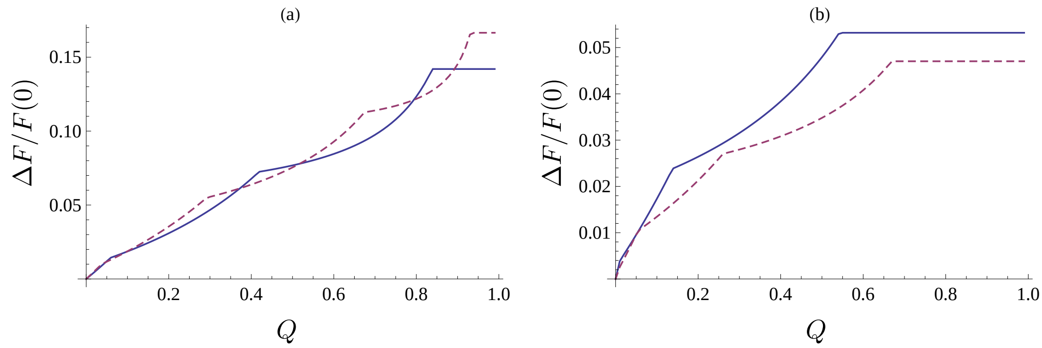


Figure 5.1. Fidelity gain $\Delta F/F(0) = [F(Q) - F(0)]/F(0)$ as a function of Q for $N = 6$ (solid line) and $N = 8$ (dashed line) and purities of $r = 0.3$ in (a) and $r = 0.7$ in (b).

Thus, we have

$$a_j = \begin{cases} 1, & j < j^*; \\ p_j^{-1}\left(Q - \sum_{j'=j_{\min}}^{j^*-1} p_{j'}\right), & j = j^*; \\ 0, & j > j^*. \end{cases} \tag{5.32}$$

In a more physical language, the optimal strategy consists actually of two successive measurements. The experimentalist first measures the total angular momentum j of the input state $\tau(\boldsymbol{n})$ and decides to abstain (provide a guess) if $j < j^*$ ($j > j^*$). If $j = j^*$, she simply decides randomly, by tossing a Bernoulli coin with probability a_{j^*} of coming up heads, and if heads (tails) show up, abstain (provide a guess). In order to provide the actual guess, if she decides to do so, she performs the optimal POVM measurement $\{\Pi(\boldsymbol{s})\}$ [or just $\{\bigoplus_\alpha \Pi_{j\alpha}(\boldsymbol{s})\}$] in Eq. (5.14) on the state $\bigoplus_\alpha \tau_{j\alpha}(\boldsymbol{n})$ that resulted from the first measurement.

5.2.2 Small number of copies

In Fig. 5.1 we plot the fidelity gain due to abstention $[F(Q) - F(0)]/F(0) = \Delta F/F(0)$ vs. Q for $N = 6$, $N = 8$, and purities of $r = 0.3$ and $r = 0.7$. The structure of Eq. (5.32) is apparent from these plots: at $Q = 0$ ($a_j = 0$ for all j) there is, naturally, no gain; kinks sequentially appear at the precise values of Q where a new coefficient a_j in (5.32) becomes positive (and j^* increases by one); the curves are convex between successive kinks, where the one a_j that has become positive, a_{j^*}, keeps increasing. This pattern repeats until the abstention rate Q reaches a critical value Q_{crit} at which $j^* = J$,

$$Q_{\text{crit}} = 1 - p_J = 1 - \frac{1}{r}\left(\frac{1+r}{2}\right)^{2J+1} + \frac{1}{r}\left(\frac{1-r}{2}\right)^{2J+1} \tag{5.33}$$

[see Eq. (5.13)]. Increasing Q further will not provide any additional gain, as the flat plateaus of Fig. 5.1 illustrate. This is so, since one can view the optimal abstention protocol as a filtering process where the low angular momentum components of the input state are filtered out. Hence, keeping the maximum value of $j = J$ is the optimal filtering beyond which no further improvement is possible. Fig. 5.1(a) shows that in noisy scenarios, e.g. $r = 0.3$, abstention can increase the fidelity quite notably, up to 15%. For higher purities the gain is more moderate, as shown in Fig. 5.1(b). The enhancement in this case is about 4-5% but with an abstention rate slightly above 50%. Further results are shown in Fig. 5.2, where we plot the fidelity gain as a function of the number of copies N for abstention rates larger than Q_{crit}, and for various values of the purity r. All the curves have a maximum at a value of N that varies with the purity. The lower the purity, the higher the value of N at which the maximum occurs (e.g., for $r = 0.3$ the maximum gain occurs at $N = 12$; for $r = 0.1$ the maximum is off scale at the right of the figure).

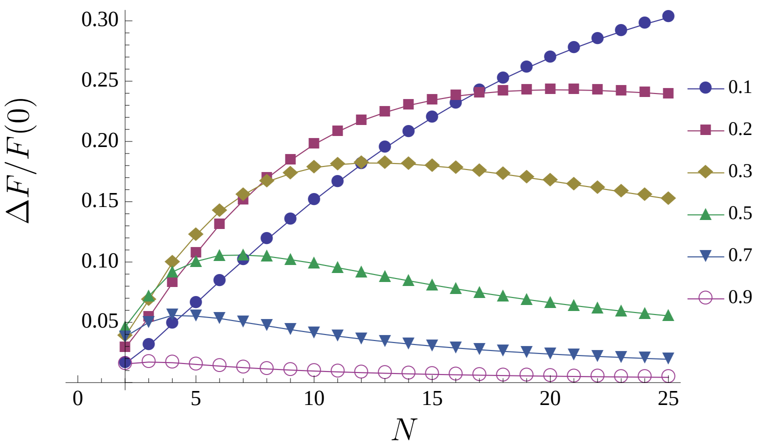


Figure 5.2. Fidelity gain as a function of N for various values of r, indicated in the legend, and for $Q \geq Q_{\text{crit}}$.

As we have seen, the possibility of abstaining enables us to reach values of the fidelity that otherwise we could only attain with lower levels of noise. To quantify this effective reduction of noise, let us define an effective purity r_{eff} by the implicit equation $F(r_{\text{eff}}, N, 0) = F(r, N, Q)$. That is, for an estimation setting, given by r, N,and Q, r_{eff} is the purity of the input states that would provide the same fidelity if the standard strategy without abstention ($Q = 0$) were used instead. Since r is related to the probability of error η in our model of noisy measurements in (5.1), an increase of the effective purity corresponds to an effective reduction of the amount of noise in the measurement through the relation $\eta_{\text{eff}} = (3/4)(1 - r_{\text{eff}})$. Figure 5.3 shows a plot of the effective purity r_{eff} as a function of Q for various values of r and N. As can be seen, r_{eff} increases faster at low values of N, but it saturates earlier (lower Q_{crit}), reaching a lower value. For low N and for a wide range of purities, $0.1 \lesssim r \lesssim 0.9$, we observe a constant effective increase of the purity, $r_{\text{eff}} \approx r + 0.2$, for reasonable values of the abstention rate Q. As N increases one has to go to higher values of the abstention rate, $Q \sim Q_{\text{crit}}$, to have a significant gain. Hence, a moderate abstention rate is most effective in noisy scenarios when a small, but fair, number of copies is available.

Finally, let us point out that the protocol we have presented requires a projection on the total angular momentum eigenspaces. This is a non-local measurement that nonetheless can be implemented efficiently Bacon *et al.* [2006]. In a more extreme scenario where there are no restriction on the abstention rate, one can attain the maximum fidelity with an even simpler strategy: perform a local Stern-Gerlach measurement on every qubit (say, of the z-component of the spin) and abstain unless all outcomes agree. This strategy renders an abstention probability of $Q = 1 - [(1 + r)/2]^N$, which might be comparable to Q_{crit} in Eq. (5.33).

5.2.3 Asymptotic regime

We next compute the analytical expressions of the fidelity in the large N limit. Here it is useful to define the variable x as

$$x = \frac{j}{J}, \quad 0 \leq x \leq 1, \tag{5.34}$$

which becomes continuous in the limit $N \to \infty$ ($J \to \infty$). In this case, we can replace p_j by the continuous probability distribution in $[0, 1]$ defined by

$$p(x) = J p_{j=xJ}, \tag{5.35}$$

so that $\int_0^1 dx\, p(x) = 1$ as N goes to infinity. Eq. (5.29) can then be approximated by its continuous version, which reads

$$F = \frac{1}{2}\left[1 + \int_0^1 dx\, p(x) \tilde{\Delta}(x)\right], \tag{5.36}$$

where

$$\tilde{\Delta}(x) = \tilde{\Delta}_{j=xJ}, \tag{5.37}$$

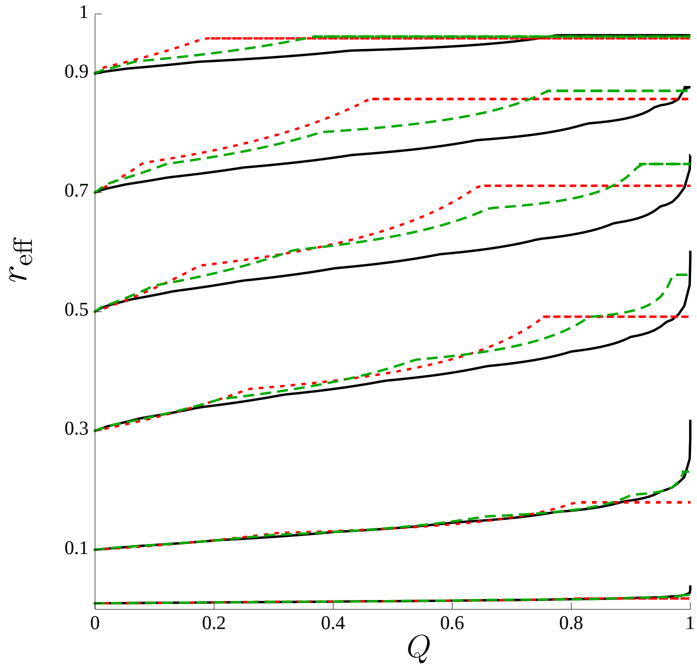


Figure 5.3. Effective purity r_{eff} as a function of the abstention rate Q for $N = 5$ (dotted), 10 (dashed), and 30 (solid), and for purities of $r = 0.01,\ 0.1,\ 0.3,\ 0.5,\ 0.7,$ and 0.9, which can be read off from the values of r_{eff} at $Q = 0$.

where recall that $\tilde{\Delta}_j$ is given in Eqs. (5.30). From Eq. (5.32) we see that asymptotically $\bar{a}_j$ becomes the step function $\theta(x - x^*)$, where $x^* = j^*/J$, and we have used the standard definition

$$\theta(x) = \begin{cases} 1, & x \geq 0; \\ 0, & x < 0. \end{cases} \tag{5.38}$$

With this, Eq. (5.30) becomes

$$\tilde{\Delta}(x) = \frac{\theta(x - x^*)}{\bar{Q}} \Delta(x), \tag{5.39}$$

and, in turn,

$$F = \frac{1}{2} \left[1 + \frac{1}{\bar{Q}} \int_{x^*}^{1} dx\, p(x) \Delta(x) \right]. \tag{5.40}$$

It also follows from (5.28) that

$$\bar{Q} = \int_0^1 dx\, p(x) \theta(x - x^*) = \int_{x^*}^{1} dx\, p(x). \tag{5.41}$$

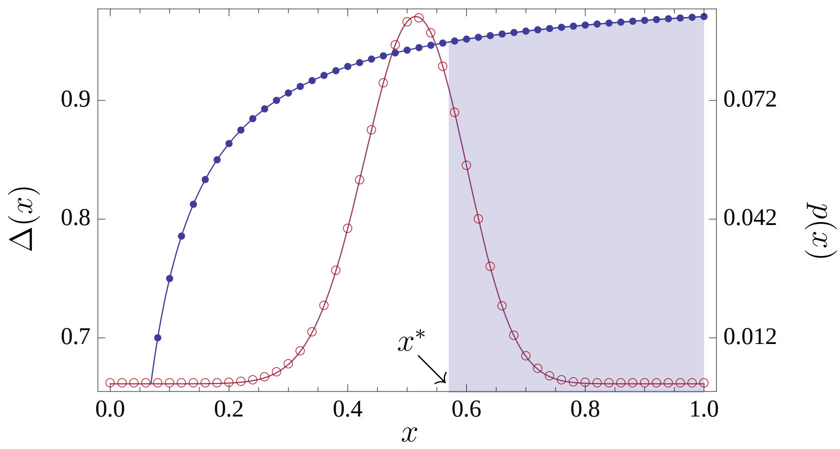


Figure 5.4. Plots of $\Delta(x)$ (blue line with solid circles) and $p(x)$ (red line with empty circles) for $N = 100$ and $r = 0.5$. The circles represent the quantities Δ_j and Jp_j as a function of $x = j/J$. The shaded area indicates the acceptance region for an abstention rate $Q \sim 93\%$.

At this point, we need to find a good approximation to $p(x)$ that would enable us to obtain the explicit form for the asymptotic fidelity. From Eq. (5.13), and using the Stirling formula, we obtain

$$p(x) \simeq \sqrt{\frac{N}{2\pi}} \frac{1}{\sqrt{1-x^2}} \frac{x(1+r)}{r(1+x)} \, \mathrm{e}^{-NH(\frac{1+x}{2} \| \frac{1+r}{2})}, \tag{5.42}$$

where $H(s \parallel t)$ is the (binary) relative entropy

$$H(s \parallel t) = s \log \frac{s}{t} + (1-s) \log \frac{1-s}{1-t}, \tag{5.43}$$

and the approximation is valid for both x and r in the open unit interval $(0, 1)$. The appearance of a relative entropy in Eq.(5.42) can be understood as follows. Our N-copy input state (diagonal in the canonical J_n basis) can be thought of as a classical coin tossing distribution of N identical coins with a bias of $(1 + r)/2$. From the theory of types Cover and Thomas [2006] it is well known that the probability to get k heads is given by the Kulback-Leibler distance (or relative entropy) between the empirical distribution $\{f = k/N, 1 - f\}$ and the distribution $\{(1+r)/2, (1-r)/2\}$. That is, $p(k) \sim \exp\{-NH[f \parallel (1+r)/2]\}$ to first order in the exponent. The number of heads k is in one-to-one correspondence with the magnetic quantum number, $m = k - J$, and the conditioned probability $p(j|m)$ is strongly peaked at $m = j$, as one can easily check. It follows that the probability that the input state has total angular momentum j, given by $p(j) = \sum_m p(j|m)p(m)$, will be asymptotically determined by the probability distribution $p(m)$, which has a convenient expression in terms of the typical and the empirical distribution of up/down outcomes.

From Eq. (5.42) it follows that $p(x)$ is peaked at the value $x = r$, i.e. at $j = rJ$, as shown in Fig. 5.4 and stated without a proof in Sec. 5.1. Actually, around the peak, $x \sim r$, the exponent becomes quadratic and $p(x)$ approaches the Gaussian distribution

$$p(x) \simeq \sqrt{\frac{N}{2\pi(1-r^2)}}\,\mathrm{e}^{-N\frac{(x-r)^2}{2(1-r^2)}}, \tag{5.44}$$

as also follows from the central limit theorem, whereas it falls off exponentially elsewhere.

It is now apparent that, asymptotically, abstention has negligible impact if components with j below rJ are filtered out ($x^* < r$), since the main contribution to the fidelity, which comes from the peak around $x \simeq r$, is not excluded from the integral in Eq. (5.40) (only the left exponentially decaying tail is). For the same reason [see Eq. (5.41)], $\bar{Q} \simeq 1$ (the abstention rate Q is exponentially small), and Eq. (5.40) yields

$$F = 1 - \frac{1}{2N}\frac{r+1}{r^2} + \ldots \qquad \text{for} \quad x^* < r, \tag{5.45}$$

which is the same expression as the asymptotic fidelity of the protocol without abstention, Eq. (5.19).

It is then clear that, in order to have a discernible improvement in the fidelity, the abstention threshold x^* must lie to the right of the peak of the probability distribution. The fidelity in (5.40) then can be written as

$$F \simeq \frac{1}{2}\left[1 + \frac{p(x^*)}{\bar{Q}}\Delta(x^*)\right] \simeq \frac{1}{2}\left[1 + \Delta(x^*)\ \right], \quad x^* > r, \tag{5.46}$$

where we have used that for $x \geq x^* > r$ and for large enough N, $p(x)$ falls off exponentially and the integral can be approximated by the value of the integrand at its lower limit. By the very same argument Eq. (5.41) gives

$$\bar{Q} \simeq p(x^*), \tag{5.47}$$

which has also been used in (5.46). Using now (5.42) we obtain that in the asymptotic limit of many copies, the rate at which our protocol provides a guess is

$$\bar{Q} \sim \exp\left[-NH\left(\tfrac{1+x^*}{2} \parallel \tfrac{1+r}{2}\right)\right]. \tag{5.48}$$

Recalling Eqs. (5.17) and (5.18) we obtain the optimal fidelity:

$$F = 1 - \frac{1}{2Nx^*}\frac{r+1}{r} + \cdots, \quad \text{for } r \le x^* \le 1, \tag{5.49}$$

for a value of Q given by (5.48). For $x^* = r$ the results (5.19) and (5.45) are recovered, whereas for $x^* \to 1$ ($Q \ge Q_{\text{crit}}$) the maximum average fidelity is attained

$$F_{\max} = 1 - \frac{1}{2N}\frac{r+1}{r} + \cdots. \tag{5.50}$$

The advantage provided by our estimation with abstention protocol can be quantified by the effective number of copies that the standard protocol without abstention would require to achieve the same fidelity: $N_{\text{eff}} = (x^*/r)N$, where $x^* \in [r, 1)$ is determined by the abstention rate Q through (5.48). For high noise levels (low purity, $r \ll 1$) our protocol provides an important saving of resources/copies, as $N_{\text{eff}}/N = 1/r \gg 1$, whereas for nearly ideal detectors the saving in this asymptotic regime is more modest.

Alternatively, the advantage discussed above can also be quantified by the effective measurement-noise reduction, or equivalently, the effective purity r_{eff} (See Sec. 5.2.2). Using (5.49) one can easily find a simple expression for the effective purity in the asymptotic limit and for large abstention rate: $r_{\text{eff}} = (r + \sqrt{4r + 5r^2})/[2(1 + r)]$. In the limit of very low noise levels the errors probability η [recall Eq. (5.1)] is effectively reduced by a factor of three, i.e., $\eta_{\text{eff}} = \eta/3$, while in the opposite limit of very noisy measurements one finds $r_{\text{eff}} = \sqrt{r}$.

5.2.4 Other regimes

In the previous section we have seen how a gain in fidelity can be obtained provided the 'acceptance' rate $\bar{Q}$ falls off exponentially as N becomes very large. Here we give an example where this gain takes place even at finite $\bar{Q}$.

At fixed noise level (purity r), the fidelity is an increasing function of N. However, one could imagine an experimental setup where the noise (purity) also increases (decreases) with N. If this is so, the asymptotic fidelity could be strictly less than one, or in other words, perfect estimation could be unattainable even with unbounded resources. This is the case in our example, were we assume that $r = a/\sqrt{N}$, a being a positive constant. Notice that the threshold x^* must also scale as $1/\sqrt{N}$ in order to have a reasonably low abstention rate. Therefore, it is convenient to use a new variable $\xi = \sqrt{N}x = \sqrt{N}\, j/J = 2j/\sqrt{N}$ instead. Then, the probability distribution in this new variable is

$$p(\xi) = \frac{\sqrt{N}}{2} p_{j=\xi\sqrt{N}/2}, \quad \text{with } r = \frac{a}{\sqrt{N}}. \tag{5.51}$$

Recalling Eq. (5.13) and using Stirling formula this equation gives

$$p(\xi) = \frac{\mathrm{e}^{-\left(\frac{\xi-a}{\sqrt{2}}\right)^2} - \mathrm{e}^{-\left(\frac{\xi+a}{\sqrt{2}}\right)^2}}{\sqrt{2\pi}a}\xi \tag{5.52}$$

to leading order in inverse powers of N. The subleading terms are of order $N^{-1/2}$ and will be neglected here. For a given threshold value $\xi^* = 2j^*/\sqrt{N}$ the abstention rate is

$$Q = \int_0^{\xi^*} p(\xi)\, d\xi = \frac{1}{2}\left(\operatorname{erf}\xi_+^* + \operatorname{erf}\xi_-^*\right) - \frac{\mathrm{e}^{-\xi_-^{*\,2}} - \mathrm{e}^{-\xi_+^{*\,2}}}{\sqrt{2\pi}a}, \tag{5.53}$$

where $\xi_\pm^* = (\xi^* \pm a)/\sqrt{2}$ and $\operatorname{erf} x$ is the error function.

From Eqs. (5.17) and (5.18) we have in this same regime and at leading order

$$\Delta(\xi) = \Delta_{j=\xi\sqrt{N}/2} = 1 - \frac{2}{1-\mathrm{e}^{2a\xi}} - \frac{1}{a\xi}. \tag{5.54}$$

With the above, the fidelity (5.29), [or rather, the counterpart of (5.36)] is

$$F = \frac{1}{2}\left[1 + \int_0^\infty d\xi\, p(\xi)\tilde{\Delta}(\xi)\right] = \frac{1}{2}\left[1 + \frac{1}{\bar{Q}}\int_{\xi^*}^\infty d\xi\, p(\xi)\Delta(\xi)\right], \tag{5.55}$$

where the last integral can be computed to be

$$\begin{aligned} \Delta^* &\equiv \int_{\xi^*}^\infty \Delta(\xi)p(\xi)\, d\xi = \\ &= \frac{1-a^2}{2a^2}\left(\operatorname{erf}\xi_-^* - \operatorname{erf}\xi_+^*\right) + \frac{\mathrm{e}^{-\xi_-^{*\,2}} + \mathrm{e}^{-\xi_+^{*\,2}}}{\sqrt{2\pi}a}. \end{aligned} \tag{5.56}$$

We can finally write the fidelity as

$$F = \frac{1}{2}\left(1 + \frac{\Delta^*}{1-Q}\right) + \mathscr{O}(N^{-1/2}). \tag{5.57}$$

As shown in (5.53) and (5.56), both Q and Δ^* are functions of the filtering threshold ξ^*, which is just a properly scaled version of the original threshold j^*. Finding the maximum fidelity for a given rate of abstention Q requires inverting Eq. (5.53) to obtain $\xi^*(Q)$, but this cannot be done analytically and one has to resort to numerical methods.

In Fig. 5.5 we plot F as a function of Q for $a = 1$. The increase of the fidelity in the asymptotic regime of large N is clearly seen: e.g., an abstention rate of a 50% yields a rise of about 10%, and it goes up to about 30% for

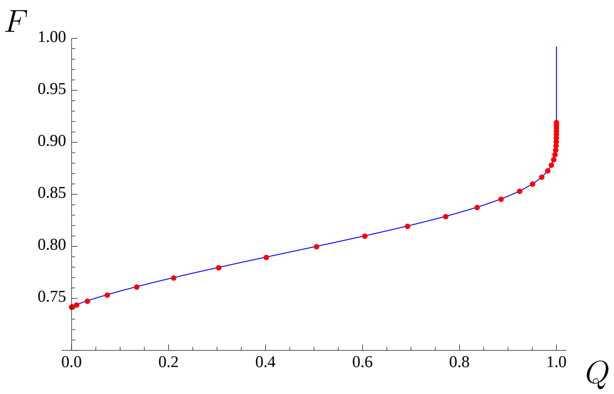


Figure 5.5. Plot of the fidelity as a function of Q for $r = a/\sqrt{N}$, with the choice $a = 1.0$, $N = 10^6$ (red circles). The solid line (in blue) is the leading term in Eq. (5.57) plotted as a function of Q [a parametric plot of the pairs (Q, F), as given by Eqs. (5.53) and (5.57)]

higher (but still reasonable) values of Q. The figure also shows the agreement between the approximate form of the fidelity given by Eqs. (5.53) to (5.57) and the numerical evaluation of its exact expression in (5.29).

It should be noted that in the regime described here a rise of the input size N fails to replicate the fidelity improvement that results from increasing the rate of abstention (no N_{eff} can be defined in this regime), thus abstention appears to be the only means by which one can improve estimation.

5.3 Conclusions

In this Chapter we have considered the probabilistic direction estimation with mixed states. We have seen that the optimal probabilistic protocol consist in measuring first the total angular momentum of the state followed by a covariant POVM. The first measurement projects the state of the system over a subspace and gives a label that heralds the precision of the proper direction estimation. By post-selecting over the outcomes of the former measurement, abstention counterbalances the adverse effect of errors in a noisy process of measurement.

We have shown that in general abstention is most useful for inputs of few copies and for error rates of the order of a few percent. We have given analytical asymptotic expressions of the fidelity valid in the limit of large number of copies. In this limit, abstention can have the effect of increasing the number of copies by a constant fraction: $N_{\text{eff}}/N = x^*/r$ $(x^* > r)$, with an acceptance rate $\bar{Q}$ given by the relative entropy: $-(1/N)\log\bar{Q} = H[(1 + x^*)/2 \parallel (1+r)/2]$. For low levels of noise this amounts to reducing the error probability η by a factor of up to three. We have also considered a scenario where the noise (per qubit) increases with the number of copies in such a way that perfect estimation is unattainable ($\lim_{N\to\infty} F < 1$). In this case one can obtain a significant enhancement of the asymptotic fidelity (few percent) even for finite abstention probabilities $Q < 1$. Moreover, in such scenario abstention appears to be the only way to improve estimation.

The case of direction estimation with mixed states has an important difference with respect the other cases discussed in this thesis: if the experimentalist is forced to give an answer, the probabilistic strategy presented in this Chapter turn out to be exactly the same as the deterministic one. However, although the averaged fidelity does not differ from the one obtained by the deterministic strategy, by measuring first the total angular momentum the experimentalist gets a label for each sample element that indicates the accuracy of this precise estimate.

CHAPTER 6

Probabilistic phase estimation of mixed states

In the previous two chapters we have seen two different facets of the probabilistic filter. In the absence of noise it can be regarded as a filter that tunes the state to a more parameter sensitive one. In this sense, probabilistic metrology will not provide an enhanced accuracy if the experimentalist has no limitations to prepare any state. Of course, in many instances the state is just given and it is worth considering a probabilistic approach. In the case of direction estimation in the presence of noise, we have seen that the filter has a different action. While the former exploits the quantum correlations of the system by modifying the profile of the state, here the protocol uses the classical correlations via a total angular momentum filter.

In this Chapter we deal with the paradigmatic case of phase estimation with mixed states. Interestingly, the optimal protocol includes a combination of the two types of the filter's actions just mentioned, and yields a precision beyond the ultimate deterministic bounds. That is, there is no preparation/probe/sginal state that under the effect of local noise gives the same accuracy as the the one that a probabilistic protocol can provide.

