Dealing with ignorance: universal discrimination, learning and quantum correlations

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

Gael Sentís Herrera

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

escrita sota la direcció del

Dr. Ramon Muñoz Tapia

Universitat Autònoma
de Barcelona

Bellaterra, febrer de 2014

Abstract

Discriminating the state of a quantum system among a number of options is one of the most fundamental operations in quantum information theory. A primal feature of quantum theory is that, when two possible quantum states are nonorthogonal, no conceivable measurement of the system can determine its state with certainty. Quantum indeterminism so demands a probabilistic approach to the task of discriminating between quantum states. The usual setting considers that the possible states of the system are *known*. In this thesis, I analyze the role of the prior information available in facing a quantum state discrimination problem, and consider scenarios where the information regarding the possible states is incomplete.

In front of a complete ignorance of the possible states' identity, I discuss a quantum *programmable* discrimination machine for qubit states that accepts this information as input programs using a quantum encoding, rather than just as a classical description. This "classical" ignorance is taken into account in the design, and, as a consequence, the machine is not case-specific but it is able to handle discrimination tasks between any pair of possible qubits, once conveniently programmed through quantum inputs. The optimal performance of programmable machines is studied in detail for general qubit states when several copies of the states are provided, in the main schemes of unambiguous and minimum-error discrimination as well as in the more general scheme of discrimination with an error margin.

Then, this type of automation in discrimination tasks is taken further. By realizing a programmable machine as a device that is trained through quantum information to perform a specific task, I propose a quantum *learning* machine for classifying qubit states that does not require a quantum memory to store the qubit programs. I prove that such learning machine classifies the state of a qubit with the minimum-error rate that quantum mechanics permits, thus allowing for several optimal uses of the machine without the need of retraining. A similar learning scheme is also discussed for coherent states of light. I present it in the context of the readout of a classical memory by means of classically correlated coherent signals, when these are produced by an imperfect source and hence their state has some uncertainty associated. I show that the retrieval of information stored in the memory can be carried out more accurately when fully general quantum measurements are used.

Finally, I analyse the mathematical structure of generalized quantum measurements, ubiquitous in all the topics covered in this thesis. I propose a constructive and efficient algorithm to decompose any given quantum measurement into a statistically equivalent convex combination of simpler (extremal) measurements, which are in principle less costly to implement in a laboratory. Being able to compute this type of measurement decompositions becomes useful in practical situations, where often a minimum-resources perspective prevails.

Resumen

Una de las tareas fundamentales de la Teoría de la Información Cuántica consiste en identificar el estado en que ha sido preparado un sistema cuántico. Cuando las posibles preparaciones forman una lista finita de estados, el problema recibe el nombre de discriminación de estados. El caso fundamental de únicamente dos estados posibles se conoce también bajo el nombre de contraste de hipótesis. Una de las características distintivas de la teoría cuántica es el hecho de que para dos estados no ortogonales no hay medida concebible que pueda identificar el estado del sistema con certeza. El indeterminismo cuántico exige, por tanto, un enfoque probabilístico para llevar a cabo esta tarea. Habitualmente se considera que los estados posibles del sistema son conocidos para el experimentador. En esta tesis analizo el papel que desempeña la información previa disponible en la tarea de discriminación y, en particular, analizo situaciones en las que dicha información es incompleta.

Suponiendo una total ignorancia de la identidad de los estados posibles, estudio la probabilidad de error de una máquina programable de discriminación para estados de qubit. Esta máquina incorpora la información sobre los estados en forma de programas de entrada donde se introducen los sistemas cuánticos en las diferentes preparaciones. Es decir, la información es utilizada en su forma genuinamente cuántica, en lugar de como una descripción clásica de los estados. Esta ignorancia clásica se tiene en cuenta en el diseño de la máquina, la cual ya no es específica para cada caso, sino que es capaz de discriminar entre cualquier par de estados de qubit, una vez ha sido convenientemente programada mediante las entradas de estados cuánticos. Estudio en detalle el rendimiento óptimo de estas máquinas para estados de qubit generales cuando se dispone de un número de copias arbitrario, tanto de los programas como del estado que se ha de identificar. Específicamente, obtengo las probabilidades de correcta identificación en los esquemas usuales de error mínimo y discriminación no ambigua, así como en el esquema más general de discriminación con margen de error.