6.1 Optimal probabilistic measurement for n-qubits

As stated in the introduction, here we are devoted to find the optimal probabilisitic protocol for estimating the parameter θ that determines the unitary evolution, $U_\theta := u_\theta^{\otimes n}$, of a probe system of n qubits in the presence of local decoherence, where $u_\theta = \exp(i\theta\,|1\rangle\langle 1|)$. In particular, the initial n-partite pure state $|\psi\rangle\langle\psi| = \psi$ (this shorthand notation will be used throughout this Chapter) is prepared and is let evolve. The state is affected by uncorrelated dephasing noise, which can be modeled by independent phase-flip errors occurring with probability $p_f = (1-r)/2$ for each qubit. Its action on the n-qubits is described by a map $\mathscr{D}$ that commutes with the Hamiltonian, so that it could as well be understood as acting before or during the phase imprinting process.

Next, the experimentalist performs a suitable measurement on $\rho_\theta = \mathscr{D}(U_\theta \psi U_\theta^\dagger)$ and, based on its outcome, decides whether to abstain or to produce an estimate $\hat\theta$ for the unknown parameter θ. Recall that this decision must be based solely on the outcome of the measurement as, naturally, the actual value of θ is unknown to the experimentalist. Our aim is to find the optimal protocol, e.g., the measurement that gives the most accurate estimates for a given probe state and for a given maximum probability of abstention.

As we have done in the other cases treated in this Thesis, we quantify the precision of the estimated phase $\hat\theta$ by the fidelity

$$f(\theta,\hat\theta) = [1+\cos(\theta-\hat\theta)]/2, \tag{6.1}$$

but in this Chapter to assess the performance of the protocol we use the worst-case fidelity over the possible values of the unknown parameter.

$$F = \inf_{\theta\in(-\pi,\pi]} \int \mathrm{d}\hat\theta\, p(\hat\theta|\theta,\mathrm{succ}) f(\theta,\hat\theta), \tag{6.2}$$

where $p(\hat\theta|\theta,\mathrm{succ})$ is the probability of estimating $\hat\theta$ after a successful event when the true value is θ. The fidelity F and the probability of success S will fully characterize our probabilistic metrology strategies. However, for covariant families of states, such as our noisy probes $\{\rho_\theta\}_{\theta\in(-\pi,\pi]}$, the worst-case fidelity is entirely equivalent to the average fidelity (see Appendix A.6). To facilitate comparison with previous point-wise results, we present ours in terms of a scaled 'infidelity', $\sigma^2 := 4(1-F)$, which approximates the mean-square error when the distribution $p(\hat\theta|\theta,\mathrm{succ})$ becomes peaked around the true value θ Berry *et al.* [2012].

Because of the symmetry of the problem, there is no loss of generality in choosing the covariant measurement defined by $\{M_{\hat\theta} = U_{\hat\theta}\Omega U_{\hat\theta}^\dagger/(2\pi)\}_{\hat\theta\in(-\pi,\pi]}$, where Ω is the so-called *seed* of the measurement. In addition, we have the invariant measurement operator $\Pi = \mathbb{1} - \int_0^{2\pi} d\hat\theta/(2\pi) U_{\hat\theta}\Omega\, U_{\hat\theta}^\dagger \leq \mathbb{1}$ that corresponds to the abstention event. With this, finding the optimal estimation scheme reduces to finding the operator Ω that maximizes the fidelity

$$F(S) \;=\; \frac{1}{S}\max_{\Omega} \int \frac{d\hat\theta}{2\pi} f(0,\hat\theta)\,\mathrm{tr}\Big(U_{\hat\theta}\Omega U_{\hat\theta}^\dagger \rho\Big) \tag{6.3}$$

for a fixed success probability

$$S = \int \frac{d\hat{\theta}}{2\pi} \mathrm{tr}\Big(U_{\hat{\theta}} \Omega U_{\hat{\theta}}^{\dagger} \rho\Big) \tag{6.4}$$

In deriving Eq. (6.3) we have used covariance to get rid of the infimum in (6.2), thereby formally fixing the value of θ to zero, and have defined $\rho = \mathscr{D}(\psi)$ accordingly.

6.2 Symmetric probes

We now focus on probe states consisting n-qubits that are initially prepared in a permutation invariant state. This family includes most of the states considered in the literature, our case-study of multiple copies of equatorial-states, and also, as we will show below, the optimal probe-state for probabilistic metrology. The input state is given by,

$$|\psi\rangle = \sum_{m=-J}^{J} c_m |J, m\rangle \,, \tag{6.5}$$

where $J = n/2$ is the maximum total spin angular momentum (hereafter spin for short) of n qubits and the set of states $\{|J, m\rangle\}_{m=-J}^{J}$ spans the fully-symmetric subspace. Given the permutation invariance of the noisy channel, the state $\rho = \mathscr{D}(\psi)$ inherits the symmetry of the probe, and can be conveniently written in a block diagonal form in the total spin bases (see Section 2.2.4 and Appendix A.3),

$$\rho = \sum_j p_j \rho^j \otimes \frac{\mathbb{1}_j}{\nu_j}, \tag{6.6}$$

where the state ρ^j has unit trace, p_j is the probability of ρ having spin j, and $\mathbb{1}_j$ stands for the identity in the ν_j-dimensional multiplicity space of the irreducible representation of spin j. The sum over j in (6.6) runs from $j_{\min} = 0$ ($j_{\min} = 1/2$) for n even (odd) to the maximum spin J. Similarly, the measurement operators, can be taken to have the same symmetry and thus be of the form $\Omega = \sum_j |\chi_j\rangle\langle\chi_j| \otimes \mathbb{1}_j$, where $|\chi_j\rangle = \sum_m f_m^j |j, m\rangle$, $0 \leq f_m^j \leq 1$. The maximum fidelity $F(S)$ for a fixed probability of success S can hence be expressed in terms of the fidelity $F_j(s_j)$ in each irreducible block and its corresponding success probability s_j,

$$F(S) = \max_{s_j} \sum_j \frac{p_j s_j}{S} F_j(s_j), \quad S = \sum_j p_j s_j \tag{6.7}$$

where

$$F_j(s_j) = \frac{1}{2}\left(1+\frac{1}{s_j}\max_{0\leq f^j_m\leq 1}\sum_m f^j_m \rho^j_{m,m+1} f^j_{m+1}\right),$$
$$\text{subject to } s_j = \sum_m (f^j_m)^2 \rho^j_{m,m}. \tag{6.8}$$

This formulation of the problem allows for the natural (and widely used in this Thesis) interpretation of the probabilistic protocol as a two step process: i) a stochastic filtering channel

$$\mathscr{F}(\rho) = \Phi\,\rho\,\Phi, \qquad \Phi = \sum_{j,m} (f^j_m)^2 \, |j,m\rangle\langle j,m| \otimes \mathbb{1}_j, \tag{6.9}$$

that coherently transforms each basis vector as $|j,m\rangle \to f^j_m\,|j,m\rangle$, so that it modulates the input to a state with enhanced phase-sensitivity, followed by ii) a canonical covariant measurement with seed $\tilde{\Omega} = \sum_j \sum_{m,m'} |j,m\rangle\langle j,m'| \otimes \mathbb{1}_j$ performed on the transformed state from which the value of the unknown phase is estimated.

By defining the vector ξ^j with components given by $\xi^j_m = f^j_m (\rho^j_{m,m}/s_j)^{1/2}$ and introducing the tridiagonal symmetric matrix H^j, with entries

$$\begin{aligned} H^j_{m,m'} &= 2\delta_{m,m'} - a^j_m \delta_{m,m'-1} - a^j_{m'}\delta_{m-1,m'}, \\ a^j_m &= \frac{\rho^j_{m,m+1}}{\sqrt{\rho^j_{m,m}\rho^j_{m+1,m+1}}}, \end{aligned} \tag{6.10}$$

we can easily recast the former optimization problem as,

$$F_j(s_j) = 1 - \frac{1}{4}\sigma^2_j, \quad \sigma^2_j := \min_{|\xi^j\rangle} \left\langle \xi^j \right| H^j \left| \xi^j \right\rangle, \tag{6.11}$$

$$\text{subject to } \langle \xi^j|\xi^j\rangle = 1 \text{ and } 0 \leq \xi^j_m \leq (\rho^j_{m,m}/s_j)^{1/2}, \tag{6.12}$$

Note that a^j_m, and in turn H^j, depend on the strength of the noise but take the same values for *all* symmetric probe states, for we have that $\rho^j_{m,m'} \propto c_m c_{m'}$. For deterministic strategies ($S=1$, i.e., $s_j = 1$ for all j) no minimization is required and one only needs to evaluate the expectation values of H^j for the 'state' $\xi^j_m = (\rho^j_{m,m})^{1/2}$. For large enough abstention, the problem becomes an unconstrained minimization, so σ^2_j is the minimal eigenvalue of H^j, and $|\xi^j\rangle$ its corresponding eigenstate. From (6.12) we find that the corresponding filtering operation only succeeds with a probability (6.12)

$$S^* = \sum_j p_j s^*_j, \qquad s^*_j = \min_m \frac{\rho^j_{m,m}}{{\xi^j_m}^2}. \tag{6.13}$$

We will refer to S^* as the critical success probability, since the fidelity will not improve by decreasing the success probability below this value, $F(S) = F(S^*)$ for $S \leq S^*$.

6.3 Asymptotic scaling: particle in a potential box

In order to compute the scaling of the precision as the number of resources becomes very large we need to solve the above optimization problem in the asymptotic limit of $n \to \infty$. We start be analysing the fidelity $F_j(s_j)$ for blocks of large j. For each such block we define the ratios $x = m/j$, $m = -j, -j+1, \ldots, j$, that approach a continuous variable as $j \to \infty$ (See Appendix A.7 for details). In this limit, $\{\sqrt{j}\xi_m^j\}$ approaches a real function of x, $\sqrt{j}\xi_m^j \to \varphi(x)$, and the expectation value (6.11) becomes,

$$\begin{aligned} \sigma_j^2 &= \frac{1}{j^2} \min_{|\varphi\rangle} \int_{-1}^{1} dx \left\{ \left[\frac{d\varphi(x)}{dx} \right]^2 + V^j(x)\varphi(x)^2 \right\}, \\ :&= \frac{1}{j^2} \min_{|\varphi\rangle} \langle\varphi| \mathcal{H}^j |\varphi\rangle, \end{aligned} \tag{6.14}$$

where we have dropped some boundary terms that are irrelevant for this discussion, $\mathcal{H}^j := -d^2/dx^2 + V^j(x)$ plays the role of a 'Hamiltonian', with a 'potential'

$$V^j(x) = 2j^2(1 - a_m^j) = j \frac{1 - r^2}{2r\sqrt{1 - (1 - r^2)x^2}}. \tag{6.15}$$

Furthermore, in Eq. (6.14) the function $\varphi(x)$ must be also differentiable and must satisfy the conditions

$$\langle\varphi|\varphi\rangle = \int_{-1}^{1} dx \, [\varphi(x)]^2 = 1, \quad \varphi(x) \leq \frac{\tilde{\varphi}(x)}{\sqrt{s_j}}, \tag{6.16}$$

where for a given large j we define $\tilde{\varphi}(x)$ through

$$\sqrt{j\rho_{mm}^j} \to \tilde{\varphi}(x), \quad x = \frac{m}{j}. \tag{6.17}$$

It is now apparent from Eqs. (6.14) through (6.17) that our optimization problem is formally equivalent to that of finding the ground state wave-function of a quantum particle in a box ($-1 \leq x \leq 1$) for a potential $V^j(x)$ and subject to boundary conditions which are fixed by the probe state, the strength of the noise, and the success probability.

6.4 Multiple-copies

Although our methods apply to general symmetric probes, for the sake of concreteness we study in full detail the paradigmatic case of a probe consisting of n identical copies of equatorial qubits:

$$|\psi_{\text{cop}}\rangle = \frac{1}{\sqrt{2^n}}(|0\rangle + |1\rangle)^{\otimes n}. \tag{6.18}$$

Decoherence turns this symmetric pure state to a full rank state with a probability of having spin j given by

$$p_j \simeq \frac{\mathrm{e}^{-J\frac{(j/J-r)^2}{1-r^2}}}{\sqrt{\pi J(1-r^2)}}, \tag{6.19}$$

which is valid around its peak at the typical value $j_0 = rJ$. For each irreducible block and before filtering we have a signal

$$\sqrt{j\rho^j_{mm}} \to \tilde{\varphi}(x) \simeq \left(\frac{jr}{\pi}\right)^{\frac{1}{4}} \mathrm{e}^{-\frac{rj}{2}x^2} \tag{6.20}$$

that peaks at $x = 0$ with variance $\langle x^2\rangle = (2rj)^{-1}$.

For deterministic protocols ($S = 1$) the constraints completely fix the solution, $\varphi(x) = \tilde{\varphi}(x)$, and the precision is obtained by computing the 'mean energy' $\sigma_j^2 = \langle \mathcal{H}^j\rangle_{\tilde{\varphi}}/j^2$, Eq. (6.14). For large j it is meaningful to use the harmonic approximation $V^j(x) \simeq V_0^j + \omega_j^2 x^2$, where $V_0^j = j(1-r^2)/(2r)$ and $\omega_j^2 = j(1-r^2)^2/(4r)$. The leading contribution to σ_j^2 comes from the 'kinetic' energy [i.e., the first term in (6.14)]: $\langle p^2\rangle_{\tilde{\varphi}} = (1/4)\langle x^2\rangle^{-1} = jr/2$, whereas the harmonic term gives a sub-leading contribution. One easily obtains $\sigma_j^2 = (2jr)^{-1}$. The leading contribution to the precision of the deterministic protocol is given by σ_j^2 at the typical spin j_0: $\sigma^2_{\text{det}} = (2Jr^2)^{-1} = (nr^2)^{-1}$, in agreement with the previous known (point-wise) bounds (see Methods).

For unlimited abstention in a block of given spin j (s_j very small) the minimization in (6.14) is effectively unconstrained and the solution (the filtered state) is given by the ground state $\varphi^{\text{g}}(x)$ of the potential $V^j(x)$. Within the harmonic approximation, we notice that the effective frequency of the oscillator grows as $\sqrt{j}$, and the corresponding gaussian ground state is confined around $x = 0$ with variance $\langle x^2\rangle = (r/j)^{1/2}(1-r^2)^{-1}$. In this situation both the kinetic and harmonic contributions to the 'energy' are sub-leading —and so are the higher order corrections to $V^j(x)$. Thus, the precision σ_j^2 for a spin j is ultimately limited by the constant term V_0^j of the potential.

Up to sub-leading order one obtains $\sigma_j^2 = (1-r^2)(2jr)^{-1}[1+(r/j)^{1/2}]$. The filtering of $\tilde{\varphi}(x)$ to produce the gaussian ground state $\varphi^{\rm g}(x)$ succeeds with probability $s_j^* \sim \mathrm{e}^{-2j\log(1+r)}$. Note that in the absence of noise ($r=0$) the potential $V^j(x)$ vanishes and the ground state is solely confined by the bounding box $-1 \leq x \leq 1$. Then, $\varphi^{\rm g}(x) = \cos(\pi x/2)$, which results in a Heisenberg limited precision (the ultimate pure-state bound) $\sigma^2 = \pi^2/n^2$ Summy and Pegg [1990].

If the optimal filtering is performed on typical blocks, $j \approx j_0$, one obtains $\sigma^2 = (1-r^2)/(nr^2)$, which coincides with the ultimate deterministic bound found in Demkowicz-Dobrzański *et al.* [2012]; Knysh *et al.* [2014]. This shows that a probabilistic protocol performed on the uncorrelated multi-copy probe state $|\psi_{\rm cop}\rangle$ can attain the precision bound of a deterministic protocol that requires a highly entangled probe. This bound is attained for a critical success probability $S^* \simeq s_{j_0}^* \sim \mathrm{e}^{-nr\log(1+r)}$.

More interestingly, we can push the limit further by post-selecting on the block with highest spin (by choosing $f_m^j \propto \delta_{j,J}$) to obtain

$$\sigma_{\rm ult}^2 := \sigma_{j\approx J}^2 = \frac{1-r^2}{nr}\left(1+\sqrt{\frac{2r}{n}}\right), \tag{6.21}$$

with a critical probability given by $S^* = p_J s_J^* \sim \mathrm{e}^{-n\log 2}$, independently of the noise strength.

Having understood the two extreme cases, i.e., the deterministic ($S=1$) and unlimited-abstention protocols, we can quantify now the asymptotic scaling for an arbitrary success probability $\sigma^2(S)$. The so-called complementary slackness condition Boyd and Vandenberghe [2004], which follows from the Karush-Kuhn-Tucker constrains [second inequality in (6.16)] guarantees that, for a given value of s_j, the solution $\varphi(x)$ to (6.14) either saturates the above inequality —in a region called coincidence set— or it must take the value of an eigenfunction of the Hamiltonian $\mathcal{H}^j$ defined after Eq. (6.14). The continuity of $\varphi(x)$ and its derivative provide some matching conditions at the border of the coincidence set and a unique solution can be easily found.

As shown in Figure 6.2, in the case of multiple copies the tails of $\varphi(x)$ coincide with the gaussian profile in (6.20) scaled by the factor $s_j^{-1/2}$ for $|x| > x_c$ (in the coincidence set), while the filter takes an active part in reshaping the peak into the optimal profile (for $|x| < x_c$). Clearly, the wider the filtered region, the higher the precision and the abstention rate. A simple expression for the leading order can be obtained if we notice that with a finite abstention probability one can change the variance of the wave function in (6.20) but not its $1/j$ scaling. Hence, as for the deterministic case, only the kinetic energy and the constant term V_0^j of the potential play

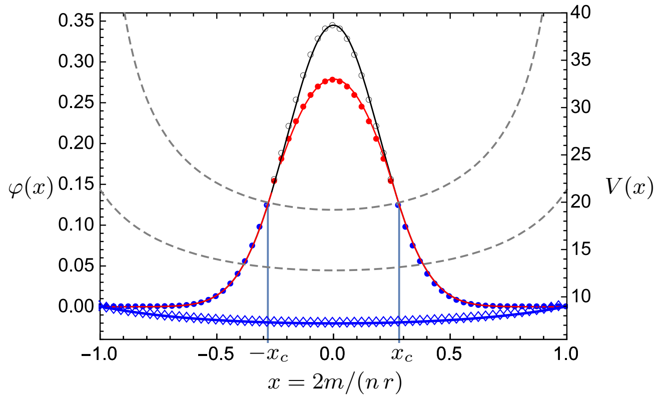


Figure 6.1. Computing the action of the probabilistic filter and its precision is formally equivalent to computing the the ground state and energy of a particle in a one-dimensional potential box. The state $\tilde{\varphi}(x)$ (empty circles) before the probabilistic filter and the state $\varphi(x)$ (solid circles) after the filter are represented together with the potential $V(x)$ (diamonds) corresponding to $j = nr/2$, see (6.15), for success probability $S = 0.75$, noise strength $r = 0.8$ and $n = 80$ probe copies. The unfiltered state (empty circles) has been rescaled so that it coincides with the filtered state in the region $|x| \geq x_c = 9/32$. The effective potential depends on the noise strength, as illustrated by the two additional dashed curves: for $r = 0.2$ (above) and $r = 0.6$ (below). Numerical (symbols) and analytical results (lines) are in full agreement.

a significant role. The solution can then be easily written in terms of the pure-state solution (See Chapter 4), which corresponds to a zero potential inside the box $-1 \leq x \leq 1$:

$$\sigma^2 \simeq \sigma^2_{j_0} = \frac{1-r^2}{nr^2} + r\sigma^2_{\text{pure}}(S) \approx \frac{1-(r^2/2)\bar{S}}{nr^2}, \tag{6.22}$$

where $\bar{S} := 1 - S$ is the probability of abstention, σ^2_{pure} is the precision for pure states ($r = 1$) and for an effective number of qubits $n_{\text{eff}} = 2j_0$. The pre-factor r takes into account the scaling of the variance of the state (6.20) as compared to the pure-state case. The first equality of (6.22) uses the fact that for finite S, only abstention on blocks about the typical spin j_0 is affordable. This also fixes the value of S to be approximately s_{j_0}. The simple expression given in the last term in (6.22) is not an exact bound, but does provide a good approximation for moderate values of $\bar{S}$ (see Figure 6.2). We notice that for low levels of noise ($r \approx 1$) one can have a considerable gain in precision already for finite abstention.

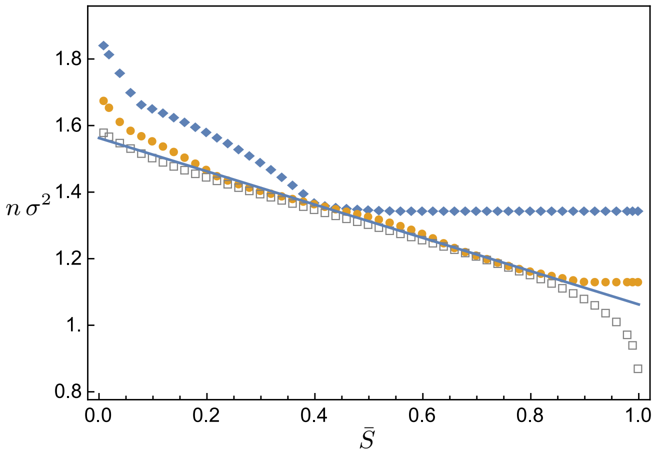


Figure 6.2. Numerical results for the rescaled precision $n\sigma^2$ for a noise strength of $r = 0.8$ as a function of the abstention probability $\bar{S} = 1 - S$ for various numbers of copies $n = \{6, 10, 20\}$ (diamonds, circles and squares). The critical success probability is clearly identified for the first ($n = 6$) curves at $\bar{S}^* = 0.46$. The solid line is the approximated analytical result (6.22).

6.5 Finite number of copies

Up to this point, we have given analytical results for asymptotically large n, the number of resources. In order to get exact values for finite n we need to resort on numerical analysis. The main observation here is that our optimization problem can be cast as a semidefinite program:

$$\sigma^2 = \min_{\Lambda : \mathcal{C}} \operatorname{tr} H\Lambda \tag{6.23}$$

subject to a set of linear conditions on the matrix Λ:

$$\mathcal{C} := \{\Lambda \geq 0,\ \operatorname{tr}\Lambda = 1,\ \Lambda^j_{mm} \leq p_j \rho^j_{m,m}/S\}, \tag{6.24}$$

where Λ, as well as H, have the block diagonal form: $\Lambda = \oplus \Lambda^j$ and $H = \oplus H^j$. Semidefinite programming problems, such as this, can be solved efficiently and with arbitrary precision Boyd and Vandenberghe [2004].

Figure 6.2 shows representative results for moderate (experimentally relevant) number n of qubits for the precision as a function of the abstention probability and noise strength $r = 0.8$. We observe that for small values of n the precision decreases quite rapidly until the critical S^* after which the precision cannot be improved. For larger n the initial gain is less dramatic, but the critical point (or plateau) is reached for higher abstention probabilities, hence allowing to reach lower precision rates. We see that for moderately large n, abstention can easily provide 60% improvement of the precision.

Figure 6.3 shows the scaling of the precision with the amount of resources (n) for low levels of noise $r = 95\%$ and for different values of the abstention probability $\bar{S}$. For low n all curves exhibit a similar (SQL n^{-1} scaling). Very soon the curve corresponding to unlimited abstention shows a big drop with a quantum-enhanced transient scaling: $n^{-(\alpha+1)}$, where $\alpha > 0$ depends on the noise strength. This curve saturates for very large $n \sim 500$ to the ultimate asymptotic limit (6.21), which is has again SQL scaling. The curves for finite S follow closely the optimal scaling up to the point where they meet the asymptotic curve (6.22). The larger the abstention probability, the later this transition happens. In addition, in the figure the ultimate scaling for $r = 99\%$ is shown to illustrate that for weaker noise levels the transient is more abrupt (larger α).

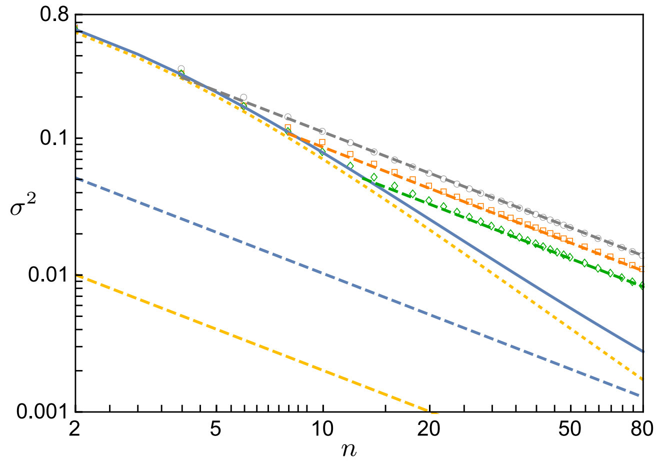


Figure 6.3. Ultimate precision scalings (for $n \to \infty$) are of fundamental interest, however from practical perspective understanding the transient behaviour equally important. The plot shows the precision σ^2 for different abstention probabilities $\bar{S} = \{0, 0.5, 0.9\}$ (from top to bottom) for $r = 0.95$. The fourth line (blue) shows the exact ultimate limit ($\bar{S}$ arbitrarily small). The dashed lines show the corresponding asymptotic limits given in Eqs. (6.22) and (6.21). In addition, for weaker noise strength $r = 0.99$ we show lines corresponding to the ultimate precision (in yellow).

6.6 Ultimate bound for metrology

So far we have studied the best precision bounds that can be attained for a fixed input state. A very relevant question of fundamental and practical interest is whether this bounds can be overcome by an appropriate choice of such state. We answer this question in the negative: the precision bound in

Eq. (6.21) is indeed the ultimate bound for metrology in the presence of local decoherence and can only be attained by a probabilistic strategy.

To this aim, we first notice that for any probe state and any measurement that attain a fidelity F with success probability S, we can find a new probe lying in the fully symmetric subspace ($j = J$) and a permutation invariant measurement that attain the very same fidelity with the very same success probability. This shows that the formulation that we have introduced, which deals with such probes, is actually completely general.

We now recall that the Hamiltonian $\mathcal{H}^j$ is independent of the choice of probe state and that such choice determines only the shape of the state $\tilde{\varphi}(x)$ before filtering, and the probability p_j of belonging to the subspace of spin j. Since the bound (6.21) is attained by the ground-state $\varphi^{\mathrm{g}}(x)$ of the potential $V^J(x)$, the choice of probe cannot further improve the precision, but only change the success probability. In particular one might increase S by choosing a probe state that gives rise to a profile $\tilde{\varphi}(x) = \varphi^g(x)$ for $j = J$, without any filtering within the block. In this case the critical success probability becomes $S^* = p_J = \mathrm{e}^{-n[\log 2 - \log(1+r)]}$ (see Appendix A.5), which is larger than that attained by $|\psi_{\mathrm{cop}}\rangle$.

At the other extreme, for deterministic strategies, the calculation of $\sigma^2_{\mathrm{opt}}(1)$ can be easily carried out performing first the sum over j and then optimizing over the $(n+1)$-dimensional probe state. In the continuum limit (for large n) such calculation can again be cast as a variational problem formally equivalent to that of finding the ground-state of particle in a box with the harmonic potential $V(y) = nr^{-2}(1-r^2)(1+y^2)$, $-1 \le y = m/J \le 1$. The corresponding ground state wave function and its energy provide the optimal probe state and precision respectively:

$$\psi_{\mathrm{op}}(y) = \left[\frac{n(1-r^2)}{(2\pi r)^2}\right]^{\frac{1}{8}} \mathrm{e}^{-\frac{\sqrt{n(1-r^2)}}{4r}y^2} \tag{6.25}$$

and

$$\sigma^2_{\mathrm{op}}(1) = \frac{1-r^2}{nr^2} + \frac{2\sqrt{1-r^2}}{n^{3/2}r}. \tag{6.26}$$

These results agree with their pointwise counterparts in Demkowicz-Dobrzański *et al.* [2012]; Knysh *et al.* [2014]. Quite surprisingly the presence of noise brings the pointwise and global approaches in agreement, as to both the attainable precision and the optimal probe state are concerned. This is in stark contrast with the noise-less case where the probe $\psi(y) = \cos(y\pi/2)$ is optimal for the global approach and gives $\sigma^2_{\mathrm{opt}} = \pi^2/n^2$, while the NOON-type state $|\psi\rangle = 2^{-1/2}(|J, J\rangle + |J, -J\rangle)$ provides the optimal point-wise precision $\sigma^2_{\mathrm{opt}} = 1/n^2$.

It remains an open question to find the optimal probe state given a finite values of $\bar{S}$. As argued above, a finite S will only be able to moderately reshape the profile without significantly changing the scaling of its width. Therefore we expect the optimal state to be fairly independent of the precise (finite) value of S, and hence very close to that obtained for the deterministic case ($S = 1$). Numerical evidence (optimizing simultaneously over probes and measurements) suggests that this is indeed the case provided S is not too small. With this we are lead to conjecture that the optimal probe state is given by

$$c_m^{\mathrm{opt}} \propto \cos\left(\frac{m\pi}{n+2}\right) \mathrm{e}^{-\sqrt{\frac{1-r^2}{r^2 n^3}} m^2} \tag{6.27}$$

independently of S (finite), which agrees with (6.25) for asymptotically large n. Note that the cosine prefactor guarantees that the solution converges to the optimal one for $r \to 1$ and it keeps the state confined in the box for all values of n and r. Such states continue to have a dominant typical value of $j = j_0$ and in those blocks both the kinetic and harmonic contributions to the energy are of sub-leading order. Hence, for probes of the form (6.27) the enhancement due to abstention is very limited, up until very high abstention probabilities where one can afford to post-select high spin states to reach the ultimate limit (6.21).

6.7 Scavenging information from discarded events

The aim of probabilistic metrology is twofold. First, it should estimate an unknown phase θ encoded in a quantum state with a precision that exceeds the bounds of the deterministic protocols. Second, it should assess the risk of failing to provide an estimate at all (e.g., it should provide the probability of success/abstention). Probabilistic metrology protocols are hence characterized by a precision versus probability of success trade-off curve $F(S)$ or, likewise $\sigma^2(S)$. As such, no attention is payed to the information on θ that might be available after an unfavorable outcome. Here, we wish to point out that one can attain $\sigma^2_{\mathrm{opt}}(S)$ and still be able to recover, or scavenge, a fairly good estimate from the discarded outcomes (see Fig. 6.4).

The optimal scavenging protocol can be easily characterized in terms of the stochastic map $\mathscr{F}$ in (6.9), which describes the state transformation after a favorable event, and that associated to the unfavorable events:

$$\bar{\mathscr{F}}(\rho_\theta) = \bar{\Phi}\rho_\theta\bar{\Phi}, \quad \bar{\Phi} = \sum_{j,m} (\bar{f}_m^j)^2 \, |j,m\rangle\langle j,m| \otimes \mathbb{1}_j, \tag{6.28}$$

where the weights $\bar{f}_m^j$ are defined through the equation $(\bar{f}_m^j)^2 = 1 - (f_m^j)^2$.

The addition of the two stochastic channels, $\bar{\mathscr{F}} + \mathscr{F}$, is trace-preserving, i.e., it describes a deterministic operation, with no post-selection. The final measurement is given by the seed $\tilde{\Omega}$ defined after (6.9) for both favorable and unfavorable events. Thus, we can easily compute the precision $\bar{\sigma}^2(S)$ for the the latter, as well as the precision $\sigma^2_{\rm all}(S)$ when all outcomes are considered. Clearly, we must have that $\sigma^2_{\rm all}(S) \geq \sigma^2_{\rm det}$ Combes *et al.* [2014], as $\sigma^2_{\rm det}$ refers to the optimal deterministic protocol.

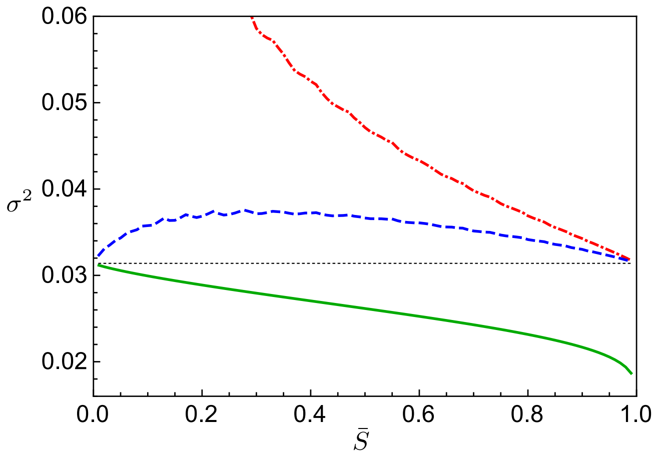


Figure 6.4. Precision σ^2 vs probability of abstention $\bar{S} := 1 - S$ from numerical optimization for $n = 50$ and $r = 0.8$. The green solid (red dash-dotted) correspond to $\sigma^2_{\rm opt}$ ($\bar{\sigma}^2$), where only the favorable (unfavorable) events are taken into account. The dashed curve $\sigma^2_{\rm all}$, where an estimate is provided on *all* outcomes, favorable or unfavorable. For low success probability ($\bar{S}$ close to unity), both $\bar{\sigma}^2$, and $\sigma^2_{\rm all}$, approach the precision of the deterministic protocol $\sigma^2_{\rm det}$ (dotted line).

As shown in Figure 6.4 a protocol that is optimized with some probability of abstention $\bar{S}$, performs slightly worse when forced to provide always a conclusive outcome. In particular we notice that if such protocol is designed to work at the ultimate limit regime, with precision $\sigma^2_{\rm ult}$, which requires a very large abstention probability ($S \to 0$) Combes *et al.* [2014], its performance coincides that of the optimal deterministic protocol. Actually, this observation follows (see Appendix A.9) from Winter's gentle measurement lemma (Lemma 9 in Winter [1999]), that states that a measurement with a highly unlikely outcome causes only a little disturbance to the measured quantum state. This is in contrast to the claims in Combes *et al.* [2014], where a random estimate is assigned to the discarded events.

6.8 Conclusions

We have given the ultimate limits in precision reachable by any (deterministic or stochastic) quantum metrology protocol in a realistic scenario with local decoherence for phase estimation. We have derived the optimal bounds that can be reached when a certain rate of abstention is allowed and hence provided a full assessment of the risks and benefits of the probabilistic strategy. This is summarised by the trade-off curve between the estimation precision and the success probability. We have elucidated the role of abstention as a filtering process selecting the total angular momentum followed by a modulating filter. It is important to remark that while the deterministic ultimate bound is ensured to be attained by the averaged estimate over a large number of repetitions of the measurement, the bound presented here is obtained without certainty. In fact, the probability of obtaining a favourable event can be vanishingly small. However, in this case the estimation accuracy about the parameter provided by the unfavourable events is essentially that of the deterministic protocol. Hence, when all events are taken into consideration the average precision of our protocol does not differ from the optimal deterministic one, but one has the possibility to obtain a much better precision estimate with some events.

CHAPTER 7

Equivalence between probabilistic estimation and super replication

The advantages of probabilistic filters are not limited to metrology. Instead, they affect a variety of tasks, including cloning [Duan and Guo, 1998; Fiurášek, 2004; Chiribella *et al.*, 2013] and amplification [Ralph *et al.*, 2009; Chiribella and Xie, 2013; Pandey *et al.*, 2013; Xiang *et al.*, 2010; Ferreyrol *et al.*, 2010; Usuga *et al.*, 2010; Zavatta *et al.*, 2011; Kocsis *et al.*, 2012]. Very recently, it has been shown that for quantum clocks the use of a filter can lead to the phenomenon of *super-replication* [Chiribella *et al.*, 2013], allowing to convert $n \gg 1$ synchronized clocks into $m \ll n^2$ replicas, whose joint state appears to be exponentially close to the ideal target of m perfect copies. Achieving such a replication rate with high fidelity is impossible without filtering, because a deterministic machine that produces more than $O(n)$ nearly perfect replicas would lead straight into a violation of the SQL.

Considering the striking difference in performance, it is natural to ask whether deterministic and probabilistic cloning machines differ in other, more fundamental features. The most fundamental feature of all is arguably the *asymptotic equivalence with state estimation* [Bruss *et al.*, 1997; Bae and Acín, 2006; Chiribella and D'Ariano, 2006; Chiribella, 2011], i. e. the fact that, in the macroscopic limit $m \to \infty$, the optimal performance of quantum cloning can be achieved by measuring the input copies and preparing the clones in a state that depends on the measurement outcome. The no-cloning theorem itself [Wootters and Zurek, 1982] can be considered as a particular instance of this equivalence: two states that can be cloned perfectly by a deterministic machine can also be cloned perfectly in the macroscopic limit and therefore they can be distinguished perfectly by a deterministic estimation strategy, which means that they must be orthogonal to one another. For deterministic machines, the cloning-estimation equivalence has been proved in full generality when the performance of cloning is assessed on small groups of $k \ll m$ clones [Chiribella and D'Ariano, 2006; Chiribella, 2011] and has been recently conjectured to hold even when the m clones are examined collectively [Yang and Chiribella, 2013; Chiribella and Yang, 2014]. However, nothing is known in the presence of postselection, where the tradeoff between performance and probability of success adds a new twist to the problem. Here, proving the equivalence requires showing that for every probabilistic cloning machine there is a PM protocol that, in the macroscopic limit, achieves the same fidelity with the same probability. But is the enhanced precision of PM sufficient to keep up with the highly-increased performance of probabilistic cloning machines?

We answer the question in the affirmative, showing that postselection does not challenge the fundamental equivalence between cloning and estimation. We first work out explicitly the example of quantum clocks, where the performance enhancements are the most prominent. We consider clock states $|\psi_t\rangle = e^{-itH}|\psi_0\rangle$ generated from an arbitrary inicial state $|\psi_0\rangle$ by time evolution with an arbitrary Hamiltonian H acting on a d-dimensional Hilbert space $\mathcal{H}$, and we exhibit PM protocols that achieve the asymptotic performance of the optimal cloning machine for every desired value of the success probability. In this comparison, we use the most restrictive criterion, namely the global fidelity between the clones and m perfectly synchronized replicas of the original clock. We evaluate the fidelity explicitly and discover that its value depends critically on the number of rationally independent eigenvalues of the Hamiltonian. The result is derived using new techniques, based on the Smith normal form [Marcus and Minc, 1964], which we expect to be useful for other problems in quantum metrology and optimal quantum information processing. Furthermore, we analyze the scenario where the performances of cloning are judged from groups of $k \ll m$ clones, establishing the equivalence between probabilistic cloning and estimation for arbitrary sets of input states and for arbitrary values of the success probability. This result extends the validity of the equivalence to all points of the optimal performance-probability tradeoff curve.

7.1 Quantum clocks

The state of a clock at time $t = 0$ can be written as $|\psi_0\rangle = \sum_{j=0}^{d-1} \sqrt{p_j}|j\rangle$, where $H|j\rangle = e_j|j\rangle$ and p_j is the probability that a measurement of energy gives outcome e_j. For n identically synchronized clocks, we denote the state at time t as $|\Psi_t^n\rangle := |\psi_t\rangle^{\otimes n}$. Without loss of generality, we assume that all probabilities $\{p_j\}_{j=0}^{d-1}$ are non-zero, that the eigenvalues $\{e_j\}_{j=0}^{d-1}$ are distinct, and that $e_0 = 0$. With these assumptions the values of the total energy can be labeled by the column vectors $\mathbf{n} = (n_1, \ldots, n_{d-1})^t$, satisfying $n_j \geq 0$ for every j and $\sum_{j=1}^{d-1} n_j \leq n$. Denoting by $\mathbf{e}$ the row vector $\mathbf{e} = (e_1, \ldots, e_{d-1})$ we express the corresponding energy as $E_{\mathbf{n}} := \sum_{j=1}^{d-1} e_j n_j := \mathbf{e}\,\mathbf{n}$. The set of the distinct such energies, i.e., the spectrum of the total Hamiltonian, will be denoted by Sp_n. We define $\mathscr{P}_n$ as the lattice of all vectors $\mathbf{n}$, and $\mathscr{P}_n^E$ as the set of the vectors that give the same energy E, i.e., $\mathscr{P}_n^E := \{\mathbf{n} \in \mathscr{P}_n : E_{\mathbf{n}} = E\}$, so that $|\mathscr{P}_n^E|$ is the degeneracy of E. Then, the state of the n clocks can be written as

$$|\Psi_t^n\rangle = \sum_{E\in\mathsf{Sp}_n} \mathrm{e}^{-iEt}\sqrt{p_{E,n}}|E,n\rangle, \quad p_{E,n} := \sum_{\mathbf{n}\in\mathscr{P}_n^E} p_{\mathbf{n},n},$$
$$|E,n\rangle := \frac{1}{\sqrt{p_{E,n}}} \sum_{\mathbf{n}\in\mathscr{P}_n^E} \sqrt{p_{\mathbf{n},n}}|\mathbf{n},n\rangle, \tag{7.1}$$

where the coefficients $p_{\mathbf{n},n}$ follow the multinomial distribution

$$p_{\mathbf{n},n} := n! \prod_{j=0}^{d-1} p_j^{n_j}/n_j!. \tag{7.2}$$

7.2 Optimal cloning

The aim of cloning is to produce m approximate copies of n ($m > n$) unknown synchronized clocks. This process if represented mathematically by a completely positive trace non-increasing map $\mathscr{C}_{n,m}$ that maps states on $\mathcal{H}^{\otimes n}$ to states on $\mathcal{H}^{\otimes m}$. The quantum operation $\mathscr{C}_{n,m}$ can always be viewed as a probabilistic filter Π followed by a trace preserving map $\mathscr{D}_{n,m}$ [Chiribella *et al.*, 2013]. Thus, the cloning process outputs a state

$$\rho_{\text{out}}^t := P_{\text{succ}}^{-1}\, \mathscr{D}_{n,m}(\Pi^{1/2}\, |\Psi_t^n\rangle\langle\Psi_t^n|\, \Pi^{1/2}) \tag{7.3}$$

with success probability $P_{\text{succ}} := \mathrm{tr}\,[\mathscr{C}_{n,m}(|\Psi_t^n\rangle\langle\Psi_t^n|)] = \|\Pi^{1/2}\, |\Psi_t^n\rangle\,\|^2$. The performance is quantified by the worst-case fidelity

$$F_{\text{CL}} = \inf_{t\in\mathbb{R}} \langle\Psi_t^m|\, \rho_{\text{out}}^t\, |\Psi_t^m\rangle\,. \tag{7.4}$$

The symmetry of the problem enables us to choose a filter of the form

$$\Pi = \sum_E \pi_E |E, n\rangle\langle E, n| \tag{7.5}$$

without loss of generality. The maximum fidelity achievable with a given filter can be upper bounded as (see Appendix B)

$$F_{\mathrm{CL}}^{\Pi} \le \max_{E\in\mathsf{Sp}_m}\{p_{E,m}\} \left(\sum_{E\in\mathsf{Sp}_n} \sqrt{p_{E,n}\frac{\pi_E}{P_{\mathrm{succ}}}}\right)^2 . \tag{7.6}$$

7.3 Cloning by means of probabilistic metrology

Let us now move to the PM scenario. Here, we simulate cloning by first obtaining an estimate of the time t, denoted by $\hat{t}$. To this end, a measurement is performed on $|\Psi_t^n\rangle$. If a low quality estimate is output, we allow the process to terminate and report fail. In case of success, the estimate is then used to prepare a guess state $|\hat{\Psi}_{\hat{t}}^m\rangle = \sum_{E\in\mathsf{Sp}_m} e^{-iE\hat{t}}\sqrt{\hat{p}_{E,m}}|E, m\rangle$ that in average over the estimates $\hat{t}$ resembles m ideal clones. The corresponding fidelity is

$$F_{\mathrm{PM}} = \inf_{t\in\mathbb{R}} \int \mathrm{d}\hat{t}\, p(\hat{t}|t, \mathrm{succ}) \left|\langle\hat{\Psi}_{\hat{t}}^m|\Psi_t^m\rangle\right|^2 , \tag{7.7}$$

where $p(\hat{t}|t, \mathrm{succ})$ is the probability of obtaining $\hat{t}$ when the true value is t conditional to successfully passing the PM stage of the process. As in the case of $\mathscr{C}_{n,m}$ above, the PM measurement can be split into a probabilistic filter $\Pi = \sum_{E\in\mathsf{Sp}_n} \pi_E|E, n\rangle\langle E, n|$, which succeeds with probability $P_{\mathrm{succ}} = \langle\Psi_t^n|\Pi|\Psi_t^n\rangle$ independent of t, followed by a deterministic measurement, as we have seen in the previous chapters. For given $\hat{p}_{E,m}$ and π_E, the supremum of the conditional fidelity over all measurements is (see Appendix B)

$$F_{\mathrm{PM}}^{\Pi} = \sum_{\mathcal{E}} \left({\sum_{E\in\mathsf{Sp}_n}}' \sqrt{p_{E,n}\, p_{E+\mathcal{E},m}\, \hat{p}_{E+\mathcal{E},m}\, \frac{\pi_E}{P_{\mathrm{succ}}}}\right)^2 , \tag{7.8}$$

where the outer sum runs over the set of energy differences, $\{\mathcal{E} = E_m - E_n \mid E_m \in \mathsf{Sp}_m,\ E_n \in \mathsf{Sp}_n\}$, and the prime ($'$) means that the sum is restricted to those $E \in \mathsf{Sp}_n$ such that $E + \mathcal{E} \in \mathsf{Sp}_m$. Since this protocol can be seen as a particular instance of cloning, and therefore $F_{\mathrm{PM}}^{\Pi} \le F_{\mathrm{CL}}^{\Pi}$, it will be enough to show that the fidelity in Eq. (7.8) achieves the upper bound (7.6) in the limit of large m. For the sake of illustration, we will focus first on the simpler case of Hamiltonian with commensurable eigenvalues and then we will treat the general case.

7.4 Commensurable energies

In this case all the energies can be written as $e_j = k_j\,\varepsilon$, where ε is a fixed unit of energy and $k_1, k_2, \ldots, k_{d-1}$ are integers. Then, the energies of m clocks are given by the linear combination of integers: $E_{\mathbf{m}}/\varepsilon = \sum_j k_j m_j$. It follows from the generalised Bezout's identity [Jones and Jones, 1998] that if the variables m_j take all integer values, the set of numbers given by such linear combination has a minimal spacing $k^* = \gcd\{k_j\}$ and, furthermore, the set coincides with the multiples of k^*, i.e., they form an infinite regular array. Note, however, that the integers m_j are constrained to be positive and must add up to less than m. This turns the infinite array into a finite one and may introduce some defects within a finite distance from its end points (see Appendix B). Note also that the degeneracy of the points in the array, namely, the degeneracy of the spectrum of m clocks scales as $|\mathscr{P}^E_m| \sim m^{d-2}$, since one condition ($E_{\mathbf{m}} = E$) is imposed on the $d-1$ variables m_j.

Let us now assume that m is asymptotically large. In this limit, the multinomial distribution $p_{\mathbf{m},m}$ approaches the multivariate normal distribution $\mathscr{N}(m\mathbf{p}, \mathbf{\Sigma})$, where $\mathbf{p} := (p_1, \ldots p_{d-1})^t$ and the covariance matrix $\mathbf{\Sigma}$ has entries $\Sigma_{jj} = mp_j(1-p_j)$, $\Sigma_{jl} = -mp_jp_l$, $j \neq l$. Then, $p_{\mathbf{m},m}$ is concentrated in an ellipsoidal region centered at $m\mathbf{p}$ of (linear) size $O(\sqrt{m})$, much smaller than that of $\mathscr{P}_m$, which is $O(m)$. When computing $p_{E,m}$, as in Eq. (7.1), we may thus allow the variables m_j to take all integer values (incurring an error that vanishes exponentially with m). Hence, as discussed above, the typical energy E varies in constant unit steps of $\Delta E^* = \varepsilon k^*$. Since the degeneracy of the spectrum scales as m^{d-2}, the sum over $\mathbf{m} \in \mathscr{P}^E_m$ that defines $p_{E,m}$ can be approximated by the integral of $\mathscr{N}(m\mathbf{p}, \mathbf{\Sigma})$ over a $(d-2)$-dimensional domain. As a result, $p_{E,m}$ is approximated by the discrete Gaussian distribution

$$p_{E,m} \approx \Delta E^* \frac{\mathrm{e}^{-\frac{(E-m\langle H\rangle)^2}{2m\mathrm{Var}(H)}}}{\sqrt{2\pi m \mathrm{Var}(H)}}, \quad \Delta E^* = \varepsilon \gcd\{k_j\}, \tag{7.9}$$

where $\mathrm{Var}(H) = \langle H^2\rangle - \langle H\rangle^2$ is the variance of H.

Thanks to Eq. (7.9) we are now in position to show that F^{Π}_{PM} approaches F^{Π}_{CL} in the macroscopic limit. Indeed, since for $m \gg n$ the probability $p_{E,m}$ is almost constant over every interval of size $O(n)$, we can pull out a factor $\max_{E\in\mathsf{Sp}_m}\{p_{E,m}\}$ from both sums in Eq. (7.8) introducing an error that vanishes exponentially in the asymptotic limit. Moreover, we can choose a guess state with $\hat{p}_{E,m}$ given by a discrete Gaussian with a width that scales as $\sqrt{m^{1-\eta}}$, $0 < \eta < 1$. This state, which may not be optimal, suffices to saturate the asymptotic expression for the fidelity, as it would any other state for which $\hat{p}_{E,m}$ is peaked and has similar width. Since the width is much larger than $O(n)$, we have $\hat{p}_{E+\mathcal{E},m} \approx \hat{p}_{\mathcal{E},m}$ for every $E \in \mathsf{Sp}_n$ and we can pull out the term $\hat{p}_{\mathcal{E},m}$ from the sum over Sp_n. Now, since the width is smaller than $\sqrt{m}$, we have $\sum_{\mathcal{E}} p_{\mathcal{E},m} \approx 1$. Hence, our choice of guess state attains the upper bound in Eq. (7.6) and we have

$$F^{\Pi}_{\mathrm{PM}} \approx F^{\Pi}_{\mathrm{CL}} \approx \frac{\Delta E^*}{\sqrt{2\pi m \mathrm{Var}(H)}} \left(\sum_{E \in \mathsf{Sp}_n} \sqrt{p_{E,n} \frac{\pi_E}{P_{\mathrm{succ}}}} \right)^2 . \tag{7.10}$$

This proof is made rigorous in Appendix B.

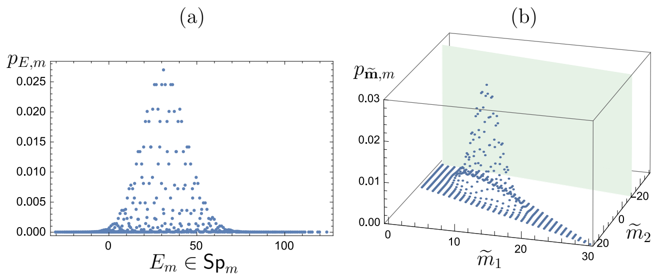


Figure 7.1. Probability mass functions $p_{E,m}$ (a) and $p_{\widetilde{\mathbf{m}},m}$ (b) for $m = 30$ quantum clocks with energies $\mathbf{e} = (0, -1, 1+\pi)$ and initial state given by $\mathbf{p} = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})^t$. In (b), $\widetilde{m}_1 = m_3$, $\widetilde{m}_2 = m_3 - m_2$, which corresponds to the choice of energy units $\boldsymbol{\varepsilon} = (\pi, 1)$. Note that $p_{E,m}$ in (a) resembles neither a Gaussian (only its outline does) nor any continuous function. Instead, it is the projection of the bell-shaped distribution of points in (b), which approach a smooth surface, onto the gray plane.

7.5 General Hamiltonians

The reader should not be misled by the simplicity of Eq. (7.9), which superficially may seem an application of the Central Limit Theorem (CLT). The CLT gives an approximation of the *cumulative* distribution of $p_{E,m}$, not of the probability mass function itself. In fact, those who believe that $p_{E,m}$ should converge to a Gaussian are in for a surprise in the case where the eigenvalues of H are not commensurable [see Fig. 7.1(a)]. In the general case, the energies $\{e_j\}$ can be expressed as integer linear combinations of a minimal number r of rationally independent units of energy $\{\varepsilon_l\}_{l=1}^r$. Rational independence means that, for every set of integer coefficients $\{c_l\}$, the relation $\sum_l c_l \varepsilon_l = 0$ implies $c_l = 0$ for every l. In terms of the units $\{\varepsilon_l\}$, we expand each energy as $e_j = \sum_l \varepsilon_l k_{lj}$ where $\{k_{lj}\}$ are integer coefficients, uniquely defined thanks to the rational independence of the units. Using the decomposition, we express the energy of m clocks as $E_{\mathbf{m}} = \boldsymbol{\varepsilon}\,\widetilde{\mathbf{m}}$ where $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_r)$, and $\widetilde{\mathbf{m}} = (\widetilde{m}_1, \ldots, \widetilde{m}_r)^t$ is the column vector with components $\widetilde{m}_l = \sum_j k_{lj} m_j$, or, more compactly, $\widetilde{\mathbf{m}} = \mathbf{K}\,\mathbf{m}$, where $\mathbf{K}$ is the $r \times (d-1)$ matrix with entries $\{k_{lj}\}$.