A continuación, este tipo de automatización en tareas de discriminación se lleva un paso más allá. Entendiendo una máquina programable como un dispositivo entrenado con información cuántica que es capaz de realizar una tarea específica, propongo una máquina de aprendizaje cuántico para clasificar estados de qubit que no requiere una memoria cuántica para almacenar los qubits de los programas, permitiendo así repetidos usos de la máquina sin necesidad de volver a entrenarla. Demuestro que dicha máquina de aprendizaje es capaz de clasificar el estado de un qubit con la mínima tasa de errores admitida por la mecánica cuántica, y por tanto puede ser reusada manteniendo un rendimiento óptimo. También estudio un esquema de aprendizaje similar para estados de luz coherente. éste se presenta en un contexto de lectura de una memoria clásica mediante señales coherentes correlacionadas clásicamente cuando éstas son producidas por una fuente imperfecta y, por lo tanto, en un estado con un cierto grado de incertidumbre asociado. Muestro que la extracción de la información almacenada en la memoria es más eficiente si la incertidumbre se trata de una forma completamente cuántica.

Por último, analizo la estructura matemática de las medidas cuánticas generalizadas, omnipresentes en todos los temas tratados en esta tesis. Propongo un algoritmo constructivo y eficiente para descomponer cualquier medida cuántica en una combinación convexa estadísticamente equivalente de medidas más simples (extremales). éstas en principio son menos costosas de implementar en un laboratorio y, por tanto, pueden ser útiles en situaciones prácticas donde a menudo prevalece una perspectiva de recursos mínimos.

Contents

List of Figures

CHAPTER 1

Prolegomenon

“We balance probabilities and choose the most likely. It is the scientific use of the imagination.”

—*Sir Arthur Conan Doyle*
The Hound of the Baskervilles

1.1 Introduction

During World War II, allied forces devoted much effort to determine the extent of German military production, specially of the brand-new Panzer IV and V tanks in the times preceding D-Day. They really wanted to have an idea of how many tanks they would encounter in battlefield, for the success of an invasion crucially depended on it. The intelligence services had gathered some information, namely espionage data of German factories' output, aerial photographies and tank counts at previous contests. Reports indicated contradictory and huge production capabilities, between 1000 and 1500 tanks *per month*. Not happy with these estimates, the allies asked statisticians to see whether their accuracy could be improved.

Only two sources of information were enough to produce incredibly accurate estimates: the number of tanks captured in battlefield, and their serial numbers. With these, statisticians estimated that an average of 246 tanks were being produced per month between 1940 and 1942, while intelligence services reported a number of 1400. When, after the war, the actual German records were captured, they showed a production number of 245 tanks per month for those three years[1]. How could the statisticians be so close?

Say the total number of tanks produced in a particular month is N. Among k captured tanks, the highest serial number turns out to be m. The statisticians assumed that the Germans, being Germans, had numbered their tanks sequentially (and they did), hence they applied the following reasoning. The first consequence of this assumption is that, at least, m tanks were produced. If only one tank is observed, a fairly reasonable guess of N would be to double its serial number, as it is more likely that m falls in the middle of the sequence of the N tanks rather than in the extremes. But this is a long shot, and more precision comes with more serial numbers. The probability that the highest serial number is m in a series of k out of N tanks is given by the number of ways that $k-1$ tanks could have all serial numbers up to $m-1$, divided by all the possible series of k tanks. Mathematically, this is expressed as

$$p(m|N,k) = \frac{\binom{m-1}{k-1}}{\binom{N}{k}} \,. \tag{1.1}$$

According to this probability, the mean value of m is $\bar{m} = (N+1)k/(k+1)$. Then, *assuming* that the observed m coincides with $\bar{m}$, one can propose the estimator $\hat{N} = m + m/k - 1$. Intuitively, this is just the highest serial number plus the average gap between serial numbers. Without going any further, this is the technique that the statisticians used to come up with the number 246. It is, though, a particular way of handling available information and uncertainty, and certainly not the only possible approach.