The matrix $\mathbf{K}$ maps the lattice $\mathscr{P}_m \subset \mathbb{Z}^{d-1}$ into a new lattice $\widetilde{\mathscr{P}}_m \subset \mathbb{Z}^r$ which, as illustrated in Fig. 1(b), conform a confined region of a Bravais lattice (again, near the boundaries some defects might appear). Due to the rational independence of the units, the points $\widetilde{\mathbf{m}}$ have the special feature that they are in one-to-one correspondence with the energies in Sp_m. Hence, instead of the ill-behaved probability distribution $p_{E,m}$ we can consider the much more manageable one $p_{\widetilde{\mathbf{m}},m} := \sum_{\mathbf{m}:\mathbf{K}\mathbf{m}=\widetilde{\mathbf{m}}} p_{\mathbf{m},m}$. Indeed, as mentioned above, for m large $p_{\mathbf{m},m}$ is a multivariate normal distribution of mean $\mathbf{m}_0 = m\mathbf{p}$ and covariance matrix $\boldsymbol{\Sigma}$ and thus so is its marginal $p_{\widetilde{\mathbf{m}},m}$, with mean $\widetilde{\mathbf{m}}_0 = \mathbf{K}\mathbf{m}_0$ and covariance matrix $\widetilde{\boldsymbol{\Sigma}} = \mathbf{K}\boldsymbol{\Sigma}\mathbf{K}^t$ (see Appendix B):

$$p_{\widetilde{\mathbf{m}},m} \approx \Delta V^* \frac{\exp\left\{-\dfrac{(\widetilde{\mathbf{m}}-\widetilde{\mathbf{m}}_0)^t \widetilde{\boldsymbol{\Sigma}}^{-1} (\widetilde{\mathbf{m}}-\widetilde{\mathbf{m}}_0)}{2}\right\}}{(2\pi)^{r/2}\sqrt{\det \widetilde{\boldsymbol{\Sigma}}}}, \tag{7.11}$$

where ΔV^* is the volume of a minimal cell of the Bravais lattice $\widetilde{\mathscr{P}}_m$. Using the Smith normal form of $\mathbf{K}$, a higher dimensional analog of Bezout's identity, one can show that $\Delta V^* = \gcd\{[\mathbf{K}]_r\}$, where $\{[\mathbf{K}]_r\}$ is the set of all minors of $\mathbf{K}$ of order r (see Appendix B). Eq. (7.11) has two major consequences: First, the probability mass function $p_{E,m}$ *does not converge to a Gaussian*, as one would naively expect from a misapplication of the CLT. Instead, it converges to the non-continuous function $p_{\widetilde{\mathbf{m}}(E),m}$, where $\widetilde{\mathbf{m}}(E)$ is the value of $\widetilde{\mathbf{m}}$ such that $E = E_{\widetilde{\mathbf{m}}}$. Second, all the steps of the proof for commensurable energies can now be reproduced by replacing the energy $E \in \mathsf{Sp}_n$ with the corresponding vector $\tilde{\mathbf{n}}(E)$. In this way we obtain $F^{\Pi}_{\mathrm{PM}} \approx F^{\Pi}_{\mathrm{CL}} \approx F^{\Pi}$, with

$$F^{\Pi} := \frac{\Delta V^*}{\sqrt{(2\pi)^r \det \widetilde{\boldsymbol{\Sigma}}}} \left(\sum_{\widetilde{\mathbf{n}} \in \widetilde{\mathscr{P}}_n} \sqrt{p_{\widetilde{\mathbf{n}},n} \frac{\pi_{\widetilde{\mathbf{n}}}}{P_{\mathrm{succ}}}} \right)^2. \tag{7.12}$$

Note that, since $\det \widetilde{\boldsymbol{\Sigma}}$ scales as m^r, the fidelity scales as $m^{-r/2}$. Hence, the asymptotic behaviour depends dramatically on the number of rationally independent units. Quite remarkably, this means that the fidelity is not continuos in the Hamiltonian: Although one can approximate arbitrarily well the Hamiltonian H with another Hamiltonian H' that has commensurable energies, the corresponding fidelities are not going to be close. The origin of the discontinuity is that the value of the fidelity depends on the *closure* of the set of clock states, due to the infimum in Eq. (7.7). When the energy eigenvalues are commensurable, the time evolution is periodic and the orbit $\{|\Psi_t^n\rangle,\ t \in \mathbb{R}\}$ is a close one-dimensional curve. But when the energies are combinations of r rationally independent units, the orbit is dense in an r-dimensional submanifold of the manifold of pure states. This phenomenon is the quantum analog of a classic feature of integrable Hamiltonian systems [Katok and Hasselblatt, 1996], where rationally independent frequencies lead to ergodic time evolutions in phase space. Here the fidelity is discontinuous because it depends on the long-time behaviour of the time evolution, during which the quantum clock can probe a higher dimensional manifold.

Having proven the asymptotic equivalence between probabilistic metrology and cloning of quantum clocks, we now give closed expressions for optimal fidelity, maximized over all possible filters. We focus on the case with large enough n to take the continuum limit of $p_{\widetilde{\mathbf{n}},n}$, but under the condition $n \ll \sqrt{m}$ in order to use the derived assymptotic expression in Eq. (7.6). We will consider two relevant regimes: First, we allow arbitrarily low probability of success, showing that the ultimate fidelity is (see Appendix B)[1]

$$F = \left[(2\pi)^r \det \widetilde{\boldsymbol{\Sigma}}\right]^{-1/2} |\widetilde{\mathscr{P}}_n|\, \Delta V^*, \tag{7.13}$$

where $|\widetilde{\mathscr{P}}_n|$ is the number of sites in the lattice $\widetilde{\mathscr{P}}_n$. Since $|\widetilde{\mathscr{P}}_n|$ scales as n^r, the fidelity scales as $F \sim (n/\sqrt{m})^r$. Second, we consider the case where the probability of success is high, i.e. $P_{\text{succ}} = 1 - \eta$ for some small η. Here the optimal fidelity acquires the particularly simple form

$$F = \left(4\frac{n}{m}\right)^{r/2} [1 + \eta(1 - 2^{-r/2})] + O(\eta^2)\,. \tag{7.14}$$

Quite surprisingly, F does not depend on the coefficients $\{p_j\}$ of the input state, but only on the number of rationally independent units r.

7.6 The fidelity of a subset of clones

We conclude by discussing the equivalence between probabilistic metrology and cloning for arbitrary sets of states. Here we assess the performance of cloning by looking at a random subset of $k \ll m$ clones, evaluating the global fidelity between the state of the k clones and the state of k ideal

[1] This expression is also correct for small n. For large n, we can replace $|\widetilde{\mathscr{P}}_n|\Delta V^*$ by V_n, the volume of the minimal polytope that contains the lattice $\widetilde{\mathscr{P}}_n$.

copies. Clearly, since the k clones are picked at random, one can assume that the optimal cloner $\mathscr{C}_{n,m}$ is invariant under permutation of the m output systems. Using a de Finetti-type argument (see Appendix B), we then prove a very general result: for every quantum operation $\mathscr{C}_m$ with permutationally invariant output there exists a PM protocol, described by a quantum operation $\widetilde{\mathscr{C}_m}$, such that *i)* $\mathscr{C}_m$ and $\widetilde{\mathscr{C}_m}$ have the same success probability and *ii)* the error probability in distinguishing between $\mathscr{C}_m$ and $\widetilde{\mathscr{C}_m}$ by inputting a state ρ and measuring k output systems is lower bounded by $p_{\mathrm{err}} \leq \frac{1}{2} + (kd^2)/[2mP_{\mathrm{succ}}(\rho)]$, where $P_{\mathrm{succ}}(\rho)$ is the probability that the operation $\mathscr{C}_m$ takes place on input ρ. This result can be applied directly to probabilistic cloning: in this case, it implies that the k-copy fidelity of an arbitrary n-to-m cloner on a generic input state $|\psi_x\rangle^{\otimes n}$ can be achieved by PM, up to an error of size $k/[mP_{\mathrm{succ}}(|\psi_x\rangle\langle\psi_x|^{\otimes n})]$ (see Appendix B). Hence, the error vanishes as k/m for every process with success probability larger than a given finite value for every possible input. This result extends the equivalence between cloning and estimation to every point of the tradeoff curve between fidelity and success probability. Furthermore, one can even remove the requirement that success probability is larger than a finite value: no matter how small is the probability of a cloning process, asymptotically the process must be equivalent to a PM protocol.

7.7 Conclusions

In conclusion, we proved that probabilistic protocols empowered by postselection do not challenge the fundamental equivalence between cloning and estimation. We worked out explicitly the case of quantum clocks, where the performance enhancements for both tasks are most dramatic, and developed a technique to evaluate the optimal asymptotic fidelity. We found out that the asymptotic fidelity depends critically on the number of rationally independent units generating the spectrum of the Hamiltonian, due to an effect that is analog to ergodicity of classical dynamical systems. Finally, we discussed the case of arbitrary families of states, establishing an equivalence between probabilistic metrology and cloning when the performance is quantified by the fidelity of $k \ll m$ randomly chosen clones.

Conclusions and outlook

In this thesis we have set up a general framework for probabilistic quantum metrology tasks. Up until now, the performance of measurement schemes has been mostly quantified using the mean square error (MSE): that is the squared error $(\phi - \hat{\phi})^2$ averaged over all possible values of the unknown parameter and outcomes. The paradigmatic shot-noise or Heisenberg scaling refers to the behaviour of the MSE for an asymptotically large number of resource systems. A particular instance of metrology experiment with a fixed amount of resources produces necessarily a single outcome. The experiment should provide an estimated value of the unknown parameter but we judge equally important to provide a precision measure that quantifies the confidence that we should place to our estimate. To quantify the precision of a single outcome we use the square error averaged over the possible values of the unknown parameter and show that with a proper choice of measurement this can be much lower than the established (deterministic) bounds on the mean square error. Such ultra-precise outcomes can only occur at the expense of producing other outcomes with a below-average precision. The simple observation that some outcomes can beat the established metrology limits is the cornerstone of this thesis. We have have found that the precision of a single outcome defeats the deterministic bound but it is still limited at a fundamental level thereby establishing the ultimate precision bound for any conceivable experiment. In order to characterize and optimize measurement schemes whose outcomes have uneven precision, we have optimized the average precision of a set of outcomes that occur with a given probability, i.e. we have found the optimal probabilistic protocol that provides the optimal trade-off curve between success probability and precision.

The starting point to analyse the probabilistic protocols is to note that they can be regarded as two consecutive steps: firstly, the system of interest goes through a stochastic channel that modifies its state, and then a deterministic measurement is performed. In order to produce a precision enhancement, some of the outcomes of the channel effectively transforms the system state into a more parameter-sensitive one. In this sense, the inconclusive outcome is associated to those outcomes of the channel that worsen the states in terms of estimation sensitivity. This treatment is not only useful for the interpretation of the problem but also helps to circumvent some technical difficulties of it as the previous results in deterministic metrology can be applied directly into the measurement design leaving only the channel to be optimised.

In order to define a precision measure for every measurement outcome we have chosen to follow a Bayesian approach to metrology. In fact, the perhaps more ubiquitous Frequentist approach, which is readily analysed by means of the quantum Cramér-Rao bound, is not suitable for this end. Solving estimation problems within the Bayesian approach is in general hard. However, in this thesis we have focussed on a set of very relevant metrology problems that enjoy a symmetry that significantly simplifies the problem and allows to apply powerful techniques in representation theory. In particular, the optimal measurement can be seen to be invariant under this symmetry, i.e. is fully characterized by a few invariant parameters. In addition the symmetry of the initial probe state will also be inherited by the POVM and will further reduce the number of invariants.

The cases of direction and reference frame estimation enjoy a fairly large symmetry (Chapter 4). The only invariant in these case is the total angular momentum j, and hence the optimal filter is fully characterized by $N/2+1$ (for N even) invariant parameters f_j (all other degrees of freedom do not effect the direction encoding). This immediately shows, that if the probes state lies in the fully symmetric subspace (maximal $j = N/2$), as in the case of N parallel pure spins, the action of the filter reduces to a single multiplying factor that obviously cannot produce and enhancement in the precision. Hence, probabilistic metrology can improve the deterministic bound if the probe state has support in several invariant subspaces. This happens for example when the direction is encoded in pairs of antiparallel spins, which involves the superposition of terms with different j. The filter then acts on these coherences, effectively modulating the amplitudes to give a state that encodes better the direction. If the filter fails than the opposite effect happens: the probe state is mapped into a state with worst performance. This typically results in an average precision (over *all* outcomes) strictly worse then the optimal deterministic one.

The case of noisy parallel spins studied in Chapter 5, also has support on various j, but it does not show any coherences between different j's (it is block diagonal in j). In that case the filter acts as a Quantum Non-Demolition (QND) measurement that essentially post-selects blocks of high angular momentum, which encode better the direction than lower-lying blocks. Here we note that the action of the filter does not alter state at all, and hence the average precision (over *all* outcomes) attains the deterministic bound. Although this measurement does not change the average precision if all the outcomes are considered, it gives crucial information that provides a more accurate assessment of the precision of the obtained outcome because information on the unknown parameter can be more reliably read from high angular momentum subspaces. In other words, if an experimentalist performs a metrology experiment providing an estimate of the direction $\hat{n}$ and the QND measurement returns a high value of j, it would be very inaccurate (and unfair) to put error bars to his estimate based on average precision instead of that corresponding to the outcome that has actually occurred.

For the case of phase estimation, which occupies a great part of this thesis, the symmetry group is smaller and there are $N+1$ (for N even) invariant subspaces labelled by the magnetic quantum number m. Furthermore, for permutationally invariant probe state, these subspaces can further be broken down into smaller blocks of fixed angular momentum j. Hence, the action of the probabilistic channel can be understood in two steps: 1) A quantum non-demolishing (QND) measurement of the total angular momentum j; 2) A modulating filter that reshapes the coherences ($\rho^j_{m,m'}$) in each representation. As in the discussion above, the role of the QND measurement is to extract information in a non-invasive way that enables an accurate assessment of the error; while the role of the modulating filter is to disturb the state (in an irreversible fashion) in order to enhance the read-out of the encoded state.

We have developed a theory to study the optimal strategies in the asymptotic regime of a large number of qubits N. In this regime evaluating the precision provided by a state amounts to the 1-D quantum mechanical problem of computing the energy of a particle in a box, where the wavefunction is fixed by the state of the N qubits. The presence of noise adds a potential well to the picture whose shape depends on the amount of noise. Then the optimal filter is given by the linear modulation map that reshapes the initial wavefunction to one that minimizes the energy. This formalism has allowed us to compute analytically the trade-off curve between the obtained precision in terms of the success probability of the filter, together with the corresponding optimal POVM. This method is very useful also in deriving results for the deterministic scenario, which typically require quite involved calculations. In addition, in the presence of noise, the fact that the mean energy of the wave-function is always lowerbounded by the minimum of the potential immediately bounds the precision to be shot-noise limited.

In general the stronger the effect of the filter the smaller is the probability of applying it successfully. For this reason, except in some pathological cases (like the phase-state in 4.3.2), in order to produce ultra-sensitive post-filter states requires maps that succeed with a quite small probability. That is ultra-precise outcomes occur occasionally. Nevertheless it is important to emphasize that, if such an event does not take place the system will remain essentially the same as before. That is, one can apply our probabilistic protocols with just a little deviation from the deterministic bounds, but with the addition that there are some lucky heralded ultra-precise estimates. In this sense, seeking ultra-sensitive measurements is a low-risk endeavour. In between, the trade-of curve gives us a smooth transition between the occasional ultra-sensitive scheme and the deterministic scheme with average precision.

Finally let us note although the asymptotic limit is very useful to characterize the performance of different protocols and provides the ultimate bounds in precision, it can conceal the true effectiveness of the probabilistic protocols in realistic scenarios (with finite N). With finite-sized samples the stochastic channel can have a wider effect. To address this of situation we have managed to cast the problem as semidefinite programming formulation, which for all practical proposes amounts to solving the problem exactly (for values of $N \sim 100$). As in the deterministic case, we have observed that there exist a transient for small number of qubits where the noise effects are set back, and one recovers the pure-state behaviour. This includes the probabilistic drastic improvement of precision from shot-noise to Heisenberg scaling for this moderate values of N, which is lost after a critical sample size is reached and the improvement is only reflected in a pre-factor, not in the scaling. Finding this critical value is a timely question, that needs to be addressed in future work, both in deterministic case and as a function of the success probability.

Although we have developed a fairly complete framework for probabilistic protocols, there are yet generalisations worth studying. The most obvious one is to consider different kinds of noise. Recently Knysh et al [Knysh *et al.*, 2014] have used the Cramer-Rao bound to derive optimal *deterministic* bounds on precision in a variety of settings, including and extending our noise sources. Surprisingly, their methods lead also to the same 1-D quantum mechanical analogy, with the same potential. This coincidence is reinforced by the conjecture that in shot-noise limited scenarios the Bayesian and Frequentist bounds for the average precision agree, which has been recently proven [Jarzyna and Demkowicz-Dobrzański, 2015]. Hence, chances are that the their results on more general noise sources can be translated to our setting again with the same effective potential. There may be some kinds of noise for which the probabilistic protocols can counterbalance its blurring effects with even more efficiency.

An avenue worth exploring would be extend the study of probabilistic schemes in the estimation of parameters that do not have a strong symmetry, as for example a loss rate in a communication channel. Furthermore, in this thesis we have restricted our analysis to systems with finite dimension, however, the framework presented here can be also applied to continuous variables systems like quantum states of light. Another very relevant question is to optimise probabilistic strategies restricted to a given constrained experimental toolbox (e.g. LOCC, gaussian operations and photon detection...). It would also be interesting to extend this study to different figures of merit, e.g. a utility function that counts only those outcomes that fall into some confidence interval an discard the rest. However, I expect that the results found with it will not differ a lot from the ones presented in this thesis.

To complete the analysis of probabilistic quantum parameter estimation, we have proved that probabilistic protocols empowered by postselection do not put at risk the fundamental equivalence between cloning and estimation. Particularly, we have compared the performance of our probabilistic estimation protocol with that of super-replication protocol presented in Chiribella *et al.* [2013] for the case of quantum clocks, where the performance enhancements for both tasks are most dramatic. We have developed a technique to evaluate the optimal asymptotic fidelity, even for Hamiltonians with rationally independent energies. We found out that the asymptotic fidelity depends critically on the number of such rationally independent units, due to an effect that is analogue to ergodicity of classical dynamical systems. Moreover, one can observe that the rationally dependent energies act as if they were independent below some number of initial copies. Again it is important to study the onset of the asymptotic regime. For finite number of copies N the system behaves as if all energies where incommensurate. As N increases he commensurability of different energies reveals itself and the fidelity goes through different scaling regimes, until a final critical point where all incommensurability is exposed and the scaling exponent remains constant. It would be interesting to relate the location of the critical points to existing mathematical notions for the degree of incommensurability.

All in all, this work opens the field of metrology to the use of stochastic maps and single-shot inference. These have been used with a wide-spread acceptance in many quantum information processing protocols, and we believe that it deserves attention in the field of quantum metrology, specially in a moment where the promises of quantum enhanced metrology have been challenged.

APPENDIX A

Technical details of Chapter 6

A.1 Notation

Throughout these Appendices we use the following notation. The n-qubit computational basis is denoted by $\{|b\rangle\}_{b=0}^{2^n-1}$, where $b = b_1 b_2 \cdots b_n$ is a binary sequence, i.e., $b_i = 0, 1$ for $i = 1, 2, \ldots, n$. We denote by $|b|$ the sum of the n digits of b, i.e., $|b| := \sum_{j=1}^{n} b_j$. The digit-wise sum of b and b' modulo 2 will be simply denoted by $b + b'$, hence $|b + b'|$ can be understood as the Hamming distance between b and b', both viewed as binary vectors.

The permutations of n objects, i.e., the elements of the symmetric group S_n, are denoted by π. We define the action of a permutation π over a binary list b as $\pi(b) := b_{\pi(1)} b_{\pi(2)} \cdots b_{\pi(n)}$. This induces a unitary representation of the symmetric group on the Hilbert space $\mathcal{H}^{\otimes n}$ of the n qubits through the definition $U_\pi |b\rangle := |\pi(b)\rangle$. The (fully) symmetric subspace of $\mathcal{H}^{\otimes n}$, which we denote by $\mathcal{H}_+^{\otimes n}$, plays an important role below. An orthornormal basis can be labelled $\beta = |b|$, where $\beta = 0, 1, \ldots, n$:

$$|\beta\rangle = \binom{n}{\beta}^{-1/2} \sum_{b \in B_\beta} |b\rangle \text{ with } \beta = 0, 1, \ldots n \tag{A.1}$$

where $B_\beta = \{b : |b| = \beta\}$. It is well-known that the symmetric subspace $\mathcal{H}_+^{\otimes n}$ carries the irreducible representation of spin $j = J := n/2$ of $SU(2)$. In this language, the magnetic number m is related to β by $m = n/2 - \beta$ (here we are mapping $\beta_i \to m_i = (-1)^{\beta_i}/2$ for qubit i). In other words, we map

$|\beta\rangle \to |n/2, n/2-\beta\rangle$, where we stick to the standard notation $|j,m\rangle$ for the spin angular momentum eigenstates.

We will be concerned with evolution under unitary transformations $U_\theta := u_\theta^{\otimes n}$, where $u_\theta = \exp(i\theta|1\rangle\langle 1|)$, $\theta \in (-\pi, \pi]$. The operator N such that $U_\theta = \mathrm{e}^{i\theta N}$ will be referred to as number operator for obvious reasons: $N|b\rangle = |b||b\rangle$. The effect of noise is taken care of by a CP map $\mathscr{D}$, so the actual evolution of an initial n-qubit state $\psi := |\psi\rangle\langle\psi|$ is $\psi \to \mathscr{D}(U_\theta \psi U_\theta^\dagger) = \rho_\theta$.

With this notation the fidelity and success probability in Eqs. (6.3) and (6.4) can be written as

$$F(S) = \frac{1}{2}\left(1 + \frac{1}{S}\max_{\Omega}\sum_{b,b'}\Omega_{b,b'}\rho_{b',b}\delta_{|b'|,|b|+1}\right), \tag{A.2}$$

$$S = \sum_{b,b'}\Omega_{b,b'}\rho_{b',b}\delta_{|b'|,|b|}, \tag{A.3}$$

where the Kronecker delta tensors result from the integration of $\hat{\theta}$.

A.2 Local dephasing: Hadamard channel

The uncorrelated dephasing noise can be modelled by phase-flip errors that occur with probability p_f. i.e., at the single qubit level, the effect of the noise is $\varrho \to (1-p_f)\varrho + p_f\,\sigma_z\varrho\,\sigma_z$, where σ_z is the standard Pauli matrix $\sigma_z = \mathrm{diag}(1,-1)$. For states of n qubits, this, so called dephasing channel $\mathscr{D}$, is most easily characterized through its action on the operator basis $\{|b\rangle\langle b'|\}_{b,b'=1}^{2^n-1}$ as

$$\mathscr{D}(|b\rangle\langle b'|) = r^{|b+b'|}|b\rangle\langle b'|, \tag{A.4}$$

where the parameter r is related to the error probability p_f through $r = 1-2p_f$. The effect of $\mathscr{D}$ on a general n-qubit state $\varrho = \sum_{b,b'}\varrho_{b,b'}|b\rangle\langle b'|$ can then be written as the Hadamard (or entrywise) product

$$\mathscr{D}(\varrho) = \sum_{b,b'} r^{|b+b'|}\varrho_{b,b'}|b\rangle\langle b'| := \mathcal{D}\circ\varrho, \tag{A.5}$$

where $\mathcal{D} := \sum_{b,b'} r^{|b+b'|}|b\rangle\langle b'|$ and hereafter we understand that the sums over sequences run over all possible values of b (and b') unless otherwise specified. Note that Hadamard product is basis-dependent.

A.3 Symmetric probes

If the probe state is fully symmetric, i.e., $|\psi\rangle \in \mathcal{H}_+^{\otimes n}$, it can be written as $|\psi\rangle = \sum_\beta \psi_\beta|\beta\rangle$, where $|\beta\rangle$ is defined in (A.1) and the components are related to those in (6.5) by $c_m = \psi_{J-m}$ and can be taken to be positive with no loss of generality (any phase can be absorbed in the measurement operators). Then, $\rho = \mathscr{D}(\psi) = \mathcal{D} \circ \psi$ in (A.2) and (A.3) becomes

$$\rho = \sum_{\beta,\beta'} \frac{\psi_{\beta'}\psi_\beta}{\binom{n}{\beta}^{1/2}\binom{n}{\beta'}^{1/2}} \sum_{b\in B_\beta} \sum_{b'\in B_{\beta'}} r^{|b+b'|}|b'\rangle\langle b|. \tag{A.6}$$

Since ρ is permutation invariant, Ω can be chosen to be so and we can easily write (A.2) and (A.3) in the spin basis. We just need the non-zero Clebsch-Gordan matrix elements $\langle j,m'|b'\rangle\langle b|j,m\rangle$, where implicitly $m = J-\beta$, $m' = J-\beta'$. If we introduce the shorthand notation $\mathcal{D}^j_{m',m} := \langle j,m'|\mathcal{D}|j,m\rangle$, then using **?**, we have

$$\begin{aligned} \mathcal{D}^j_{m',m} &= \sum_{b\in B_\beta} \sum_{b'\in B_{\beta'}} r^{|b+b'|}\langle j,m'|b'\rangle\langle b|j,m\rangle \\ &= (1-r^2)^{J-j} r^{m-m'} \sum_k [\Delta_k^{(j)}]_m^{m'} r^{2k}, \end{aligned} \tag{A.7}$$

where

$$\left[\Delta_k^{(j)}\right]_m^{m'} := \frac{\sqrt{(j-m)!(j+m)!(j-m')!(j+m')!}}{(j-m-k)!(j+m'-k)!(m-m'+k)!k!}, \tag{A.8}$$

and the sums run over all integer values for which the factorials make sense. Recalling (6.6), a simple expression, involving just a sum over k in (A.8), for $\rho^j_{m,m'} = p_j^{-1}\mathrm{tr}(|j,m\rangle\langle j,m'| \otimes \mathbb{1}_j\ \rho)$ follows by combining the above results. In short,

$$\rho^j_{m',m} = \frac{1}{p_j} \frac{c_{m'}c_m}{\binom{n}{J-m'}^{1/2}\binom{n}{J-m}^{1/2}} \mathcal{D}^j_{m',m}, \tag{A.9}$$

where

$$p_j = \nu_j \sum_m \frac{c_m^2}{\binom{n}{J-m}} \mathcal{D}^j_{m,m}, \tag{A.10}$$

and the multiplicity is given by,

$$\nu_j = \binom{n}{J-j} \frac{2j+1}{J+j+1} \tag{A.11}$$

and a^j_m in (6.10) becomes

$$a^j_m = \frac{\mathcal{D}^j_{m,m+1}}{\sqrt{\mathcal{D}^j_{m,m}\mathcal{D}^j_{m+1,m+1}}}. \tag{A.12}$$

A.4 Relevant expressions for the multi-copy state

If the input state is of the form given in Eq. (6.18), the expressions (A.9) and (A.10) become

$$\rho^j_{m',m} = \frac{\mathcal{D}^j_{m',m}}{\sum_m \mathcal{D}^j_{m,m}} \text{ and } p_j = \nu_j\ 2^{-n} \sum_m \mathcal{D}^j_{m,m}\,, \tag{A.13}$$

where

$$\sum_m \mathcal{D}^j_{m,m} = (1-r^2)^{J-j}\frac{(1+r)^{2j+1}-(1-r)^{2j+1}}{2\,r} \tag{A.14}$$

The probability to find the state in the fully symmetric subspace ($j=J$) is important when assessing the success probability of the the ultimate bounds. Since the multiplicity for the maximum spin J is equal to one, it can be readily seen that p_J scales as

$$p_J \sim \mathrm{e}^{-n[\log 2 - \log(1+r)]}\,, \tag{A.15}$$

The critical probability s^*_j within a block can also be computed in the asymptotic limit $j \gg 1$ from Eq. (6.13)

$$s^*_j = \frac{\rho^j_{j,j}}{(\xi^j_j)^2} \sim \mathrm{e}^{-2j\log 1+r} \tag{A.16}$$

where ξ^j_m is the gaussian ground state, with $(\xi^j_j)^2 \sim \exp(-(1-r^2)\sqrt{j/4r})$. For $m=m'=j$ equation (A.7) gives $D^j_{j,j}=(1-r^2)^{J-j}$ which together with (A.13) and (A.14) gives $\rho^j_{j,j} \sim \exp[2j\log(r+1)]$. This scaling dominates over that of ξ^j_m , and hence determines the scaling of s^*_j. From here we obtain critical value for the overall success probability $S^* = p_J\ s^*_J \sim \mathrm{e}^{-n\log\ 2}$.

A.5 Ultimate bound without in-block filtering

In Section 6.6 section we discuss the possibility to prepare a probe state such that after the action of noise becomes an optimal state within the fully symmetric subspace $j=J$. Here we give what its critical success probability, which only entails computing p_J.

For this purpose we first recall that the optimal filtered state ξ^j_m, defined before Eq. (6.10), has to fulfil

$$\xi^j_m = f^j_m c_m \sqrt{\frac{\nu_j \mathcal{D}^j_{m,m}}{s_j p_j \binom{n}{J-m}}}\ . \tag{A.17}$$

The probability of falling in the block of maximum spin J for a given filtered state ξ^j can be easily derived from (A.17) recalling that the probe state ψ is normalized, and thus $\sum_m c_m^2 = 1$. Solving (A.17) for c_m^2/p_J and summing over m we obtain

$$\frac{1}{p_J} = s_J \sum_{m=-J}^{J} \frac{\binom{n}{J-m}}{\mathcal{D}^J_{m,m}} \left(\frac{\xi^J_m}{f^J_m}\right)^2 . \tag{A.18}$$

Now, for our strategy all j but the maximum one, $j = J$, are filtered out, and no further filtering is required within the block J, i.e. we have $f^J_m = 1$, for all $2J + 1$ values of m. Then $s_J = 1$ and

$$p_J = \left\{ \sum_{m=-J}^{J} \frac{\binom{n}{J-m}}{\mathcal{D}^J_{m,m}} \left(\xi^J_m\right)^2 \right\}^{-1} . \tag{A.19}$$

In the asymptotic limit the probability p_J can be estimated by noticing that the optimal distribution $(\xi^j_m)^2$ is much wider than $\binom{n}{J-m}/\mathcal{D}^J_{m,m}$ and can be replaced by $(\xi^J_0)^2$. Around $m = 0$, we can use the asymptotic formulas

$$\mathcal{D}^j_{m,m} \sim (1-r^2)^{J-j} \frac{(1+r)^{2j+1}}{2\sqrt{\pi r j}} \mathrm{e}^{-rm^2/j}, \tag{A.20}$$

$$\binom{n}{J-m} \sim \frac{2^n}{\sqrt{\pi J}} \mathrm{e}^{-m^2/J}. \tag{A.21}$$

They can be derived using the Stirling approximation and saddle point techniques. Eq. (A.20) also requires the Euler-Maclaurin approximation to turn the sum over k in (A.7) into an integral that can be evaluated using again the saddle point approximation. Retaining only exponential terms, $S^* = p_J \sim (1+r)^n/2^n = \mathrm{e}^{-n[\log 2 - \log(1+r)]}$.

A.6 Performance metrics: Equivalence of worst-case and average fidelity, and point-wise vs. global approach

Here we give a simple proof that for the estimation problem at hand, the worst-case fidelity in Eq. (6.2) and the average fidelity

$$F_{\text{av}} := \int \frac{d\theta}{2\pi} \int d\hat{\theta}\, p(\hat{\theta}|\theta, \text{succ}) f(\theta, \hat{\theta}) \tag{A.22}$$

(the integration limits $-\pi$, π are understood) take the same value, and so do the corresponding success probabilities. We recall that $f(\theta, \hat{\theta}) = [1 + \cos(\theta -$

$\hat{\theta})]/2$ and $p(\hat{\theta}|\theta, \mathrm{succ}) = \mathrm{tr}\,\rho_\theta M_{\hat{\theta}}$, and note that we have assumed a flat prior probability for $\rho_\theta = U_\theta \rho U_\theta^\dagger$. Obviously, $F \leq F_{\mathrm{av}}$. We just need to show that the opposite inequality also hold.

It is known that in order to maximize F_{av} one can choose a covariant measurement, so that $M_{\hat{\theta}} = U_{\hat{\theta}} \Omega U_{\hat{\theta}}^\dagger$ for a given seed Ω. Because of covariance, we note that for any phase θ' we have $\mathrm{tr}\,\rho_\theta M_{\hat{\theta}} = \mathrm{tr}\,\rho_{\theta'} M_{\hat{\theta}+\Delta\theta}$, where $\Delta\theta = \theta' - \theta$, thus $p(\hat{\theta}|\theta, \mathrm{succ}) = p(\hat{\theta} + \Delta\theta|\theta', \mathrm{succ})$. Likewise, we have $f(\theta, \hat{\theta}) = f(\theta', \hat{\theta} + \Delta\theta)$. By shifting variables $\hat{\theta} + \Delta\theta \to \hat{\theta}$ in Eq. (A.22), the integrant becomes independent of the variable θ, which can be trivially integrated to give

$$F_{\mathrm{av}} = \int d\hat{\theta}\, p(\hat{\theta}|\theta', \mathrm{succ}) f(\theta', \hat{\theta}) \tag{A.23}$$

for any θ'. It follows that

$$F_{\mathrm{av}} = \inf_{\theta \in (-\pi,\pi]} \int d\hat{\theta}\, p(\hat{\theta}|\theta, \mathrm{succ}) f(\theta, \hat{\theta}) \leq F, \tag{A.24}$$

where the last inequality states that the measurement that maximizes F_{av} need not maximize the worst-case fidelity F. We conclude that $F = F_{\mathrm{av}}$.

Proceeding along the same lines, we note that

$$S_{\mathrm{av}} = \int \frac{d\theta}{2\pi} \int d\hat{\theta}\, p(\hat{\theta}|\theta, \mathrm{succ}) = \int d\hat{\theta}\, p(\hat{\theta}|\theta', \mathrm{succ}) \tag{A.25}$$

for any θ'. Hence $S_{\mathrm{av}} = S$.

We point out here that most of the cited work dealing with quantum metrology, rather than the minimax approach used here, follows a *pointwise* approach that is aimed at improving the phase sensitivity around a rough estimate of θ. This very powerful and general approach is based on the quantum Cramer-Rao bound [Braunstein and Caves, 1994], and one can often argue that the so-obtained sensitivity can be attained by a suitable two-step adaptive protocol. However, the working hypotheses and the Cramer-Rao bound entail some subtleties that are often ignored, which can lead to erroneous conclusions [Berry *et al.*, 2012; Giovannetti and Maccone, 2012], wrong bounds, or misleading accounting of resources, even in the asymptotic regime of many such resources (see for instance Hayashi [2011]). In the particular case of probabilistic metrology the direct application of the pointwise approach can lead to unphysical results, as pointed out in Chiribella *et al.* [2013]. Here, we follow a *global approach* [Holevo, 1982; Aspachs *et al.*, 2009; Kołodyński and Demkowicz-Dobrzański, 2010] where no a priori knowledge about the phase is assumed, so instead the phase θ takes random equidistributed values in the interval $(-\pi, \pi]$. Within this approach the allocation of resources is straight-forward and the results are valid both for asymptotically large and finite amount of resources.

A.7 The continuum limit: Particle in a potential box

Proceeding as in Chapter 4, one can easily derive from Eqs. (6.8), (6.10) and (6.11) the following equation:

$$\begin{aligned}\langle\xi^j|H^j|\xi^j\rangle &= \sum_{m=-j}^{j-1}\left\{(\xi_{m+1}-\xi_m)^2+\frac{V_m^j}{j^2}\xi_m^2\right\}\\ &+ a_{-j}\xi_{-j}^2+a_j\xi_j^2,\end{aligned} \tag{A.26}$$

where $V_m^j = 2j^2(1-a_m)$ and we have dropped the superscript j in ξ_m^j to simplify the expression. In the asymptotic limit, as j becomes very large, $m/j = x$ approaches a continuum variable that takes values in the interval $[-1,1]$. Accordingly, the values $\{\sqrt{j}\xi_m\}$ approach a real function that we denote by $\varphi(x)$. With this, the former equation becomes

$$\begin{aligned}\langle\xi^j|H^j|\xi^j\rangle &= \frac{1}{j^2}\int_{-1}^{1}dx\left\{\left[\frac{d\varphi(x)}{dx}\right]^2+V^j(x)\varphi(x)^2\right\},\\ :&= \frac{1}{j^2}\left\langle\varphi\right|\mathcal{H}^j\left|\varphi\right\rangle,\end{aligned} \tag{A.27}$$

where

$$V^j(x)=2j^2(1-a_{xj}^j),\quad \mathcal{H}^j:=-d^2/dx^2+V^j(x), \tag{A.28}$$

and we have dropped the boundary term $[\varphi^2(-1)+\varphi^2(-1)]/j$ that stems from the second line in (A.26). Minimization of $\langle\xi^j|H^j|\xi^j\rangle$ require the vanishing of this term, and Eq. (6.14) readily follows. The formula

$$V^j(x)=j\frac{1-r^2}{2r\sqrt{1-(1-r^2)x^2}} \tag{A.29}$$

[also in Eq. (6.15)] follows from the asymptotic expression of a_m^j, defined in (6.10). Our starting point is Eq. (A.12) and (A.7). The sum over k in the latter can be evaluated using the Euler-Maclaurin formula and the saddle point approximation.

A.8 Symmetric probe is optimal and no benefit in probe-ancilla entanglement

We next show that permutation invariance enables us to choose with no loss of generality the probe state $|\psi\rangle$ from the symmetric subspace $\mathcal{H}_+^{\otimes n}$ and the seed Ω to be fully symmetric.

We first write Eqs. (A.2) and (A.3) in a more compact form. We define Δ as the non-trivial term of the fidelity through the relation $F(S) = (1 + S^{-1} \max_{\psi,\Omega} \Delta)/2$, where the maximization is performed also over the probe states since here we are concerned with the ultimate precision bound. We also introduce a slight modification of $\mathcal{D}$ that includes the Kronecker delta tensor: $\mathcal{D}_l := \sum_{b,b'} r^{|b+b'|}\delta_{|b'|,|b|+l}|b\rangle\langle b'|$, $l = 0, 1$. Then,

$$\Delta = \mathrm{tr}\left[(\psi \circ \Omega)\mathcal{D}_1\right], \quad S = \mathrm{tr}\left[(\psi \circ \Omega)\mathcal{D}_0\right], \tag{A.30}$$

where we have used that $\mathrm{tr}[A(B \circ C)] = \mathrm{tr}[(C \circ A)B]$ if $B = B^t$. The result we wish to show follows from the invariance of the noise under permutations of the n qubits, namely, from $U_\pi \mathcal{D}_l U_\pi^\dagger = \mathcal{D}_l$, for any $\pi \in S_n$, which implies that the very same value of Δ and S attained by some given measurement seed Ω and some initial state ψ, i.e., attained by $\psi \circ \Omega$, can also be attained by $U_\pi(\psi \circ \Omega)U_\pi^\dagger$, and likewise by the average $(n!)^{-1} \sum_{\pi\in S_n} U_\pi(\psi \circ \Omega)U_\pi^\dagger$.

The proof starts with yet a few more definitions: given a fully general probe state $|\psi\rangle$, we define the $n{+}1$ normalized states $|\phi_\beta\rangle = \sum_{b\in B_\beta}(\psi_b/\psi_\beta)|b\rangle$, $\beta = 0, 1, \ldots, n$ where $\psi_\beta^2 = \sum_{b\in B_\beta} |\psi_b|^2$, and write $|\psi\rangle = \sum_{\beta=0}^n \psi_\beta|\phi_\beta\rangle$. Additionally, we define

$$|\phi\rangle = \sum_{\beta=0}^{n} \binom{n}{\beta}^{1/2} |\phi_\beta\rangle, \quad \phi = |\phi\rangle\langle\phi|. \tag{A.31}$$

We obviously have $\phi \circ \Omega \geq 0$, as the Hadamard product of two positive operators is also a positive operator, and $(n!)^{-1} \sum_{\pi\in S_n} U_\pi\,(\phi \circ \Omega)\,U_\pi^\dagger \geq 0$, as this expression is a convex combination of positive operators. Similarly, the seed condition $\mathbb{1} - \Omega \geq 0$ implies $(n!)^{-1} \sum_{\pi\in S_n} U_\pi\,[\phi \circ (\mathbb{1} - \Omega)]\,U_\pi^\dagger \geq 0$. But

$$\frac{1}{n!}\sum_{\pi\in S_n} U_\pi(\phi \circ \mathbb{1})U_\pi^\dagger = \sum_{\beta=0}^{n} \frac{\binom{n}{\beta}}{n!} \left(\sum_{\pi\in S_n} U_\pi \phi_\beta U_\pi^\dagger\right) \circ \mathbb{1}, \tag{A.32}$$

since the diagonal entries of ϕ and ϕ_β transform among themselves under permutations. The right hand side can be written as

$$\sum_{\beta=0}^{n}\sum_{b\in B_\beta}\frac{\binom{n}{\beta}}{n!}\sum_{\pi\in S_n}\frac{|\psi_{\pi^{-1}(b)}|^2}{\psi_\beta^2}|b\rangle\langle b| = \sum_{\beta=0}^{n}\sum_{b\in B_\beta}|b\rangle\langle b| = \mathbb{1}, \tag{A.33}$$

where we have used that, for any $b \in B_\beta$, the set $\{\pi^{-1}(b)\}_{\pi\in S_n}$ contains exactly $\beta!(n-\beta)!$ times each one of the elements of B_β. It follows from Eqs. (A.32) and (A.33) that $\Omega^{\mathrm{sym}} := (n!)^{-1} \sum_{\pi\in S_n} U_\pi\,(\phi \circ \Omega)\,U_\pi^\dagger$ satisfies $0 \leq$

$\Omega^{\rm sym} \leq \mathbb{1}$ and is invariant under permutations of the n qubits. It is, therefore, a legitimate fully symmetric measurement seed. Moreover,

$$\frac{1}{n!}\sum_{\pi\in S_n} U_\pi(\psi\circ\Omega)U_\pi^\dagger = \sum_{\beta,\beta'} \frac{\psi_\beta\psi_{\beta'}}{\binom{n}{\beta}^{1/2}\binom{n}{\beta'}^{1/2}} \mathbb{1}_\beta \Omega^{\rm sym}\mathbb{1}_{\beta'}, \tag{A.34}$$

where $\mathbb{1}_\beta$ is the projector into the subspace with $|b| = \beta$, namely $\mathbb{1}_\beta := \sum_{b\in B_\beta} |b\rangle\langle b|$. Thus, recalling the definition of $|\beta\rangle$ in Eq. (A.1), the righthand side of (A.34) can be readily written as

$$\left(\sum_{\beta,\beta'} \psi_\beta\psi_{\beta'}|\beta\rangle\langle\beta'|\right) \circ \Omega^{\rm sym} = \psi^{\rm sym} \circ \Omega^{\rm sym}, \tag{A.35}$$

where $|\psi^{\rm sym}\rangle := \sum_{\beta=0}^{n} \psi_\beta|\beta\rangle \in \mathcal{H}_+^{\otimes n}$. It follows from these results and Eq. (A.30) that the very same fidelity and success probability attained by any pair $(|\psi\rangle, \Omega)$ of probe state and measurement seed is also attained by the state $|\psi^{\rm sym}\rangle \in \mathcal{H}_+^{\otimes n}$ and the fully symmetric seed $\Omega^{\rm sym}$. This completes the proof.

Now that we have learned that no boost in performance can be achieved by considering probe states more general than those in the symmetric subspace $\mathcal{H}_+^{\otimes n}$ (in the subspace of maximum spin $j = J$), we may wonder if entangling the probe with some ancillary system could enhance the precision. Here we show that this possibility can be immediately ruled out, thus extending the generality of our result. For this purpose we take the general probe-ancilla state $|\Psi_{\rm PA}\rangle = \sum_b \psi_b\,|b\rangle\,|\chi_b\rangle$, where $|\chi_b\rangle$ are normalized states (not necessarily orthogonal) of the ancillary system. The action of the phase evolution and noise on the probe leads to a state of the form $\rho_{\rm PA}(\theta) = \sum_{b,b'} r^{|b+b'|}\,{\rm e}^{i\theta(|b|-|b'|)}\,\psi_b\psi_{b'}\,|b\rangle\langle b'| \otimes |\chi_b\rangle\langle\chi_{b'}|$. This state could as well be prepared without the need of an ancillary system by taking instead an initial probe state $|\psi\rangle = \sum_b \psi_b\,|b\rangle$ and performing the trace-preserving completely positive map defined by $|b\rangle \to |b\rangle\,|\chi_b\rangle$ before implementing the measurement. This map can, of course, be interpreted as part of the measurement. It would correspond to a particular Neumark dilation of some measurement performed on the probe system alone, and hence it is included in our analysis.

A.9 Scavenging at the ultimate precision limit

The gentle measurement lemma Winter [1999] states that if measurement outcome occurs with a very high probability, e.g., an unfavorable event in some probabilistic protocol, which happens with probability $\bar{S} = {\rm tr}[\bar{\mathscr{F}}(\rho_\theta)] =$

$1-\epsilon$, then, in that event, the measurement causes very little disturbance to the state namely $\| \rho_\theta - \bar{\rho}_\theta \|_1 \leq \sqrt{2}\epsilon$, where $\bar{\rho} = \bar{\mathscr{F}}(\rho_\theta)/\bar{S}$.

Indeed, from (6.3), we find that the fidelity of the scavenged events, $\bar{F}(S)$, rapidly approaches that of the optimal deterministic machine $F(S=0)$:

$$\begin{aligned} F(0)-\bar{F}(S) &= \max_{0\leq\Omega\leq\mathbb{1}} \operatorname{tr} W\rho_\theta - \max_{0\leq\bar{\Omega}\leq\mathbb{1}} \operatorname{tr} \bar{W}\bar{\rho}_\theta \\ &\leq \min_{0\leq\Omega\leq\mathbb{1}} \operatorname{tr}[W(\rho_\theta-\bar{\rho}_\theta)] \leq \| \rho_\theta - \bar{\rho}_\theta \|_1 \leq \sqrt{2}S, \end{aligned} \tag{A.36}$$

where W (likewise $\bar{W}$) is shorthand for the matrix with entries $W_{b,b'} = \Omega_{b,b'}\delta_{|b|,|b'|+1}$.

We recall that, as $F(S)$ approaches the ultimate bound $F_{\text{ult}} = 1-(1-r^2)/(4nr)$, the success probability S decreases exponentially. Eq. (A.36) thus shows that such likely failure is not ruinous since in that event one can still recover the deterministic bound, i.e., one has $\bar{F} = F(0) = 1-(1-r^2)/(4nr^2)$.

APPENDIX B

Technical details of Chapter 7

B.1 Cloning fidelity and its upper bound

In this section we derive the upper bound in Eq. (7.6)

$$F_{\mathrm{CL}}^{\Pi} \leq \max_{E\in\mathsf{Sp}_m}\{p_{E,m}\} \left(\sum_{E\in\mathsf{Sp}_n}\sqrt{p_{E,n}\frac{\pi_E}{P_{\mathrm{succ}}}}\right)^2 . \tag{B.1}$$

to the probabilistic cloning fidelity. For the sake of self-completeness, we also derive the fidelity itself. We start from the definition of the worst-case cloning fidelity in Eq. (7.4):

$$F = \inf_{t\in\mathbb{R}} \frac{\langle\Psi_t^m|\mathscr{C}_{n,m}\left(|\Psi_t^n\rangle\langle\Psi_t^n|\right)|\Psi_t^m\rangle}{P_{\mathrm{succ}}} \tag{B.2}$$

where in full generality the quantum operation $\mathscr{C}_{n,m}$ has been decomposed as a probabilistic filter followed by a deterministic (trace preserving) map, i.e., as $\mathscr{C}_{n,m} = \mathscr{D}_{n,m}\circ\Pi$ [Chiribella *et al.*, 2013], and the probability of success is $P_{\mathrm{succ}} = \langle\Psi_t^n|\Pi|\Psi_t^n\rangle$. The covariance of quantum clocks enables us to drop the infimum in the last equation and consider without loss of generality only invariant filters of the form $\Pi = \sum_{E\in\mathsf{Sp}_n}\pi_E|E,n\rangle\langle E,n|$, as well as covariant maps, which satisfy

$$\mathscr{D}_{n,m}(U_t^n \ \cdot \ U_t^{n\dagger}) = U_t^m\mathscr{D}_{n,m}(\ \cdot\ )U_t^{m\dagger}, \tag{B.3}$$

where $U_t^n = \sum_{E\in\mathsf{Sp}_n} \mathrm{e}^{-iEt}|E,n\rangle\langle E,n|$ is the time evolution operator. Using the Choi-Jamiolkowski isomorphism, Eq. (B.3) is equivalent to

$$\mathcal{D} = U_t^m \otimes U_t^{n\,*}\; \mathcal{D}\; \left(U_t^m \otimes U_t^{n\,*}\right)^\dagger , \tag{B.4}$$

which in turn implies the direct sum decomposition $\mathcal{D}{=}\sum_{\mathcal{E}} \mathcal{D}_{\mathcal{E}}$, where

$$\mathcal{D}_{\mathcal{E}} {:=} \sum_{E,E'\in\mathsf{Sp}_n}{}' d^{\mathcal{E}}_{E,E'} |E+\mathcal{E},m\rangle\langle E'+\mathcal{E},m| \otimes |E,n\rangle\langle E',n| \tag{B.5}$$

and the 'primed' sum include only those terms for which $E{+}\mathcal{E}, E'{+}\mathcal{E} \in \mathsf{Sp}_m$. Taking the above into account, the probabilistic cloning fidelity can be cast as

$$\begin{aligned} F \;&=\; \frac{1}{P_{\rm succ}} \sum_{\mathcal{E}} \sum_{E,E'\in\mathsf{Sp}_n}{}' d^{\mathcal{E}}_{E,E'} \\ &\times\; \sqrt{\pi_E p_{E,n} p_{E+\mathcal{E},m}} \sqrt{\pi_{E'} p_{E',n} p_{E'+\mathcal{E},m}}\,. \end{aligned} \tag{B.6}$$

An upper bound to the fidelity (B.6) can be obtained by pulling the maximum value of $p_{E,m}$ out of the sums. Recalling the positivity of the Choi-Jamiolkowski operator, $\mathcal{D} \geq 0$, one has $\mathcal{D}_{\mathcal{E}} \geq 0$ and $|d^{\mathcal{E}}_{E,E'}|^2 \leq d^{\mathcal{E}}_{E,E} d^{\mathcal{E}}_{E',E'}$. Thus,

$$\begin{aligned} F \;&\leq\; \max_{E\in\mathsf{Sp}_m}\{p_{E,m}\} \\ &\times \sum_{E,E'\in\mathsf{Sp}_n} \sum_{\mathcal{E}}{}' \sqrt{\frac{\pi_E d^{\mathcal{E}}_{E,E} p_{E,n}}{P_{\rm succ}}} \sqrt{\frac{\pi_{E'} d^{\mathcal{E}}_{E',E'} p_{E',n}}{P_{\rm succ}}}, \end{aligned} \tag{B.7}$$

where we have interchanged the order of summation. We can now use the Schwarz inequality to write

$$\sum_{\mathcal{E}}{}' \sqrt{d^{\mathcal{E}}_{E,E}} \sqrt{d^{\mathcal{E}}_{E',E'}} \leq \left(\sum_{\mathcal{E}}{}' d^{\mathcal{E}}_{E,E}\right) \left(\sum_{\mathcal{E}}{}' d^{\mathcal{E}}_{E',E'}\right). \tag{B.8}$$

The first (second) sum on the right hand side of the last equation can be extended to values of $\mathcal{E}$ for which $E+\mathcal{E} \in \mathsf{Sp}_m$ ($E'+\mathcal{E} \in \mathsf{Sp}_m$), as $d^{\mathcal{E}}_{E,E} \geq 0$ ($d^{\mathcal{E}}_{E',E'} \geq 0$). Then, each of these sums is unity since $\mathscr{D}_{n,m}$ is trace preserving. Substituting in (B.7), we obtain the bound in Eq. (B.1).

B.2 Fidelity of probabilistic metrology

In this section we derive the expression for the PM fidelity in Eq. (B.9):

$$F = \sum_{\mathcal{E}} \left(\sum_{E\in\mathsf{Sp}_n}{}' \sqrt{p_{E,n}\, p_{E+\mathcal{E},m}\, \hat{p}_{E+\mathcal{E},m}\, \frac{\pi_E}{P_{\rm succ}}} \right)^2 . \tag{B.9}$$

Since later we will show that asymptotically the r.h.s. of Eq. (B.9) achieves the optimal cloning fidelity, here it is enough to show that the r.h.s. of Eq. (B.9) can be achieved by some suitable measurement.

For every $\sigma > 0$, consider the operators $\Phi^n_{\hat{t},\sigma} = |\Phi^n_{\hat{t},\sigma}\rangle\langle\Phi^n_{\hat{t},\sigma}|$, where

$$|\Phi^n_{\hat{t},\sigma}\rangle := \sqrt{p_\sigma(\hat{t}\,)}\,\sum_{E\in\mathsf{Sp}_n} \mathrm{e}^{-iE\hat{t}}|E,n\rangle\,, \tag{B.10}$$

and $p_\sigma(t)$ is a suitable probability distribution. The latter is chosen as follows: For commensurable energies, when the evolution is periodic, $p_\sigma(t)$ is the uniform distribution $p_\sigma(t) = 1/T$ over the period T [$p_\sigma(t) = 0$ for $t < 0$ and $T < t$]. For incommensurable energies, $p_\sigma(t)$ is the Gaussian distribution $p_\sigma(t) = (2\pi\sigma^2)^{-1/2}\exp\{-t^2/(2\sigma^2)\}$. Now, for commensurable energies, the operators $\{\Phi^n_{\hat{t},\sigma}\}_{\hat{t}\in\mathbb{R}}$ define a quantum measurement: indeed, it is immediate to check the normalization condition $\int \mathrm{d}\hat{t}\,\Phi^n_{\hat{t},\sigma} = \mathbb{1}_n$, where $\mathbb{1}_n$ denotes the identity on the subspace containing the state of the n input copies. For incommensurable energies, the operators $\{\Phi^n_{\hat{t},\sigma}\}_{\hat{t}\in\mathbb{R}}$ form an "approximate measurement", satisfying the approximate normalization condition $\int \mathrm{d}\hat{t}\,\Phi^n_{\hat{t},\sigma} = \mathbb{1}_n + O(e^{-\sigma^2})$. Note that in the limit $\sigma \to \infty$ the approximate measurement becomes arbitrarily close to a legitimate measurement, as the normalization defect disappears in such limit.