There is an alternative solution to this problem that, involving different assumptions, accounts for how *our* knowledge is modified when more data becomes available. This solution aims at obtaining the whole probability distribution of the number of tanks $p(N|m,k)$ [that is the inverse of Eq. (1.1)], thus it goes beyond just giving an estimate.

Before any tank is found, we know nothing about N. We can represent this complete ignorance as a uniform probability for any value of N (maybe up to a reasonable maximum, but this is not important). Now, say one tank is found with the serial number 230. Then, two facts and one assumption comprise *our* state of knowledge: N is at least 230 (fact), the *a priori* probability of that number appearing was $1/N$ (fact[2]), and, as said before, any number N of tanks was equally probable (assumption). The composition

[1] These numbers were obtained from [Ruggles and Brodie, 1947].

[2] As long as we keep the problem in its simplest form, e.g., not taking into account that older tanks have a greater probability to be found.

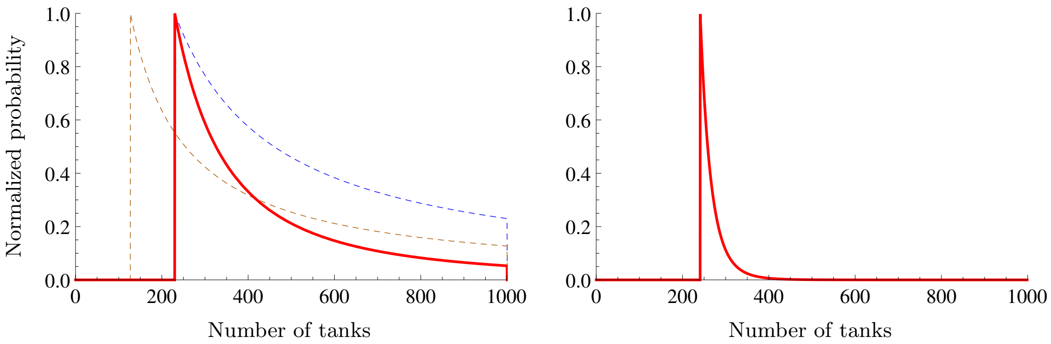


Figure 1.1. (left) Normalized probability of the total number N of tanks when the first tank found is numbered 230 (blue dashed curve) or 127 (brown dashed curve), and when the two tanks are taken into account (red solid curve). To ease presentation, a maximum of 1000 tanks is assumed.
(right) Normalized probability when a series of 10 tanks is observed, with a highest serial number of 241.

of these three pieces of information yields a probability distribution for N, represented by the blue dashed curve in Fig. 1.1. The most likely number of tanks is $N = 230$, but numbers around 900 still have a lot of probability, so we better wait for more data. Say another tank is found, this time with serial number 127. A similar probability distribution represents this new information (brown dashed curve). It could seem that this does not tells us anything new, since we already know that there are at least 230 tanks, but the combination of the old and the new evidence, which, roughly speaking, amounts to multiply the two distributions, is much more eloquent. The red solid curve on the left side of Fig. 1.1 represents our updated state of knowledge after taking into account the second tank. It is still peaked at 230 tanks, but now the greater numbers are significantly suppressed.

Observing more tanks means a greater concentration of the probability near the peak value: the right side of Fig. 1.1 shows the probability distribution for N given a series of 10 tanks, where the highest serial number is 241; from this relatively small amount of data we have been able to localize N around a mean value of 270, with a standard deviation of $\simeq 30$ tanks. For arbitrary k and m (given $k > 2$), the probability distribution is

$$p(N|m,k) = \frac{k-1}{k}\frac{\binom{m-1}{k-1}}{\binom{N}{k}}\,, \tag{1.2}$$

peaked at $N = m$ and with a mean value $\bar{N} = \frac{(m-1)(k-1)}{k-2}$. Although this reasoning follows a fundamentally different route than the first above, when

k is large enough, both $\bar{N}$ and the estimator $\hat{N}$ computed before converge. What this means is that, despite we started from strong—and different!—assumptions in both approaches ($m = \bar{m}$ in the first, and an equal chance of any total number of tanks in the second), their effect in the final result fades away as more data arrives.