Let us denote by F_σ the value obtained by inserting the approximate measurement $\{\Phi^n_{\hat{t},\sigma}\}_{\hat{t}\in\mathbb{R}}$ in

$$F = \inf_{t\in\mathbb{R}} \int \mathrm{d}\hat{t}\, p(\hat{t}|t)\, \left|\langle\hat{\Psi}^m_{\hat{t}}|\Psi^m_t\rangle\right|^2 . \tag{B.11}$$

Since our set of approximate measurements becomes closer and closer to the set of allowed measurements as $\sigma \to \infty$, the limit value $F_* := \lim_{\sigma\to\infty} F_\sigma$ is an achievable value of the fidelity. We now show that F_* is equal to the r.h.s. of Eq. (B.9): recalling the expansion of $|\Psi^n_t\rangle$ in Eq. (7.1) and the definition of $|\hat{\Psi}^n_{\hat{t}}\rangle$ in the first paragraph of Section 7.2, we obtain

$$F_\sigma = \inf_t \int \mathrm{d}\hat{t}\, p_\sigma(\hat{t}\,) \left|\sum_{\mathcal{E}} \mathrm{e}^{-i\mathcal{E}(t-\hat{t})} f_{\mathcal{E}}\right|^2 , \tag{B.12}$$

with

$$f_{\mathcal{E}} := \sum_{E\in\mathsf{Sp}_n} \sqrt{\frac{\pi_E p_{E,n} p_{E+\mathcal{E},m} \hat{p}_{E+\mathcal{E},m}}{P_{\mathrm{succ}}}}\,. \tag{B.13}$$

Integrating over $\hat{t}$ we then obtain

$$\begin{aligned} F_* &= \lim_{\sigma\to\infty} \inf_t \sum_{\mathcal{E},\mathcal{E}'} \mathrm{e}^{-\frac{\sigma^2(\mathcal{E}-\mathcal{E}')^2}{2}}\, \mathrm{e}^{-it(\mathcal{E}-\mathcal{E}')}\, f_{\mathcal{E}} f_{\mathcal{E}'} \\ &= \sum_{\mathcal{E}} f_{\mathcal{E}}^2\,, \end{aligned}$$

which coincides with the r.h.s. of Eq. (B.9).

B.3 Geometry of the problem and Smith variables

B.3.1 Rationally independent energy units and lattices

We recall that as a set of points the partitions $\{(n_0,\mathbf{n})\}$, where $\mathbf{n}\in\mathscr{P}_n$, is a regular lattice on the simplex $\Delta_n^{d-1}=\{(x_0,\dots,x_{d-1})\in\mathbb{R}^d : x_j\geq 0, j=0,\dots,d-1 \text{ and } \sum_{j=0}^{d-1}x_j=n\}$, of edge length n. Each site of this lattice is of the form $(n-\sum_{j=1}^{d-1}n_j,\mathbf{n})$. The vectors (1 column integer matrices) $\mathbf{n}\in\mathscr{P}_n$ form themselves also a regular lattice, defined by the inequalities $0\leq n_j\leq n-\sum_{l=1}^{j-1}n_l$, $j=1,\dots,d-1$, inside the corner of a $(d-1)$-dimensional cube of side length n: $\Delta_{c,n}^{d-1}=\{(x_1,\dots,x_{d-1})\in\mathbb{R}^{d-1} : 0\leq x_j\leq n-\sum_{l=1}^{j-1}x_l, j=1,\dots,d-1\}$. Note that if $n\leq m$, one has the inclusion $\mathscr{P}_n\subset\mathscr{P}_m$.

Recall also that the spectrum of $H^{\otimes n}$, is given by $\mathsf{Sp}_n=\{E\mid\exists=\mathbf{n}\in\mathscr{P}_n : \mathbf{e}\,\mathbf{n}=E\}$. All vectors $\mathbf{n}$ that give rise to a particular value E of the energy, i.e., those in the set $\mathscr{P}_n^E=\{\mathbf{n}\in\mathscr{P}_n : E_{\mathbf{n}}=E\}$, necessary lie in the affine hyperplane obtained by translating the hyperplane orthogonal to $\mathbf{e}$. Since $\mathscr{P}_n$ is a regular lattice, it is not clear a priori how many of its sites fall on the hyperplane defined by $\mathbf{e}$; the actual degeneracy of E strongly depends on the commensurability of the energies $\{e_j\}$ in the spectrum of H. To identify all distinct values of $E_{\mathbf{n}}\in\mathsf{Sp}_n$ we use the fact that the energies of the Hamiltonian H can always be written as a linear combination with integer coefficients of a minimal set of rationally independent 'energy units' $\{\varepsilon_l\}_{l=1}^r$, $r\leq d-1$, namely

$$e_j=\sum_l k_{lj}\varepsilon_l \qquad k_{lj}\in\mathbb{Z}\,. \tag{B.14}$$

By *minimal* we mean that no subset of $\{\varepsilon_l\}_{l=1}^r$ is sufficient to write every energy in the spectrum of H as in Eq. (B.14). Note that the rational independence of ε_l implies that k_{lj} is fixed once the choice of energy units is made. With this, the energies of Sp_n can be written as $E_{\mathbf{n}}=\sum_{l=1}^r\varepsilon_l\tilde{n}_l$, where $\tilde{n}_l=\sum_{j=1}^{d-1}k_{lj}n_j$. It is then useful to introduce the vector notation $\tilde{\mathbf{n}}=\mathbf{K}\,\mathbf{n}$, where $\mathbf{K}$ is the $r\times(d-1)$ matrix whose integer entries are k_{lj}, and $E_{\mathbf{n}} = \boldsymbol{\varepsilon}\, \tilde{\mathbf{n}}$, where $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_r)$. Because of the rational independency of the energy units, we conclude that there is a bijection between the distinct energies in Sp_n and the points in the set $\widetilde{\mathscr{P}}_n = \{\tilde{\mathbf{n}} = \mathbf{K}\, \mathbf{n} : \mathbf{n} \in \mathscr{P}_n\}$.

We view each column $\mathbf{K}_j = (k_{1j}, \ldots, k_{rj})^t \in \mathbb{Z}^r$ of $\mathbf{K}$ as a set of vectors that span the (infinite) Bravais lattice

$$\widetilde{\mathscr{P}}_\infty = \Big\{\tilde{\mathbf{n}} = \textstyle\sum_{j=1}^{d} n_j\, \mathbf{K}_j,\ n_j \in \mathbb{Z}\Big\}. \tag{B.15}$$

We have the obvious inclusion $\widetilde{\mathscr{P}}_n \subset \widetilde{\mathscr{P}}_\infty$. For finite n, $\widetilde{\mathscr{P}}_n$ departs from the Bravais lattice $\widetilde{\mathscr{P}}_\infty$ in two ways: i) the lattice $\widetilde{\mathscr{P}}_n$, defined by the linear transformation $\mathbf{K}$ acting on $\mathscr{P}_n$, inherits its boundaries and hence lies inside the convex r-dimensional polytope $\widetilde{\Delta}^r_n := \mathbf{K}\Delta^{d-1}_{c,n}$; ii) near the boundaries of $\widetilde{\Delta}^r_n$, some points of $\widetilde{\mathscr{P}}_\infty$ are missing in $\widetilde{\mathscr{P}}_n$ (since none of the corresponding inverse images satisfy the constrains that define $\mathscr{P}_n$, given in the first paragraph of this section), so we typically have $\widetilde{\mathscr{P}}_n \subsetneq \widetilde{\mathscr{P}}_\infty \cap \widetilde{\Delta}^r_n$. These boundary related issues have a minor effect in our analysis for asymptotically large n, as we argue below.

B.3.2 Smith vectors/variables. Volume of primitive cell

For $r < d-1$, the vectors $\mathbf{K}_j$ are not linearly independent and, therefore, they cannot be a minimal set of primitive vectors of the lattice $\widetilde{\mathscr{P}}_\infty$. On the other hand, having a minimal set of primitive vectors is necessary in order to compute the volume ΔV^* of the unit cell [see, e.g., Eqs. (7.11) and (7.12)]. To this purpose, we use the Smith normal form of $\mathbf{K}$ [Marcus and Minc, 1964]:

$$\mathbf{K} = \mathbf{TAPS}, \tag{B.16}$$

where $\mathbf{T} \in \mathbb{Z}^{r\times r}$ and $\mathbf{S} \in \mathbb{Z}^{(d-1)\times(d-1)}$ are unimodular matrices, i.e. invertible matrices over the integers with $\det \mathbf{T} = \det \mathbf{S} = \pm 1$; $\mathbf{A} \in \mathbb{Z}^{r\times r}$ is a diagonal matrix with entries

$$(\mathbf{A})_{ll} = \frac{\gcd\{[\mathbf{K}]_l\}}{\gcd\{[\mathbf{K}]_{l-1}\}}, \tag{B.17}$$

where gcd stands for greatest common divisor and $\{[\mathbf{K}]_l\}$ is the set of all minors of $\mathbf{K}$ of order l; $\mathbf{P}$ is the $r \times (d-1)$ matrix with entries

$$(\mathbf{P})_{ll} = 1, \quad 1 \le l \le r, \tag{B.18}$$

and zero otherwise. Using (B.16) we can define the lattice $\widetilde{\mathscr{P}}_n$ in terms of the new vectors/variables

$$\mathbf{s} = (s_1, \ldots, s_r)^t := \mathbf{PSn}, \tag{B.19}$$

which we coin *Smith vectors/variables*. Given a matrix $\mathbf{K}$ the Smith form guarantees that the choice of Smiths variables is unique, up to linear combinations within the degenerate subspaces of $\mathbf{A}$ (if any). Note that since the matrix $\mathbf{S}$ (and analogously $\mathbf{T}$) is unimodular, the gcd of each of its rows and columns is unity. Indeed, for the first column (similarly for the other columns/rows) one has $\mathbf{S}_1 = \gcd\{(\mathbf{S})_{j,1}\}\,\mathbf{a}_1$, for some $\mathbf{a}_1 \in \mathbb{Z}^{d-1}$. Then, $1 = \det \mathbf{S} = \gcd\{(\mathbf{S})_{j,1}\} \det \bar{\mathbf{S}}$, where $\bar{\mathbf{S}}$ is obtained by substituting $\mathbf{a}_1$ for the first column of $\mathbf{S}$. Since $\det \bar{\mathbf{S}} \in \mathbb{Z}$, necessarily $\det \bar{\mathbf{S}} = \gcd\{(\mathbf{S})_{j,1}\} = 1$. Being the case that the gcd of each row of $\mathbf{S}$ is unity, Bezout lemma ensures that each component of $\mathbf{s}$, s_l, will take *all* integer values. That is, in stark contrast to the variables $\widetilde{n}_i$, the Smith variables s_l take values independently of each other, with unit spacing between consecutive values. This means that the Bravais lattice $\widetilde{\mathscr{P}}_\infty$ is defined in terms of the Smith variables as

$$\widetilde{\mathscr{P}}_\infty = \left\{\tilde{\mathbf{n}} = \textstyle\sum_{l=1}^{r} s_l\,\mathbf{v}_l,\ s_l \in \mathbb{Z}\right\}, \tag{B.20}$$

where $\{\mathbf{v}_l\}_{l=1}^{r}$ are a set of (linear independent) primitive vectors of the lattice defined by each of the r columns of $\mathbf{TA}$. It follows that the volume of the unit cell can be computed as

$$\Delta V^* = |\det(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r)| = \det \mathbf{A} = \gcd\{[\mathbf{K}]_r\}, \tag{B.21}$$

where we have used Eq. (B.17).

B.3.3 Parametrizing the energy in terms of the Smith variables

Both the vectors $\tilde{\mathbf{n}}$ and the Smith vectors $\mathbf{s}$ are in one-to-one correspondence with the distinct energies in Sp_n, and thus they are interchangeable in all our arguments. As a matter of fact, they would coincide had we chosen the energy units as

$$\boldsymbol{\varepsilon}^{\mathrm{S}} := \boldsymbol{\varepsilon}\,\mathbf{T}\,\mathbf{A} \tag{B.22}$$

(the script S stands for Smith), so that the total energy becomes $E_{\mathbf{s}} = \boldsymbol{\varepsilon}^{\mathrm{S}}\,\mathbf{s}$. Note that, if $\boldsymbol{\varepsilon}$ is minimal, then also $\boldsymbol{\varepsilon}^S$ must be minimal, since the two vectors are related by an invertible matrix with integer entries.

Despite the one-to-one correspondence, the Smith variables $\mathbf{s}$ are more convenient than the variables $\tilde{\mathbf{n}}$. Indeed, they define a cubic lattice with unit spacing, which facilitates, e.g., taking the continuum limit. Note that one has

$$\begin{aligned}
\det \widetilde{\mathbf{\Sigma}} =& \det(\mathbf{K}\mathbf{\Sigma}\mathbf{K}^t) \\
=& \det\left[(\mathbf{T}\,\mathbf{A}\,\mathbf{P}\,\mathbf{S})\,\mathbf{\Sigma}\,(\mathbf{T}\,\mathbf{A}\,\mathbf{P}\,\mathbf{S})^t\right] \\
=& \det{}^2\mathbf{A}\,\det\mathbf{\Sigma}^{\mathrm{S}} \qquad \mathbf{\Sigma}^{\mathrm{S}} := \mathbf{P}\,\mathbf{S}\,\mathbf{\Sigma}(\mathbf{P}\,\mathbf{S})^t \\
\equiv& (\Delta V^*)^2 \det\mathbf{\Sigma}^{\mathrm{S}},
\end{aligned}$$

which implies the relation

$$\Delta V^* \left(\det\widetilde{\mathbf{\Sigma}}\right)^{-\frac{1}{2}} = \left(\det\mathbf{\Sigma}^{\mathrm{S}}\right)^{-\frac{1}{2}} . \tag{B.23}$$

Using this fact, the volume of the unit cell in Eqs. (7.11) and (7.12) in main text can absorbed into the matrix $\mathbf{\Sigma}^{\mathrm{S}}$, which is the covariance matrix of the probability distribution of Smith variables $p_{\mathbf{s},n}$.

B.3.4 Equivalence of different choices of energy units

The choice of rationally independent energy units that define $\mathbf{K}$ is not unique. Here we show that such ambiguity has no physical implications. Our strategy is to show that different choices of energy units lead to Smith variables that are related by unimodular matrices.

Let $\boldsymbol{\varepsilon}$ and $\boldsymbol{\varepsilon}'$ be two minimal sets of rationally independent energy units spanning the spectrum of H. In the following we will use the same notation of the previous section, attaching primes to all the matrices and quantities defined in terms of $\boldsymbol{\varepsilon}'$. Note that, by minimality, there must be an invertible transformation $\mathbf{R}$ such that $\boldsymbol{\varepsilon} = \boldsymbol{\varepsilon}'\,\mathbf{R}$. Comparing the two relations $E_{\mathbf{n}} = \boldsymbol{\varepsilon}'\,\mathbf{K}'\,\mathbf{n}$ and $E_{\mathbf{n}} = \boldsymbol{\varepsilon}\,\mathbf{K}\,\mathbf{n} = \boldsymbol{\varepsilon}'\,\mathbf{R}\,\mathbf{K}\,\mathbf{n}$ and using the rational independence of the units $\boldsymbol{\varepsilon}'$ we obtain the relation

$$\begin{aligned}
\mathbf{K}'\,\mathbf{n} &= \mathbf{R}\,\mathbf{K}\,\mathbf{n} \\
&= \mathbf{R}\,\mathbf{T}\,\mathbf{A}\,\mathbf{P}\,\mathbf{S}\,\mathbf{n} \\
&= \mathbf{R}\,\mathbf{T}\,\mathbf{A}\,\mathbf{s} \\
&= \mathbf{M}\,\mathbf{s}, \qquad \mathbf{M} := \mathbf{R}\,\mathbf{T}\,\mathbf{A}\,,
\end{aligned} \tag{B.24}$$

the second and third equalities coming from the Smith form in Eq. (B.16) and the definition of the Smith vectors in Eq. (B.19), respectively. Now, recall that $\mathbf{K}'$ is a matrix of integers, and, therefore $\mathbf{K}'\,\mathbf{n}$ is a vector of integers for every $\mathbf{n} \in \mathbb{Z}^{d-1}$. Since the Smith variables s_l are independent and take all possible integer values, by choosing $\mathbf{s} = (1, 0, \ldots, 0)^t$, the relation $\mathbf{K}'\,\mathbf{n} = \mathbf{M}\,\mathbf{s}$ implies that the first column of $\mathbf{M}$ must have integer entries. By the same argument, all columns of $\mathbf{M}$ must have integer entries, i.e. $\mathbf{M}$ is an integer matrix. Note that $\mathbf{M}$ is invertible (since it is defined as the product of three invertible matrices), although its inverse needs not be a matrix with integer entries.

Let us write $\mathbf{M}$ in the Smith form $\mathbf{M} = \mathbf{U}\,\mathbf{B}\,\mathbf{V}$, where $\mathbf{U}$ and $\mathbf{V}$ are unimodular and $\mathbf{B}$ is an invertible diagonal matrix. Inserting this expression in Eq. (B.24), we obtain

$$\begin{aligned}\mathbf{K}'\,\mathbf{n} &= \mathbf{U}\,\mathbf{B}\,\mathbf{V}\,\mathbf{s}\\ &= \mathbf{U}\,\mathbf{B}\,\mathbf{V}\,\mathbf{P}\,\mathbf{S}\,\mathbf{n}\,,\end{aligned}$$

having used Eq. (B.19). Since the relation holds for arbitrary integer vectors, we obtained $\mathbf{K}' = \mathbf{UBVPS}$. Note that, by definition of the matrix $\mathbf{P}$ in Eq. (B.18), we have

$$\mathbf{V}\,\mathbf{P} = \mathbf{P}\,\mathbf{W}\,, \tag{B.25}$$

where $\mathbf{W}$ is the $(d-1)\times(d-1)$ matrix defined as, e.g., $\mathbf{W} = \mathbf{V} \oplus \mathbb{1}_{d-1-r}$. Hence, we have $\mathbf{K}' = \mathbf{UBPWS}$. Now, comparing this equation with the Smith form $\mathbf{K}' = \mathbf{T}'\mathbf{A}'\mathbf{P}'\mathbf{S}'$ we obtain $\mathbf{T}' = \mathbf{U}$, $\mathbf{A}' = \mathbf{B}$, $\mathbf{P}' = \mathbf{P}$ and $\mathbf{S}' = \mathbf{W}\,\mathbf{S}$. In conclusion, the Smith variables defined by $\mathbf{K}'$ are related to the Smith variables defined by $\mathbf{K}$ through the relation

$$\begin{aligned}\mathbf{s}' &\equiv \mathbf{P}'\,\mathbf{S}'\,\mathbf{n}\\ &= \mathbf{P}\,\mathbf{W}\,\mathbf{S}\,\mathbf{n}\\ &= \mathbf{V}\,\mathbf{P}\,\mathbf{S}\,\mathbf{n}\\ &\equiv \mathbf{V}\,\mathbf{s}\,,\end{aligned}$$

having used Eq. (B.25) in the third equality. Clearly, the covariance matrix of $\mathbf{s}'$ is given by $\mathbf{\Sigma}^{\mathrm{S}'} = \mathbf{V}\mathbf{\Sigma}^{\mathrm{S}}\mathbf{V}^t$, where $\mathbf{\Sigma}^{\mathrm{S}}$ is the covariance matrix of $\mathbf{s}$, and, thanks to the unimodularity of $\mathbf{V}$, we have $\det\mathbf{\Sigma}^{\mathrm{S}'} = \det\mathbf{\Sigma}^{\mathrm{S}}$. Using Eq. (B.23) we conclude that

$$\begin{aligned}\Delta V^*(\det\widetilde{\mathbf{\Sigma}})^{-1/2} &= (\det\mathbf{\Sigma}^{\mathrm{S}})^{-1/2}\\ &= (\det\mathbf{\Sigma}^{\mathrm{S}'})^{-1/2}\\ &= \Delta V^{*\prime}(\det\widetilde{\mathbf{\Sigma}}')^{-1/2}\,.\end{aligned}$$

This proves that physically relevant quantities, such as the most probable energy in Eq. (7.11) or the fidelity in Eq. (7.12), do not depend on the choice of energy units used to represent the spectrum of H.

B.3.5 Boundary defects

Ideally, one would like to have $\widetilde{\mathscr{P}}_n = \widetilde{\mathscr{P}}_\infty \cap \widetilde{\Delta}^r_n$, however, near the boundary of the polytope $\widetilde{\Delta}^r_n$ some sides are missing, giving rise to loss of regularity, as already mentioned. We argue here that the thickness of the defective region does not scale with n. This ensures that in the asymptotic limit of large n, e.g., the sums over $\widetilde{\mathscr{P}}_\infty$ can always be safely approximated by integrals over $\widetilde{\Delta}^r_n$, even when the probability distribution $p_{\mathbf{n},n}$ is flat, as required to obtain Eq. (7.13). The thickness of the defective region depends solely on the intrinsic properties of the Hamiltonian H, including the number of rationally independent energy units, through the matrix $\mathbf{K}$.

To simplify the argument let us assume that $\widetilde{\mathscr{P}}_\infty$ is a cubic lattice with unit spacing. It has been shown above that this assumption does not entail any loss of generality, as it just requires choosing energy units as in (B.22). Let $\mathbf{K}^l = (k_{l\,1}, \ldots, k_{l\,d-1})$ stand for the row l of the matrix $\mathbf{K}$. Then, for each $l = 1, \ldots, r$,

$$\tilde{n}_l = \mathbf{K}^l \cdot \mathbf{x}, \quad \mathbf{x} \in \mathbb{R}^{d-1}, \tag{B.26}$$

where $\tilde{n}_l$ is the l-th component of $\tilde{\mathbf{n}} \in \widetilde{\mathscr{P}}_\infty \cap \widetilde{\Delta}^r_n$, defines a discrete family of hyperplanes that are orthogonal to $\mathbf{K}^l$. The kernel of $\mathbf{K} \in \mathbb{Z}^{r\times(d-1)}$ can always be spanned by a set $\{\mathbf{u}_k\}_{k=1}^{d-r-1}$ of independent vectors of $\mathbb{Z}^{d-1}$ that are orthogonal to $\{\mathbf{K}^l\}_{l=1}^r$. Upper bounds, b, to the minimum length of these (integer) vectors (high-dimensional extensions of the so-called Siegel's bound) can be found in Bombieri and Vaaler [1983]; Borosh [1976] and depends entirely on the matrix $\mathbf{K}$ (b is thus independent of n). We note that $\{\mathbf{u}_k\}_{k=1}^{d-r-1}$ define a lattice within $\ker \mathbf{K}$ and any $(d-r-1)$-dimensional ball or radius b (or larger) contains at least one site of it.

A site $\tilde{\mathbf{n}} \in \widetilde{\mathscr{P}}_\infty \cap \widetilde{\Delta}^r_n$ belongs to $\widetilde{\mathscr{P}}_n$ iff there is at least one site $\mathbf{n} \in \mathscr{P}_n$ that satisfies (B.26) for all l. The Smith normal form and Bezout lemma (see previous subsection) ensure that there are infinitely many vectors $\mathbf{n}$ in $\mathscr{P}_\infty$ satisfying (B.26) for a given $\tilde{\mathbf{n}}$ (but they are not necessarily in $\widetilde{\mathscr{P}}_n$). The difference between any two such vectors, $\mathbf{n} - \mathbf{n}'$ belongs to $\ker \mathbf{K}$, i.e., satisfies $\mathbf{n} - \mathbf{n}' = \sum_{k=1}^{d-r-1} \eta_k \mathbf{u}_k$, $\eta_k \in \mathbb{Z}$. Therefore, if a given $\tilde{\mathbf{n}}$ is such that the intersection of the polytope $\widetilde{\Delta}^r_n$ with the hyperplanes defined by (B.26) contains a ball of radius b, it is ensured that $\tilde{\mathbf{n}} \in \widetilde{\mathscr{P}}_n$. This is so because at least one site in $\mathbf{n}_0 + \ker \mathbf{K}$, where $\mathbf{n}_0 \in \mathscr{P}_\infty$ is a solution (any of the infinitely many solutions) of (B.26), will be contained in this ball, as argued above. Since $\Delta^{d-1}_{c,n}$ is convex, by increasing n (the length of its edge) the volume of its intersection with the hyperplanes (B.26), which is also a convex polytope, grows as n^{d-r-1} and it will eventually contain a ball of radius b. Only sides whose distance to the boundary is kept fixed and is small enough can escape from this fate and may thus not be in $\widetilde{\mathscr{P}}_n$. This concludes our proof.

By the same argument, since the volume of the intersection of $\Delta^{d-1}_{c,m}$ with the hyperplanes (B.26) grows with m (but at a fixed distance from the boundary), the number of balls or radius b it will eventually contain grows as m^{d-r-1} and so does the number of point in $\ker \mathbf{K}$, i.e., the numbers of points in $\mathscr{P}_m$ associated to a given $\widetilde{\mathbf{m}}$, as required to obtain, e.g., Eq. (7.11).

As a final remark, we note also that the pattern of defects near the boundary of $\tilde{\Delta}^r_n$ is exactly the same for every n, provided that n is sufficiently large. The reason is that all the lattices $\mathscr{P}_n$ are similar (i.e., have the same shape but different size). Hence, the regions of them that are at a fixed distance from the boundary are identical and, as argued above, these regions determine the pattern of defects of $\widetilde{\mathscr{P}}_n$.

B.4 Equivalence between PM and macroscopic cloning

In this section we show that the PM fidelity attains the upper bound (B.1) to the cloning fidelity, and thus prove the equivalence between PM and macroscopic cloning of quantum clocks. The proof uses the one-to-one correspondence between Sp_m and $\widetilde{\mathscr{P}}_m$, as well as the fact that the probability distribution $p_{\widetilde{\mathbf{m}},m}$ approaches a multivariate Gaussian distribution as m goes to infinity. The former, enables us to write Eq. (B.9) as

$$F^{\Pi}_{\mathrm{PM}} = \sum_{\widetilde{\boldsymbol{\mu}}} \left({\sum_{\widetilde{\mathbf{n}}\in\widetilde{\mathscr{P}}_n}}' \xi_{\widetilde{\mathbf{n}}} \sqrt{\hat{p}_{\widetilde{\mathbf{n}}+\widetilde{\boldsymbol{\mu}},m}\, p_{\widetilde{\mathbf{n}}+\widetilde{\boldsymbol{\mu}},m}} \right)^2 , \tag{B.27}$$

where $\xi_{\widetilde{\mathbf{n}}} := \sqrt{p_{\widetilde{\mathbf{n}},n}\pi_{\widetilde{\mathbf{n}}}/P_{\mathrm{succ}}}$ and the outer sum runs over $\widetilde{\mathscr{P}}_{n,m} := \{\widetilde{\boldsymbol{\mu}} = \widetilde{\mathbf{m}} - \tilde{\mathbf{n}} \mid \tilde{\mathbf{n}} \in \widetilde{\mathscr{P}}_n, \widetilde{\mathbf{m}} \in \widetilde{\mathscr{P}}_m\}$. To keep our notation as uncluttered as possible we suppress the tildes throughout the rest of this section. We further assume that the multivariate normal distribution $p_{\mathbf{m},m}$ peaks at $\mathbf{m} = \mathbf{0}$, and so does $p_{\mathbf{n},n}$, but we make no other assumption on the form of $p_{\mathbf{n},n}$ (it could, e.g., be flat, in which case any point in $\mathscr{P}_n$ could be chosen to be $\mathbf{0}$). This may require shifting the vectors in $\mathscr{P}_m$ by a fixed $\mathbf{m}_0 \in \mathscr{P}_m$ (similarly, by a fixed $\mathbf{n_0} \in \mathscr{P}_n$ for those in $\mathscr{P}_n$). The primed summation is restricted to vectors $\mathbf{n}$ such that $\boldsymbol{\mu}+\mathbf{n}\in \mathscr{P}_m$.