The two methods used to solve the "German tank problem", paradigms of statistics, attempt to provide useful answers in uncertain scenarios. Their fundamentals are rooted in different interpretations of information, but they share a common feature: in front of uncertainty, they build on assumptions. Both use, in a way or another—but, maybe, the second method is more explicit—, what we *think* is reasonable, what we *know* beforehand, what we *expect* to observe. Statistics gives us a lesson: any prediction we may make necessarily passes first through us, subjective observers of an uncertain world, and, "despite" that, we are able to predict with relative success. Well enough said by Pierre-Simon Laplace, "probability theory is nothing but common sense reduced to calculation."

The theory that, perhaps, best advocates the importance of the observer as an active agent in the generation of knowledge is quantum mechanics. The building block of the theory is the quantum state, a mathematical entity that does not differ too much from any of the curves in Fig. 1.1, that is, a representation of what one knows and does not know about a particular quantum system. Quantum mechanics, in contrast to its classical counterpart, is thus an intrinsically probabilistic theory, where uncertainty is considered to be a fundamental property of nature, and, moreover, where the act of observation is an intrusive process that necessarily disturbs what is being observed. In a quantum context, the concepts "information" and "uncertainty" adopt new meanings, and the role of the observer is inseparable from any experiment. Statistics arises as the main tool we have to make predictions about the—quantum—world. The example of the German tanks showed the importance of considering our state of knowledge in an uncertain situation—our certainties and our ignorance—as a crucial part of statistical analysis. In a nutshell, this thesis takes the lesson into the analysis of quantum information processes.

In the remainder of this Chapter, I summarize the main results of my research. Chapter 2 starts by giving the reader a philosophical hint on the jumble of interpretations of probability to choose thereafter one of them, a Bayesian view. Then, I introduce some fundamental concepts in quantum theory widely used throughout the whole document, such as quantum states and quantum measurements. In Chapter 3, I describe the main framework in which my research is situated, that is the problem of discriminating between quantum states. I focus on binary discrimination problems. I give an overview of the basics of the topic, starting from its classical analogue, i.e., the problem of distinguishing two probability distributions. Although the Chapter reviews known results, in Section 3.3.3 I present an alternative derivation of the discrimination with an error margin that can be more directly generalized to encompass the setting discussed in Section 4.4. Chapters from 4 to 7 comprise the body of results that I have obtained during my PhD. The dissertation finalizes with an outlook on future work, followed by the bibliography.

1.2 Summary of results

Programmable quantum state discrimination

The central topic of this thesis is quantum state discrimination, a fundamental primitive in quantum statistics where one has to correctly identify the state of a system that is in one of two possible states. The usual approach to the problem considers that the possible states are *known*. By contrast, a programmable discrimination machine performs this task when the pair of possible states is completely unknown. The machine is visualized as a device with one data and two program ports, each fed with a number of identically prepared qubits—the data and the programs—, and it aims at correctly identifying the data state with one of the two program states. The machine is thus designed to work for *every* possible pair of states. In the first part of Chapter 4, I derive the optimal performance of programmable discrimination machines for general qubit states when an arbitrary number of copies of program and data states are available. Two scenarios are considered: one in which the purity of the possible states is *a priori* known, and the fully universal one where the machine operates over generic mixed states of unknown purity. Analytical results are found for both the unambiguous and minimum-error discrimination strategies. This allows to calculate the asymptotic performance of programmable discrimination machines when a large number of copies are provided and to recover the standard state discrimination and state comparison values as different limiting cases. These results are reported in

> G. Sentís, E. Bagan, J. Calsamiglia, and R. Muñoz Tapia, "Multicopy programmable discrimination of general qubit states", *Physical Review A* **82**, 042312 (2010); **83**, 039909(E) (2011).

In the second part of the Chapter, I generalize the problem by allowing an error margin. This generalized scheme has the unambiguous and the minimum-error schemes as extremal cases, when the error margin is set to zero or it is sufficiently large, respectively. Analytical results are given in the two situations where the margin is imposed on the average error probability—weak condition—or it is imposed separately on the two probabilities of assigning the state of the data to the wrong program—strong condition. It is a general feature of the proposed scheme that the success probability rises sharply as soon as a small error margin is allowed, thus providing a significant gain over the unambiguous scheme while still having high confidence results. The contents of this second part are published in

> G. Sentís, E. Bagan, J. Calsamiglia, and R. Muñoz Tapia, "Programmable discrimination with an error margin", *Physical Review A* **88**, 052304 (2013).

Quantum learning of qubit states

In Chapter 5, by taking a closer look to the structure of the optimal measurement in programmable discrimination, I introduce a quantum learning machine for binary classification of qubit states that does not require a quantum memory. I show that this machine performs with the minimum-error rate allowed by quantum mechanics, that is, the one provided by a programmable machine, for *any* size of the training set. This result is robust under (an arbitrary amount of) noise and under (statistical) variations in the composition of the training set, provided it is large enough. Such learning machine can be used an arbitrary number of times without retraining. Its required classical memory grows only logarithmically with the number of training qubits, while its excess risk decreases as the inverse of this number, and twice as fast as the excess risk of an "estimate-and-discriminate" machine, which estimates the (unknown) states of the training qubits and classifies the data qubit with a discrimination protocol tailored to the obtained estimates. These results are reported in

> G. Sentís, J. Calsamiglia, R. Muñoz Tapia, and E. Bagan, "Quantum learning without quantum memory", *Scientific Reports* **2**, 708 (2012).

Quantum learning of coherent states

Chapter 6 extends the learning concepts presented in Chapter 5 to the domain of continuous-variables systems in a particular setting. Using a simple model of a classical memory, consisting in an array of cells with two possible reflectivities, I propose a readout scheme that uses an imperfect coherent light source to illuminate each cell and retrieves the stored binary information by determining the state of the reflected signal. Assuming that a number of extra modes coming from the same source are at one's disposal, I show that a fully quantum processing of the signal together with the extra modes provides better results than any strategy that first tries to diminish the incomplete knowledge of the source specifications by estimating the amplitude of the extra modes, and then determines the state of the signal based on the obtained estimate. In particular, I prove this for any Gaussian estimation measurement, and I conjecture that this is the case for any local strategy based on a simple example. A quantum-enhanced readout of a classical memory is thus observed when using classically correlated coherent signals and the value of their amplitude is not completely determined. The results of this Chapter will be reported in

> G. Sentís, G. Adesso, and M. Guţă, "Quantum reading with coherent light", in preparation.

Decomposition of quantum measurements

The thesis closes with a study of a transversal character: the convex structure of quantum measurements. Present in all previous chapters as solutions of particular optimization problems, generalized quantum measurements, or, more accurately, their mathematical representations, form a convex set. This means that, if a certain measurement belongs to the inner region of the convex set, it is actually implementable as a convex combination of other measurements. The statistics reproduced by the original measurement is identical to the one reproduced by any of its decompositions. In Chapter 7, I design an efficient and constructive algorithm to decompose any generalized quantum measurement into a convex combination of extremal measurements (i.e., measurements that cannot be decomposed as combinations of other measurements). I show that, if one allows for a classical post-processing step, only extremal rank-1 positive operator-valued measures are needed. For a measurement with N elements on a d-dimensional space, the algorithm will decompose it into at most $(N-1)d+1$ extremals, whereas the best previously known upper bound scaled as d^2. Since the decomposition is not unique, I show how to tailor the algorithm to provide particular types of decompositions that exhibit some desired property. This work is published in