For any $\rho > \nu > 0$, where $\nu = \max\{|\mathbf{n}| : \mathbf{n} \in \mathscr{P}_n\}$, define the set $\mathcal{R}_\rho = \{\boldsymbol{\mu} : \forall\mathbf{n}\in\mathscr{P}_n,\ \boldsymbol{\mu}+\mathbf{n}\in\mathscr{P}_m \text{ and } |\boldsymbol{\mu}+\mathbf{n}|\le\rho\}$. Then

$$F^{\Pi}_{\mathrm{PM}} > \sum_{\boldsymbol{\mu}\in\mathcal{R}_\rho} \left(\sum_{\mathbf{n}} \xi_{\mathbf{n}} \sqrt{\hat{p}_{\mathbf{n}+\boldsymbol{\mu},m} p_{\mathbf{n}+\boldsymbol{\mu},m}} \right)^2 . \tag{B.28}$$

Note that we can drop the prime in the last sum over $\mathbf{n}$. We have

$$F_{\mathrm{PM}}^{\Pi} > p_{\mathbf{0},m}\,\mathrm{e}^{-\frac{\rho^2}{2m\sigma_1^2}} \sum_{\boldsymbol{\mu}\in\mathcal{R}_\rho} \left(\sum_{\mathbf{n}} \xi_{\mathbf{n}} \sqrt{\hat{p}_{\mathbf{n}+\boldsymbol{\mu},m}} \right)^2 , \tag{B.29}$$

where $m\sigma_1^2$ is the smallest eigenvalue of the covariance matrix of $p_{\mathbf{m},m}$. Here, we explicitly display the m dependence of the covariance matrix eigenvalues; thus σ_1^2 does not scale with m. Let us choose the 'guessed' distribution as

$$\hat{p}_{\mathbf{m},m} = \hat{p}_{\mathbf{0},m}\,\mathrm{e}^{-\zeta\frac{|\mathbf{m}|^2}{2m}}; \qquad \sum_{\mathbf{m}\in\mathscr{P}_m} \hat{p}_{\mathbf{m},m} = 1. \tag{B.30}$$

Then,

$$\begin{aligned} \hat{p}_{\mathbf{n}+\boldsymbol{\mu},m} &= \hat{p}_{\mathbf{0},m}\,\mathrm{e}^{-\zeta\frac{|\mathbf{n}+\boldsymbol{\mu}|^2}{2m}} \geq \hat{p}_{\mathbf{0},m}\,\mathrm{e}^{-\zeta\frac{(|\mathbf{n}|+|\boldsymbol{\mu}|)^2}{2m}} \\ &= \mathrm{e}^{-\zeta\frac{|\mathbf{n}|^2+2|\mathbf{n}||\boldsymbol{\mu}|}{2m}}\,\hat{p}_{\boldsymbol{\mu},m} \geq \mathrm{e}^{-\zeta\frac{3|\mathbf{n}|^2+2|\mathbf{n}|\rho}{2m}}\,\hat{p}_{\boldsymbol{\mu},m}, \end{aligned} \tag{B.31}$$

where we have used that $|\boldsymbol{\mu}| \leq |\mathbf{n}| + \rho$ if $\boldsymbol{\mu} \in \mathcal{R}_\rho$. Thus, the following bound holds

$$F_{\mathrm{PM}}^{\Pi} > p_{\mathbf{0},m}\,\mathrm{e}^{-\frac{\rho^2}{2m\sigma_1^2}-\zeta\frac{3\nu^2+2\nu\rho}{2m}} \left(\sum_{\mathbf{n}} \xi_{\mathbf{n}} \right)^2 \sum_{\boldsymbol{\mu}\in\mathcal{R}_\rho} \hat{p}_{\boldsymbol{\mu},m}. \tag{B.32}$$

We now need to lower bound the last sum. For this, we write

$$\sum_{\boldsymbol{\mu}\in\mathcal{R}_\rho} \hat{p}_{\boldsymbol{\mu},m} = 1 - \sum_{\boldsymbol{\mu}\in\overline{\mathcal{R}}_\rho\cap\mathscr{P}_m} \hat{p}_{\boldsymbol{\mu},m}. \tag{B.33}$$

For $\boldsymbol{\mu} \in \overline{\mathcal{R}}_\rho \cap \mathscr{P}_m$ one has $\rho < |\mathbf{n}+\boldsymbol{\mu}| \leq \nu + |\boldsymbol{\mu}|$, then, recalling that $\rho > \nu$,

$$\begin{aligned} \sum_{\boldsymbol{\mu}\in\mathcal{R}_\rho} \hat{p}_{\boldsymbol{\mu},m} &> 1 - \sum_{\boldsymbol{\mu}\in\overline{\mathcal{R}}_\rho\cap\mathscr{P}_m} \hat{p}_{\mathbf{0},m}\,\mathrm{e}^{-\zeta\frac{(\rho-\nu)^2}{2m}} \\ &> 1 - |\mathscr{P}_m|\,\hat{p}_{\mathbf{0},m}\,\mathrm{e}^{-\zeta\frac{(\rho-\nu)^2}{2m}}. \end{aligned} \tag{B.34}$$

Note that $|\mathscr{P}_m| \leq (m+d-1)!/[m!(d-1)!] \sim m^{d-1}$, for large m. With all the above,

$$\begin{aligned} F_{\mathrm{PM}}^{\Pi} &> \left(\max_{\mathbf{m}} p_{\mathbf{m},m} \right) \mathrm{e}^{-\frac{\rho^2}{2m\sigma_1^2}-\zeta\frac{3\nu^2+2\nu\rho}{2m}} \\ &\times \left[1 - Cm^{d-1}\mathrm{e}^{-\zeta\frac{(\rho-\nu)^2}{2m}} \right] \left(\sum_{\mathbf{n}} \xi_{\mathbf{n}} \right)^2 \end{aligned} \tag{B.35}$$

for some positive constant C. Therefore, if $\rho = m^{\frac{1-\epsilon}{2}}$, and $\zeta = m^{\delta}$, with $(1+\epsilon)/2 > \delta > \epsilon > 0$, then $\rho^2/m = m^{-\epsilon}$, $\zeta/m = m^{-1+\delta}$, $\zeta\rho/m = m^{-\frac{1+\epsilon}{2}+\delta}$ and $\zeta\rho^2/m = m^{\delta-\epsilon}$. Thus, for large m we have

$$F^{\Pi}_{\mathrm{PM}} > \left(\max_{\mathbf{m}} p_{\mathbf{m},m}\right)\left(\sum_{\mathbf{n}} \xi_{\mathbf{n}}\right)^2 . \tag{B.36}$$

This result holds provided $p_{\mathbf{m},m}$ is a multivariate normal distribution picked at some $\mathbf{m}_0 \in \mathscr{P}_m$. The actual distribution is multinomial on the points of $\mathscr{P}_m$. However, as m becomes asymptotically large, the induced distribution $p_{\mathbf{m},m}$ becomes arbitrarily close to the multivariate normal assumed in the proof above. Recalling that the right hand side of (B.36) is also an upper bound to F^{Π}_{CL} and, thus to F^{Π}_{PM}, we finally conclude that (we restore the suppressed tildes)

$$F^{\Pi}_{\mathrm{PM}} = F^{\Pi}_{\mathrm{CL}} = \left(\max_{\widetilde{\mathbf{m}}\in\widetilde{\mathscr{P}}_m} p_{\widetilde{\mathbf{m}},m}\right)\left(\sum_{\widetilde{\mathbf{n}}\in\widetilde{\mathscr{P}}_n} \xi_{\widetilde{\mathbf{n}}}\right)^2 \tag{B.37}$$

for asymptotically large m and fixed n. This leads to Eq. (7.12), of which Eq. (7.10) is a particular case for $r = 1$.

B.5 Fidelity calculations

In this section we give some details of the calculation leading to Eqs. (7.13) and (7.14). As already mentioned, Smith vectors $\mathbf{s} \in \mathbb{Z}^r$ [recall Eq. (B.19)] are most suited to this purpose because they form a cubic lattice of unit step size, i.e., their minimal cell has volume $\Delta V^* = 1$. In this sense, they are just a particular instance of vectors $\widetilde{\mathbf{m}} \in \widetilde{\mathscr{P}}_m$. Hence, to avoid further proliferation of notation, we will use here the generic symbols $\widetilde{\boldsymbol{\Sigma}}$ and $\widetilde{\mathscr{P}}_m$ to refer to the covariance matrix of the multivariate Gaussian distribution $p_{\mathbf{s},m}$ and the lattice of the Smith vectors $\mathbf{s}$ respectively. Since $\widetilde{\boldsymbol{\Sigma}}$ scales with m, we write $\widetilde{\boldsymbol{\Sigma}} = m\widetilde{\boldsymbol{\Sigma}}_1$, where $\widetilde{\boldsymbol{\Sigma}}_1$ is independent of m. Then, the expression for the asymptotic fidelity in (7.12) becomes

$$F^{\Pi} = \frac{1}{\sqrt{(2\pi m)^r \det \widetilde{\boldsymbol{\Sigma}}_1}}\left(\sum_{\mathbf{s}\in\widetilde{\mathscr{P}}_n} \sqrt{p_{\mathbf{s},n}\frac{\pi_{\mathbf{s}}}{P_{\mathrm{succ}}}}\right)^2 . \tag{B.38}$$

The last sum can be evaluated in the asymptotic limit of large n (recall however that we assume $n \ll m$), as we show below. In this case, also $p_{\mathbf{s},n}$

approaches a multivariate normal distribution, as that in Eq. (7.11), and the sum over $\widetilde{\mathscr{P}}_n$ can be approximated by an integral over the polytope $\widetilde{\Delta}^r_n$.

As a warmup act, we first compute the fidelity F^Π for the deterministic protocol, i.e., when $P_\text{succ} = 1$. Since no filtering is applied, we have $\pi_\mathbf{s} = 1$ for all $\mathbf{s} \in \widetilde{\mathscr{P}}_n$. The sum in (B.38) simplifies to

$$\int_{\widetilde{\Delta}^r_n} d\mathbf{s}\, \sqrt{p_{\mathbf{s},n}} \simeq \int_{\mathbb{R}^r} d\mathbf{s}\, \sqrt{p_{\mathbf{s},n}} = 2^{r/2}\Big[(2\pi n)^r \det \widetilde{\mathbf{\Sigma}}_1\Big]^{1/4}, \tag{B.39}$$

as $2^{-r/2}(2\pi)^{-r/4}(\det \widetilde{\mathbf{\Sigma}})^{-1/4} \times \sqrt{p_{\mathbf{s},n}}$ is also a properly normalized multivariate normal distribution with covariance matrix $2\widetilde{\mathbf{\Sigma}}$. Substituting in Eq. (B.38) we obtain

$$F = \left(4\frac{n}{m}\right)^{r/2}. \tag{B.40}$$

This expression agrees with Eq. (7.14) in the deterministic limit, when $\eta \to 0$.

In the probabilistic case, $P_\text{succ} < 1$, we need to optimize the filter parameters $\{\pi_\mathbf{s}\}$. To simplify the notation, let us use the definition of the normalized state $|\xi\rangle$, with components $\xi_\mathbf{s} := \sqrt{p_{\mathbf{s},n}\, \pi_\mathbf{s}/P_\text{succ}}$. Then, the maximum fidelity, $F = \max_\Pi F^\Pi$, is obtained when the sum in (B.38) takes its maximum value:

$$\text{max} \qquad \sum_\mathbf{s} \xi_\mathbf{s}\ , \tag{B.41}$$

$$\text{subject to} \qquad \sum_\mathbf{s} \xi_\mathbf{s}^2 = 1 \tag{B.42}$$

$$\text{and} \qquad \xi_\mathbf{s} \le \sqrt{\frac{p_{\mathbf{s},n}}{P_\text{succ}}}, \quad \mathbf{s} \in \widetilde{\mathscr{P}}_n, \tag{B.43}$$

where (B.42) is the normalization constraint, and (B.43) comes from the positivity and trace preserving requirements on the stochastic filter. To solve (B.41), (B.42) and (B.43), we use Lagrange multipliers and the Karush-Kuhn-Tucker conditions. The problem reduces to solving the stationary conditions

$$\begin{aligned}\frac{\partial}{\partial \xi_\mathbf{s}}\left(\sum_{\mathbf{s}'} \xi_{\mathbf{s}'}\right) \;=\; & \frac{\partial}{\partial \xi_\mathbf{s}}\left[\frac{1}{2\zeta}\left(\sum_{\mathbf{s}'} \xi_{\mathbf{s}'}^2 - 1\right)\right. \\ & \left. + \sum_{\mathbf{s}'} \sigma_{\mathbf{s}'}\left(\xi_{\mathbf{s}'} - \sqrt{\frac{p_{\mathbf{s}',n}}{P_\text{succ}}}\right)\right],\end{aligned} \tag{B.44}$$

where the sums extend to $\mathbf{s} \in \widetilde{\mathscr{P}}_n$, and $1/(2\zeta)$ and $\{\sigma_\mathbf{s}\}_{\mathbf{s}\in\widetilde{\mathscr{P}}_n}$ are multipliers. Conditions (B.42) and (B.43) are called primal feasibility conditions. In addition, one has to impose that $\sigma_{\mathbf{s}} \geq 0$, known as dual feasibility condition, and

$$\sigma_{\mathbf{s}} \left(\xi_{\mathbf{s}} - \sqrt{\frac{p_{\mathbf{s},n}}{P_{\mathrm{succ}}}} \right) = 0, \tag{B.45}$$

known as complementary slackness condition, both for all $\mathbf{s} \in \widetilde{\mathscr{P}}_n$. The latter, implies that at any site of $\widetilde{\mathscr{P}}_n$, either $\sigma_{\mathbf{s}} = 0$ or $\xi_{\mathbf{s}} = \sqrt{p_{\mathbf{s},n}/P_{\mathrm{succ}}}$, in which case we say that $\mathbf{s}$ belongs to the coincidence set $\mathscr{C}$, i.e., $\mathscr{C} := \{\mathbf{s} \in \widetilde{\mathscr{P}}_n : \xi_{\mathbf{s}}^2 = p_{\mathbf{s},n}/P_{\mathrm{succ}}\}$. If $\mathbf{s} \notin \mathscr{C}$, Eq. (B.44) readily gives the constant solution $\xi_{\mathbf{s}} = \zeta$.

If $P_{\mathrm{succ}} < \min_{\mathbf{s}\in\widetilde{\mathscr{P}}_n} p_{\mathbf{s},n}$, then $p_{\mathbf{s},n}/P_{\mathrm{succ}} > 1 \geq \xi_{\mathbf{s}}^2$, thus $\mathscr{C} = \emptyset$. In this case, normalization implies $\zeta = |\widetilde{\mathscr{P}}_n|^{-1/2}$, where $|\widetilde{\mathscr{P}}_n|$ is the number of sites in $\widetilde{\mathscr{P}}_n$. Substituting in (B.38), we have

$$F = \frac{|\widetilde{\mathscr{P}}_n|}{\sqrt{(2\pi m)^r \det \widetilde{\boldsymbol{\Sigma}}_1}} \,. \tag{B.46}$$

Recalling Eq. (B.23) and $m\widetilde{\boldsymbol{\Sigma}}_1 = \widetilde{\boldsymbol{\Sigma}}$, this equation becomes Eq. (7.13), which holds for any choice of vectors $\tilde{\mathbf{n}}$. Notice that both, Eqs. (B.46) and (7.13), also hold for small n. For large n, $|\widetilde{\mathscr{P}}_n|\Delta V^*$ (recall $\Delta V^* = 1$ for Smith vectors) approaches V_n, the volume of the polytope $\tilde{\Delta}_n^r$, as the irregularities or defects of the lattice $\widetilde{\mathscr{P}}_n$ can only arise within a finite distance from its boundary (see previous sections). Closed formulas for $|\widetilde{\mathscr{P}}_n|$ or V_n depend on the Hamiltonian H and do not seem to generalize easily. In the main text, only the obvious scalings $|\widetilde{\mathscr{P}}_n| \sim V_n \sim n^r$ are used to show that $F \sim (n/\sqrt{m})^r$.

For $P_{\mathrm{succ}} > \min_{\mathbf{s}\in\widetilde{\mathscr{P}}_n} p_{\mathbf{s},n}$ the problem becomes more involved, as the constraint (B.43) is now non-trivial and $\mathscr{C} \neq \emptyset$. Since $p_{\mathbf{s},n}/P_{\mathrm{succ}}$ is bell-shaped, the complement of the coincidence set is the 'ellipsoid' $\overline{\mathscr{C}}_\alpha := \{\mathbf{s} \in \widetilde{\mathscr{P}}_n : (\mathbf{s} - \mathbf{s}_0)^t \widetilde{\boldsymbol{\Sigma}}^{-1} (\mathbf{s} - \mathbf{s}_0) \leq \alpha^2$, for some α (Fig. B.1 shows a two-dimensional version of it), and we have the solution

$$\xi_{\mathbf{s}} = \begin{cases} \sqrt{p_{\alpha,n}/P_{\mathrm{succ}}} & \text{if} \quad \mathbf{s} \in \overline{\mathscr{C}}_\alpha \\ \sqrt{p_{\mathbf{s},n}/P_{\mathrm{succ}}} & \text{if} \quad \mathbf{s} \notin \overline{\mathscr{C}}_\alpha \,, \end{cases} \tag{B.47}$$

where we have defined

$$p_{\alpha,n} := \frac{\mathrm{e}^{-\alpha^2/2}}{\sqrt{(2\pi)^r \det \widetilde{\boldsymbol{\Sigma}}}}, \tag{B.48}$$

and we note that the parameter α that gives the size of the 'ellipsoid' $\overline{\mathscr{C}}_\alpha$ is determined by normalization: $1 = \sum_{\mathbf{s}} \xi_{\mathbf{s}}^2 \simeq \int d^r\mathbf{s}\, \xi_{\mathbf{s}}^2$. Note also that the

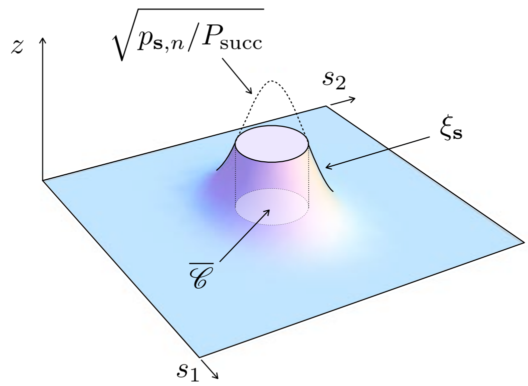


Figure B.1. Plot of $\xi_{\mathbf{s},n}$ (z axis) for large n (the truncated bell-shaped surface). The figure also shows the complement of the coincidence set and $\sqrt{p_{\mathbf{s},n}/P_{\text{succ}}}$ for $P_{\text{succ}} > \min_{\mathbf{s}\in\widetilde{\mathscr{P}}_n} p_{\mathbf{s},n}$.

solution $\xi_{\mathbf{s}}$ is a 'truncated' multivariate normal distribution, as shown in Fig. B.1. The above integral thus splits into two straightforward ones over the two regions in (B.47), and we obtain the relation:

$$P_{\text{succ}} = \frac{1}{\Gamma(\frac{r}{2})}\left(\Gamma(\tfrac{r}{2},\tfrac{\alpha^2}{2}) + \frac{\alpha^r \mathrm{e}^{-\alpha^2/2}}{2^{r/2-1}r}\right), \tag{B.49}$$

where $\Gamma(a,x)$ is the upper incomplete Gamma function. Proceeding along the same line, one can compute $\sum_{\mathbf{s}}\xi_{\mathbf{s}}$ to obtain

$$F = \frac{1}{\Gamma(\frac{r}{2}+1)}\left(\frac{n}{2m}\right)^{r/2}\frac{\left[2^{r-1}r\Gamma(\frac{r}{2},\frac{\alpha^2}{4}) + \alpha^r \mathrm{e}^{-\alpha^2/4}\right]^2}{2^{r/2-1}r\Gamma(\frac{r}{2},\frac{\alpha^2}{2}) + \alpha^r \mathrm{e}^{-\alpha^2/2}}. \tag{B.50}$$

Eqs. (B.49) and (B.50) give the solution to our optimization problem in parametric form, in terms of α. Finding the fidelity F as a explicit function of P_{succ} would require inverting the relation (B.49), which cannot be done analytically for an arbitrary success probability. We can however obtain an analytic expression of F for a success probability close to one, i.e., for $P_{\text{succ}} = 1-\eta$, $\eta \ll 1$, by simply expanding the right hand side of Eq. (B.49) to leading order in α. Such expansion reads:

$$\eta = 1 - P_{\text{succ}} = \frac{\alpha^{2+r}}{2^{1+r/2}\,\Gamma(2+r/2)} + \mathcal{O}(\alpha^{r+3}). \tag{B.51}$$

Solving for α and substituting in the expansion (at leading order in α) of the right hand side of Eq. (B.50) one obtains

$$F = \left(4\frac{n}{m}\right)^{r/2}\left[1 + (1-2^{-r/2})\,\eta\right] + \mathcal{O}(\eta^2). \tag{B.52}$$

Eqs. (B.49) through (B.52) hold provided P_{succ} does not exponentially vanish with n. If it does, as in (7.13), where we assumed that $P_{\text{succ}} < \min_{\mathbf{s}\in\widetilde{\mathscr{P}}_n} p_{\mathbf{s},n}$, we obtain the better scaling $F \sim (n/\sqrt{m})^r$. The reason for this different scaling is that a multivariate normal distribution only approximates $p_{\mathbf{s},n}$ accurately around its peak, whereas it falls off exponentially with n at the tails (where $\min_{\mathbf{s}\in\widetilde{\mathscr{P}}_n} p_{\mathbf{s},n}$ lies).

B.6 PM approximation of probabilistic processes with permutationally invariant output

Here we prove that for every quantum operation $\mathscr{C}_m$, with input in $\mathcal{H}_{\text{in}}$ and permutationally invariant output in $\mathscr{H}^{\otimes m}$ there exists a PM protocol, described by a quantum operation $\widetilde{\mathscr{C}_m}$, such that *i)* $\mathscr{C}_m$ and $\widetilde{\mathscr{C}_m}$ have the same success probability and *ii)* the error probability in distinguishing between $\mathscr{C}_m$ and $\widetilde{\mathscr{C}_m}$ by inputting a state ρ and measuring k output systems is lower bounded by $p_{\text{err}} \leq \frac{1}{2} + (kd^2)/[2mP_{\text{succ}}(\rho)]$, where $P_{\text{succ}}(\rho)$ is the probability that the operation $\mathscr{C}_m$ takes place on input ρ.

The proof follows the same lines of the proofs of theorems 4 and 5 in Ref. **?**. Let $\mathscr{C}_m$ be a quantum operation transforming states on $\mathcal{H}_{\text{in}}$ into sates on $\mathcal{H}^{\otimes m}$. First, we suppose that the output states of $\mathscr{C}_m$ have support contained in the symmetric subspace of $\mathcal{H}^{\otimes m}$. In this case, it is useful to consider the universal measure-and-prepare channel $\mathcal{M}$ defined by

$$\mathscr{M}(\rho) = \int \mathrm{d}\psi\; d_m^+ \,\mathrm{tr}\,[\psi^{\otimes m}\rho]\;\psi^{\otimes m}$$

where $\mathrm{d}\psi$ is the invariant measure over the pure states, $d_m^+ = \binom{m+d-1}{d-1}$ is the dimension of the symmetric subspace. Note that the map $\mathscr{M}$ is trace-preserving for all states with support in the symmetric subspace. Moreover, $\mathscr{M}$ provides a good approximation of the partial trace **?**:

$$\|\mathrm{tr}_{\,m-k} - \mathrm{tr}_{\,m-k} \circ \mathscr{M}\|_\diamond \leq \frac{2kd}{m}\,, \tag{B.53}$$

where $\|\mathscr{L}\|_\diamond$ denotes the diamond norm of a linear, Hermitian-preserving map $\mathscr{L}$, defined as

$$\|\mathscr{L}\|_\diamond := \sup_{r\in\mathbb{N}} \sup_{\substack{|\Psi\rangle\in\mathcal{H}_{\text{in}}\otimes\mathbb{C}^r \\ \||\Psi\rangle\|=1}} \mathrm{tr}\,\left|(\mathscr{L}\otimes\mathscr{I}_r)(|\Psi\rangle\langle\Psi|)\right|\,,$$

$\mathscr{I}_r$ denoting the identity map for an r-dimensional quantum system. Using Eq. (B.53), it is immediate to construct the desired PM protocol. The protocol consists in performing the quantum operation $\mathscr{C}_m$ and subsequently applying the measure-and-prepare channel $\mathscr{M}$. Mathematically, it is described by the quantum operation $\widetilde{\mathscr{C}_m} = \mathscr{M} \circ \mathscr{C}_m$. Since $\mathscr{M}$ is trace-preserving (in the symmetric subspace), one has

$$\mathrm{tr}[\widetilde{\mathscr{C}_m}(\rho)] = \mathrm{tr}[\mathscr{C}_m(\rho)]$$

for every input state ρ, that is, $\widetilde{\mathscr{C}_m}$ and $\mathscr{C}_m$ have the same success probability. Moreover, one has the bound

$$\begin{aligned}\left\| \mathrm{tr}_{\,m-k} \circ \mathscr{C}_m - \mathrm{tr}_{\,m-k} \circ \widetilde{\mathscr{C}_m} \right\|_\diamond & \\ &\leq \| \left(\mathrm{tr}_{\,m-k} - \mathrm{tr}_{\,m-k} \circ \mathscr{M}\right) \mathscr{C}_m \|_\diamond \\ &\leq \|\mathrm{tr}_{\,m-k} - \mathrm{tr}_{\,m-k} \circ \mathscr{M}\|_\diamond \|\mathscr{C}_m\|_\diamond \\ &\leq \frac{2kd}{m}\,, \end{aligned} \tag{B.54}$$

having used the fact that $\|\mathscr{C}_m\|_\diamond \leq 1$ by definition. In words, the k-copy restrictions of $\widetilde{\mathscr{C}_m}$ and $\mathscr{C}_m$ are close to each other provided that $k \ll m$.

Now, suppose that the output of $\mathscr{C}_m$ has support outside the symmetric subspace. In this case, the invariance under permutations implies that $\mathscr{C}_m$ has a Stinespring dilation of the form

$$\mathscr{C}_m = \mathrm{tr}_{\,\mathcal{H}_E} \circ \mathscr{K}_m \qquad \mathscr{K}(\rho) := K_m \rho K_m^\dagger,$$

where $\mathcal{H}_E = \mathcal{H}^{\otimes m}$ and K_m is an operator with range contained in the symmetric subspace of $\mathcal{H}^{\otimes m} \otimes \mathcal{H}_E \simeq (\mathcal{H} \otimes \mathcal{H})^{\otimes m}$ (for a proof see e.g. **?**). Hence, we can take the measure-and-prepare quantum operation $\widetilde{\mathscr{K}_m} := \mathscr{M}_E \circ \mathscr{K}_m$, where $\mathscr{M}_E$ is the universal measure-and-prepare channel on $\mathcal{H}^{\otimes m} \otimes \mathcal{H}_E$, and we can define $\widetilde{\mathscr{C}_m} := \mathrm{tr}_{\,\mathcal{H}_E} \circ \widetilde{\mathscr{K}_m}$. By definition, the success probability of $\widetilde{\mathscr{C}_m}$ is equal to the success probability of $\mathscr{C}_m$. Moreover, one has the relation

$$\begin{aligned}\left\| \mathrm{tr}_{\,m-k} \circ \mathscr{C}_m - \mathrm{tr}_{\,m-k} \circ \widetilde{\mathscr{C}_m} \right\|_\diamond & \\ &= \left\| \mathrm{tr}_{\,m-k} \circ \mathrm{tr}_{\,\mathcal{H}_E} \circ \left(\mathscr{K}_m - \widetilde{\mathscr{K}_m}\right) \right\|_\diamond \\ &= \left\| \mathrm{tr}_{\,\mathcal{H}_E^k} \circ \left(\mathrm{tr}_{\,m-k} \otimes \mathrm{tr}_{\,\mathcal{H}_E^{m-k}}\right) \circ \left(\mathscr{K}_m - \widetilde{\mathscr{K}_m}\right) \right\|_\diamond \end{aligned}$$

where $\mathrm{tr}_{\,\mathcal{H}_E^k}$ denotes the partial trace over k ancillary Hilbert spaces. Hence, one gets the bound

$$\begin{aligned}
&\left\|\operatorname{tr}_{m-k} \circ \mathscr{C}_m - \operatorname{tr}_{m-k} \circ \widetilde{\mathscr{C}_m}\right\|_\diamond \\
&\qquad \le \left\|\operatorname{tr}_{\mathcal{H}_E^k}\right\|_\diamond \left\|\left(\operatorname{tr}_{m-k} \otimes \operatorname{tr}_{\mathcal{H}_E^{m-k}}\right)\left(\mathscr{K}_m - \widetilde{\mathscr{K}_m}\right)\right\|_\diamond \\
&\qquad \le \frac{2d^2 k}{m}\,, \qquad (\mathrm{B.55})
\end{aligned}$$

having used Eq. (B.54) with $\widetilde{\mathscr{C}_m}$, $\mathscr{C}_m$, and d replaced by $\widetilde{\mathscr{K}_m}$, $\mathscr{K}_m$, and d^2, respectively.

Finally, the error probability in distinguishing between $\widetilde{\mathscr{C}_m}$ and $\mathscr{C}_m$ by inputting a state ρ and measuring k output system is equal to the error probability in distinguishing between the two states

$$\widetilde{\rho}_{m,k} = \frac{\operatorname{tr}_{M-k}[\widetilde{\mathscr{C}_m}(\rho)]}{\operatorname{tr}[\widetilde{\mathscr{C}_m}(\rho)]} \quad \text{and} \quad \rho_{m,k} = \frac{\operatorname{tr}_{M-k}[\mathscr{C}_m(\rho)]}{\operatorname{tr}[\mathscr{C}_m(\rho)]}\,,$$

respectively. Assuming equal prior probabilities for the two quantum operations, Helstrom theorem gives the bound $p_{\mathrm{err}} = \frac{1}{2}\left[1 + \frac{1}{2}\|\widetilde{\rho}_{m,k} - \rho_{m,k}\|_1\right]$ and therefore we have

$$\begin{aligned}
p_{\mathrm{err}} &\le \frac{1}{2}\left[1 + \frac{\|\operatorname{tr}_{m-k} \circ (\widetilde{\mathscr{C}_m} - \mathscr{C}_m)\|_\diamond}{2P_{\mathrm{succ}}(\rho)}\right] \\
&\le \frac{1}{2}\left[1 + \frac{kd^2}{mP_{\mathrm{succ}}(\rho)}\right],
\end{aligned}$$

where $P_{\mathrm{succ}}(\rho) := \operatorname{tr}[\mathscr{C}_m(\rho)] \equiv \operatorname{tr}[\widetilde{\mathscr{C}_m}]$. ■

B.7 Approximation of the optimal k-copy cloning fidelity

Suppose that we are given n copies of the state $|\psi_x\rangle \in \mathcal{H}$, $x \in \mathsf{X}$ and that we want to produce m approximate copies, whose quality is assessed by checking a random group of k output systems. Let $\mathscr{C}_{n,m}$ be the quantum operation describing the cloning process. Since the k systems are chosen at random, we can restrict our attention to quantum operations that are invariant under permutation of the output spaces. Conditional on the occurrence of the quantum operation $\mathscr{C}_{n,m}$ and on preparation of the input $|\psi_x\rangle$, the k-copy cloning fidelity is given by

$$
\begin{aligned}
F_{k,x}[\mathscr{C}_{n,m}] &= \frac{O_{k,x}[\mathscr{C}_{n,m}]}{P_x[\mathscr{C}_{n,m}]} \\
O_{k,x}[\mathscr{C}_{n,m}] &:= \langle\varphi_x|^{\otimes k} \operatorname{tr}_{m-k}[\mathscr{C}_{n,m}\left(\varphi_x^{\otimes n}\right)] \, |\varphi_x\rangle^{\otimes k} \\
P_x[\mathscr{C}_{n,m}] &:= \operatorname{tr}\left[\mathscr{C}_{n,m}\left(\varphi_x^{\otimes n}\right)\right],
\end{aligned}
\tag{B.56}
$$

where ψ_x denotes the rank-one projector $\psi_x := |\psi_x\rangle\langle\psi_x|$. Now, constructing the quantum operation $\widetilde{\mathscr{C}_{n,m}}$ as in theorem 1 and using Eq. (B.55) we have

$$
\begin{aligned}
\left|O_{k,x}[\mathscr{C}_{n,m}] - O[\widetilde{\mathscr{C}_{n,m}}]\right| &\leq \frac{2kd^2}{m} \\
P_x[\mathscr{C}_{n,m}] = P_x[\widetilde{\mathscr{C}_{n,m}}] &\equiv P_{\text{succ}}\left(\psi_x^{\otimes n}\right),
\end{aligned}
$$

and, therefore,

$$
\left|F_{k,x}[\mathscr{C}_{n,m}] - F_{k,x}[\widetilde{\mathscr{C}_{n,m}}]\right| \leq \frac{2kd^2}{mP_{\text{succ}}\left(\psi_x^{\otimes n}\right)} \qquad \forall x \in \mathsf{X}\,.
$$

In conclusion, as long as the probability of success $P_{\text{succ}}(\psi_x^{\otimes n})$ is lower bounded by a finite value independent of m, the k-copy fidelities of the cloning processes $\mathscr{C}_{n,m}$ and $\widetilde{\mathscr{C}_{n,m}}$. Since the bound holds for arbitrary quantum operations, in particular it holds for the quantum operation describing the optimal cloner with given probability of success.

It is immediate to extend the derivation to the Bayesian scenario where the input state $|\psi_x\rangle^{\otimes n}$ is given with probability p_x and one considers the average fidelity and average success probability. Indeed, the average k-copy fidelity is given by

$$
\begin{aligned}
F_k[\mathscr{C}_{n,m}] &= \frac{O_k[\mathscr{C}_{n,m}]}{P[\mathscr{C}_{n,m}]} \\
O_k[\mathscr{C}_{n,m}] &:= \sum_x p_x \, O_{k,x}[\mathscr{C}_{n,m}] \\
P[\mathscr{C}_{n,m}] &:= \sum_x p_x \, P_x[\mathscr{C}_{n,m}]
\end{aligned}
$$

and one has the bound

$$
F_k[\mathscr{C}_{n,m}] \leq \frac{2kd^2}{mP_{\text{succ}}}\,,
$$

P_{succ} being the average success probability. Again, for every fixed value of the success probability, the fidelity of the optimal cloner is achieved by a PM protocol.

B.8 Lower bound on the average probability of success

We now show that for every fixed n, the probability of success of the optimal cloner is lower bounded by a finite value. Precisely, we prove the following

Lemma 1. *The quantum operation $\mathscr{C}^*_{n,m}$ corresponding to the n-to-m cloner that maximizes the fidelity in Eq. (B.56) can be chosen without loss of generality to have success probability $P^*_{\rm succ}$ equal to $1/\|\tau^{-1}\|_\infty$, where τ is the average input state $\tau := \sum_x p_x \varphi_x^{\otimes n}$ and $\|\tau^{-1}\|_\infty$ is the maximum eigenvalue of τ^{-1}.*

Proof. For a generic quantum operation $\mathscr{C}_{n,m}$, the k-copy fidelity can be expressed in terms of its Choi operator $C_{n,m}$ as

$$
\begin{aligned}
F_k[\mathscr{C}_{n,m}] &= \frac{\mathrm{tr}\,[\Omega\, C_{n,m}]}{\mathrm{tr}\,[(I^{\otimes m} \otimes \bar{\tau}) C_{n,m}]} \\
\Omega &:= \frac{1}{\binom{m}{k}} \sum_{\mathsf{S}} \sum_x p_x \left(\psi_x^{\otimes k}\right)_{\mathsf{S}} \otimes \bar{\psi}_x^{\otimes n} ,
\end{aligned}
$$

where the outer summation runs over all k-element subsets S of the output Hilbert spaces, $\left(\psi_x^{\otimes k}\right)_{\mathsf{S}}$ denotes the operator $\psi_x^{\otimes k}$ acting on the Hilbert spaces in the set S, and $\bar{\tau}$ $(\bar{\psi}_x)$ is the complex conjugate of τ (ψ_x). Following the arguments of Fiurášek [2004]; Chiribella and Xie [2013], one can easily show that the maximum fidelity over all quantum operations is given by

$$
F^*_k = \left\| \left(I^{\otimes m} \otimes \bar{\tau}^{-\frac{1}{2}}\right) \Omega \left(I^{\otimes m} \otimes \bar{\tau}^{-\frac{1}{2}}\right) \right\|_\infty .
$$

The maximum is achieved by choosing a Choi operator $C^*_{n,m}$ of the form

$$
C^*_{n,m} = \gamma \ \left(I^{\otimes m} \otimes \bar{\tau}^{-\frac{1}{2}}\right) |\Psi\rangle \langle\Psi| \left(I^{\otimes m} \otimes \bar{\tau}^{-\frac{1}{2}}\right)
$$

where $\gamma \geq 0$ is a proportionality constant and $|\Psi\rangle$ is the eigenvector of $\left(I^{\otimes m} \otimes \bar{\tau}^{-\frac{1}{2}}\right) \Omega \left(I^{\otimes m} \otimes \bar{\tau}^{-\frac{1}{2}}\right)$ with maximum eigenvalue. With this choice, the success probability $P[\mathscr{C}^*_{n,m}]$ is given by

$$
\begin{aligned}
P[\mathscr{C}^*_{n,m}] &= \mathrm{tr}\,[C^*_{n,m} \left(I^{\otimes m} \otimes \bar{\tau}\right)] \\
&= \gamma .
\end{aligned}
$$

We now show that γ can be always chosen to be larger than $1/\|\tau^{-1}\|_\infty$. To this purpose, note that the only constraint on γ is that the quantum operation $\mathscr{C}^*_{n,m}$ must be trace non-increasing. Now, for a generic state ρ one has

$$\begin{aligned}
\mathrm{tr}\,[\mathscr{C}^*_{n,m}(\rho)] &= \gamma\ \mathrm{tr}\,[C^*_{n,m}\left(I^{\otimes M} \otimes \bar{\rho}\right)] \\
&= \gamma\ \langle\Psi| \left(I^{\otimes M} \otimes \bar{\tau}^{-\frac{1}{2}}\bar{\rho}\bar{\tau}^{-\frac{1}{2}}\right)|\Psi\rangle \\
&\leq \gamma\ \|\bar{\tau}^{-\frac{1}{2}}\bar{\rho}\bar{\tau}^{-\frac{1}{2}}\|_\infty \\
&\leq \gamma\ \|\tau^{-1}\|_\infty \|\rho\|_\infty \\
&\leq \gamma\ \|\tau^{-1}\|_\infty .
\end{aligned}$$

Hence, the choice $\gamma = 1/\|\tau^{-1}\|_\infty$ leads to a legitimate (trace non-increasing) quantum operation. ■

Bibliography

Arrad, G., Vinkler, Y., Aharonov, D., and Retzker, A. "Increasing Sensing Resolution with Error Correction". *Physical Review Letters*, 112(15) 150801 (2014).

Aspachs, M., Calsamiglia, J., Muñoz Tapia, R., and Bagan, E. "Phase estimation for thermal Gaussian states". *Physical Review A*, 79(3) (2009).

Bacon, D., Chuang, I.L., and Harrow, A.W. "Efficient Quantum Circuits for Schur and Clebsch-Gordan Transforms". *Physical Review Letters*, 97(17) 170502 (2006).

Bae, J. and Acín, A. "Asymptotic Quantum Cloning Is State Estimation". *Physical Review Letters*, 97(July) 030402 (2006).

Bagan, E., Baig, M., Brey, A., Munoz-Tapia, R., and Tarrach, R. "Optimal strategies for sending information through A quantum channel". *Physical review letters*, 85(24) 5230–3 (2000).

Bagan, E., Baig, M., and Muñoz Tapia, R. "Aligning Reference Frames with Quantum States". *Physical Review Letters*, 87(25) 1–4 (2001).

Bagan, E., Baig, M., and Muñoz Tapia, R. "Quantum reverse engineering and reference-frame alignment without nonlocal correlations". *Physical Review A*, 70(3) 030301 (2004).

Bagan, E., Ballester, M.A., Gill, R.D., Monras, A., and Muñoz Tapia, R. "Optimal full estimation of qubit mixed states". *Physical Review A*, 73(3) 032301 (2006).

Bagan, E., Monras, A., and Muñoz Tapia, R. "Comprehensive analysis of quantum pure-state estimation for two-level systems". *Physical Review A*, 71(6) 1–15 (2005).

Bagan, E., Muñoz Tapia, R., Olivares-Rentería, G.A., and Bergou, J.A. "Optimal discrimination of quantum states with a fixed rate of inconclusive outcomes". *Physical Review A*, 86 1–4 (2012).

Ban, M., Kurokawa, K., Momose, R., and Hirota, O. "Optimum measurements for discrimination among symmetric quantum states and parameter estimation". *International Journal of Theoretical Physics*, 36(6) 1269–1288 (1997).

Bernardo, J.M. and Smith, A.F.M. *Bayesian Theory*. Wiley Series in Probability and Statistics. John Wiley & Sons, Inc., Hoboken, NJ, USA (1994).

Berry, D.W., Hall, M.J.W., Zwierz, M., and Wiseman, H.M. "Optimal Heisenberg-style bounds for the average performance of arbitrary phase estimates". *Physical Review A*, 86(5) 053813 (2012).

Boixo, S., Flammia, S.T., Caves, C.M., and Geremia, J. "Generalized Limits for Single-Parameter Quantum Estimation". *Physical Review Letters*, 98(9) 090401 (2007).

Bombieri, E. and Vaaler, J. "On Siegel's lemma". *Inventiones Mathematicae*, 73(1) 11–32 (1983).

Borkar, V.S. *Probability Theory*. Universitext. Springer New York, New York, NY (1995).

Borosh, I. "A Sharp Bound for Positive Solutions of Homogeneous Linear Diophantine Equations". *Proceedings of the American Mathematical Society*, 60(1) 19 (1976).

Boyd, S. and Vandenberghe, L. *Convex optimization*. Cambridge University Press, Cambridge (2004).

Braunstein, S.L. and Caves, C.M. "Statistical distance and the geometry of quantum states". *Physical Review Letters*, 72(22) 3439–3443 (1994).

Bruss, D., Ekert, A., and Macchiavello, C. "Optimal universal quantum cloning and state estimation". *Physical Review Letters*, 81(12) 4 (1997).

Caves, C.M., Fuchs, C.a., and Schack, R. "Unknown quantum states: The quantum de Finetti representation". *Journal of Mathematical Physics*, 43(2002) 4537–4559 (2002).

Chaves, R., Brask, J.B., Markiewicz, M., Kołodyński, J., and Acín, A. "Noisy Metrology beyond the Standard Quantum Limit". *Physical Review Letters*, 111(12) 120401 (2013).

Chin, A.W., Huelga, S.F., and Plenio, M.B. "Quantum Metrology in Non-Markovian Environments". *Physical Review Letters*, 109(23) 233601 (2012).

Chiribella, G. *Theory of Quantum Computation, Communication, and Cryptography*, volume 6519 of *Lecture Notes in Computer Science*. Springer Berlin Heidelberg, Berlin, Heidelberg (2011).

Chiribella, G. and D'Ariano, G.M. "Quantum Information Becomes Classical When Distributed to Many Users". *Physical Review Letters*, 97(25) 250503 (2006).

Chiribella, G. and Xie, J. "Optimal Design and Quantum Benchmarks for Coherent State Amplifiers". *Physical Review Letters*, 110(21) 213602 (2013).

Chiribella, G. and Yang, Y. "Optimal asymptotic cloning machines". *New Journal of Physics*, 16(6) 063005 (2014).

Chiribella, G., Yang, Y., and Yao, A.C.C. "Quantum replication at the Heisenberg limit". *Nature Communications*, 4 2915 (2013).

Choi, M.D. "Completely positive linear maps on complex matrices". *Linear Algebra and its Applications*, 10(3) 285–290 (1975).

Cirac, J.I., Ekert, A.K., and Macchiavello, C. "Optimal Purification of Single Qubits". *Physical Review Letters*, 82(21) 4344–4347 (1999).

Combes, J. and Ferrie, C. "Cost of postselection in decision theory". *Physical Review A*, 92 022117 (2015).

Combes, J., Ferrie, C., Jiang, Z., and Caves, C.M. "Quantum limits on postselected, probabilistic quantum metrology". *Physical Review A*, 89(5) 052117 (2014).

Cover, T.M. and Thomas, J.A. *Elements of Information Theory*. Wiley-Interscience (2006).

Croke, S., Andersson, E., Barnett, S.M., Gilson, C.R., and Jeffers, J. “Maximum Confidence Quantum Measurements”. *Physical Review Letters*, 96(7) 070401 (2008).

De Finetti, B. *Theory of probability.* John Wiley & Sons (1990).

de Finetti, B. *Philosophical Lectures on Probability*, volume 340 of *Synthese Library.* Springer Netherlands, Dordrecht (2008).

Demkowicz-Dobrzański, R., Kołodyński, J., and Guţă, M. “The elusive Heisenberg limit in quantum-enhanced metrology.” *Nature communications*, 3 1063 (2012).

Dieks, D. “Overlap and distinguishability of quantum states”. *Physics Letters A*, 126(5-6) 303–306 (1988).

Duan, L.M. and Guo, G.C. “Probabilistic Cloning and Identification of Linearly Independent Quantum States”. *Physical Review Letters*, 80 4999–5002 (1998).

Dür, W., Skotiniotis, M., Fröwis, F., and Kraus, B. “Improved Quantum Metrology Using Quantum Error Correction”. *Physical Review Letters*, 112(8) 080801 (2014).

Edmonds, A.R. *Angular Momentum in Quantum Mechanics.* Princeton University Press, Princeton, New Jersey (1957).

Escher, B.M., de Matos Filho, R.L., and Davidovich, L. “General framework for estimating the ultimate precision limit in noisy quantum-enhanced metrology”. *Nature Physics*, 7(5) 406–411 (2011).

Ferreyrol, F., Barbieri, M., Blandino, R., Fossier, S., Tualle-Brouri, R., and Grangier, P. “Implementation of a Nondeterministic Optical Noiseless Amplifier”. *Physical Review Letters*, 104(March) 123603 (2010).

Fiurášek, J. “Optimal probabilistic cloning and purification of quantum states”. *Physical Review A*, 70(3) 032308 (2004).

Fiurášek, J. “Optimal probabilistic estimation of quantum states”. *New Journal of Physics*, 8(9) 192–192 (2006).

Giovannetti, V., Lloyd, S., and Maccone, L. “Quantum Metrology”. *Physical Review Letters*, 96(1) 010401 (2006).

Giovannetti, V. and Maccone, L. “Sub-Heisenberg Estimation Strategies Are Ineffective”. *Physical Review Letters*, 108(May) 210404 (2012).

Gühne, O. and Tóth, G. “Entanglement detection”. *Physics Reports*, 474(1-6) 1–75 (2009).

Hayashi, a., Hashimoto, T., and Horibe, M. “State discrimination with error margin and its locality”. *Physical Review A*, 78(1) 012333 (2008).

Hayashi, M. “Comparison Between the Cramer-Rao and the Mini-max Approaches in Quantum Channel Estimation”. *Communications in Mathematical Physics*, 304 689–709 (2011).

Helstrom, C.W. *Quantum Detection and Estimation Theory.* Academic Press, New York (1976).

Holevo, A. *Probabilistic and Statistical Aspects of Quantum Theory.* North-Holland, Amsterdam (1982).

Huelga, S.F., Macchiavello, C., Pellizzari, T., Ekert, A.K., Plenio, M.B., and Cirac, J.I. “Improvement of Frequency Standards with Quantum Entanglement”. *Physical Review Letters*, 79(20) 3865–3868 (1997).

Ivanovic, I. “How to differentiate between non-orthogonal states”. *Physics Letters A*, 123(6) 257–259 (1987).

Jaeger, G. and Shimony, A. “Optimal distinction between two non-orthogonal quantum states”. *Physics Letters A*, 197(2) 83–87 (1995).

Jarzyna, M. and Demkowicz-Dobrzański, R. “True precision limits in quantum metrology”. *New Journal of Physics*, 17(1) 013010 (2015).

Jeske, J., Cole, J.H., and Huelga, S.F. “Quantum metrology subject to spatially correlated Markovian noise: restoring the Heisenberg limit”. *New Journal of Physics*, 16(7) 073039 (2014).

Jones, G.A. and Jones, J.M. *Elementary Number Theory.* Springer (1998).

Katok, A. and Hasselblatt, B. *Introduction to the modern theory of dynamical systems.* Cambridge University Press, Cambridge (1996).

Kay, S. “Fundamentals of Statistical Signal Processing: Estimation Theory, 1993” (1993).

Kessler, E.M., Lovchinsky, I., Sushkov, A.O., and Lukin, M.D. “Quantum Error Correction for Metrology”. *Physical Review Letters*, 112(15) 150802 (2014).

Knysh, S., Smelyanskiy, V.N., and Durkin, G.a. “Scaling laws for precision in quantum interferometry and the bifurcation landscape of the optimal state”. *Physical Review A*, 83(2) 021804 (2011).

Knysh, S.I., Chen, E.H., and Durkin, G.A. “True Limits to Precision via Unique Quantum Probe”. *arXiv: 1402.0495* (2014).

Kocsis, S., Xiang, G.Y., Ralph, T.C., and Pryde, G.J. “Heralded noiseless amplification of a photon polarization qubit”. *Nature Physics*, 9(1) 23–28 (2012).

Kołodyński, J. and Demkowicz-Dobrzański, R. “Phase estimation without a priori phase knowledge in the presence of loss”. *Physical Review A*, 82 053804 (2010).

Kraus, K. “General state changes in quantum theory”. *Annals of Physics*, 64(2) 311–335 (1971).

Marcus, M. and Minc, H. *A survey of matrix theory and matrix inequalities*. Dover, New York (1964).

Marek, P. “Optimal probabilistic measurement of phase”. *Physical Review A*, 88(4) 045802 (2013).

Massar, S. and Popescu, S. “Optimal Extraction of Information from Finite Quantum Ensembles”. *Physical Review Letters*, 74(8) 1259–1263 (1995).

Napolitano, M., Koschorreck, M., Dubost, B., Behbood, N., Sewell, R.J., and Mitchell, M.W. “Interaction-based quantum metrology showing scaling beyond the Heisenberg limit”. *Nature*, 471(7339) 486–489 (2011).

Nielsen, M.A. and Chuang, I.L. *Quantum Computation and Quantum Information*. Cambridge University Press, Cambridge (2010).

Ostermann, L., Ritsch, H., and Genes, C. “Protected State Enhanced Quantum Metrology with Interacting Two-Level Ensembles”. *Physical Review Letters*, 111(12) 123601 (2013).

Pandey, S., Jiang, Z., Combes, J., and Caves, C.M. “Quantum limits on probabilistic amplifiers”. *Physical Review A*, 88(3) 033852 (2013).

Peres, A. "How to differentiate between non-orthogonal states". *Physics Letters A*, 128(1-2) 19 (1988).

Peres, A. and Scudo, P.F. "Transmission of a Cartesian Frame by a Quantum System". *Physical Review Letters*, 87(16) 167901 (2001).

Plotnitsky, A. and Khrennikov, A. "Reality without Realism: On the Ontological and Epistemological Architecture of Quantum Mechanics". *arxiv preprints ArXiv:1502.06310*, page 33 (2015).

Ralph, T.C., Lund, A.P., and Lvovsky, A. "Nondeterministic Noiseless Linear Amplification of Quantum Systems". In "AIP Conference Proceedings", pages 155–160 (2009).

Sentís, G., Bagan, E., Calsamiglia, J., and Muñoz Tapia, R. "Programmable discrimination with an error margin". *Physical Review A - Atomic, Molecular, and Optical Physics*, 88 1–7 (2013a).

Sentís, G., Gendra, B., Bartlett, S.D., and Doherty, A.C. "Decomposition of any quantum measurement into extremals". *Journal of Physics A: Mathematical and Theoretical*, 46(37) 10 (2013b).

Stirzaker, D. *Elementary Probability*. Cambridge University Press, 2n edition (2003).

Sugimoto, H., Hashimoto, T., Horibe, M., and Hayashi, A. "Discrimination with error margin between two states: Case of general occurrence probabilities". *Physical Review A*, 80(5) 052322 (2009).

Summy, G. and Pegg, D. "Phase optimized quantum states of light". *Optics Communications*, 77 75–79 (1990).

Szańkowski, P., Trippenbach, M., and Chwedeńczuk, J. "Parameter estimation in memory-assisted noisy quantum interferometry". *Physical Review A*, 90(6) 063619 (2014).

Usuga, M.a., Müller, C.R., Wittmann, C., Marek, P., Filip, R., Marquardt, C., Leuchs, G., and Andersen, U.L. "Noise-powered probabilistic concentration of phase information". *Nature Physics*, 6(10) 767–771 (2010).

Vidal, G., Latorre, J.I., Pascual, P., and Tarrach, R. "Optimal minimal measurements of mixed states". *Physical Review A*, 60(1) 20 (1999).

Winter, A. "Coding theorem and strong converse for quantum channels". *IEEE Transactions on Information Theory*, 45(7) 2481–2485 (1999).

Wootters, W.K. and Zurek, W.H. “A single quantum cannot be cloned”. *Nature*, 299 802–803 (1982).

Xiang, G.Y., Ralph, T.C., Lund, A.P., Walk, N., and Pryde, G.J. “Heralded noiseless linear amplification and distillation of entanglement”. *Nature Photonics*, 4 316–319 (2010).

Yang, Y. and Chiribella, G. “Is global asymptotic cloning state estimation?” *LIPICS*, 22 220 (2013).

Zavatta, a., Fiurášek, J., and Bellini, M. “A high-fidelity noiseless amplifier for quantum light states”. *Nature Photonics*, 5(November 2010) 52–60 (2011).