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

Fundamentals

> "What exactly qualifies some physical systems to play the role of 'measurer'? Was the wavefunction of the world waiting to jump for thousands of millions of years until a single-celled living creature appeared? Or did it have to wait a little longer, for some better qualified system . . . with a PhD?"
>
> —*John Stewart Bell*
> Against 'Measurement'

This Chapter aims at providing working definitions of key concepts in quantum mechanics that will be used extensively throughout this dissertation, such as probability distributions, quantum states and quantum measurements. A deep understanding of such concepts is an arduous quest with a variety of ends, for it belongs ultimately to the realms of interpretation and philosophy, and it is certainly not the purpose of this introduction to cover these matters in full. However, it is both fascinating and beneficial to shed some light on the conceptual background where the statistical problems posed in the following chapters lie. This Chapter starts sketching the viewpoint considered here, that is the Bayesian interpretation of probability, what comes with it, and which are its alternatives, to detail thereafter the mathematical definitions and formalism later used.

2.1 Epistemology of probability

We constantly handle probabilities in our everyday lives. We make estimations when we lack certainty, we make decisions based on statements that include expressions like “better odds”, “more probable”, or “less likely”. We invoke common sense and probability to give a rational justification to our actions, yet the definition of *probability*, or, more accurately, its interpretation[1], is far from consensus. A probability theory aspires to provide the procedure one should follow in facing any nondeterministic problem if one wants to be *rational*, but that rationality comes in accordance with the interpretation of probability that the theory assumes. Choosing one particular theory carries unavoidably an epistemological compromise, namely a specific answer to the question: what *is* a probability, and what does it tell us about reality?

In modern statistics we can distinguish two major schools of thought that address such a question: frequentism and Bayesianism. However, the first attempt of a formal answer dates from 1812 and is attributed to Laplace’s principle of indifference[2]. In his *Théorie analytique des probabilités*, Laplace wrote

> The theory of chance consists in reducing all the events of the same kind to a certain number of cases equally possible, that is to say, to such as we may be equally undecided about in regard to their existence, and in determining the number of cases favorable to the event whose probability is sought. The ratio of this number to that of all the cases possible is the measure of this probability, which is thus simply a fraction whose numerator is the number of favorable cases and whose denominator is the number of all the cases possible.

The principle simply prescribes the use of the “uniform prior probability distribution” of all possible cases when no evidence indicates otherwise. That is to say, if I roll a die that I’m convinced is unbiased, I should assign a probability 1/6 to each face appearing (needless to say, the principle fails at assessing any problem with no natural symmetry). This is recognized nowadays as Bayesian thinking. In Laplace’s treatise one finds no justification, for him was just common sense, but it actually implies a definite interpretative viewpoint: it locates the essence of probability in the perception of the observer, linking it with a personal belief. In a more recent language, the principle of indifference corresponds to the simplest noninformative prior,

[1] For an account of the mainstream interpretations of probability, see [Gillies, 2000].

[2] This denomination was actually coined much later by John M. Keynes [Keynes, 1921].

that is the—in principle—least compromising assumption one can make over uncertain future phenomena. But an assumption nonetheless.

Frequentism appeared in the scene as a strong critique to intuitive arguments of the sort. The felt necessity to deprive probability of any trace of subjectivism rendered what William Feller calls "the statistical, or empirical, attitude towards probability", initiated mainly by the contributions of Ronald A. Fisher and Richard E. von Mises [Feller, 1950]. The frequentist standpoint conceives the probability of an event as the *relative frequency* of this event happening in an infinite number of trials. Aseptic and strictly empirical. The frequentist methods present several difficulties[3] that need not be reviewed here, but one main shortage worth remarking arises from the very definition of probability just exposed: probabilities are discussed only in relation to well-defined repeatable random experiments, hence situations that are nonrepeatable are out of the question. A typical example used to highlight this fact is the impossibility for a frequentist statistician to say anything about the probability of the Sun exploding tomorrow. One might argue that statistical inference over an "imaginary" ensemble of realizations of such an experiment would still be possible, but then isn't that quite the same as a subjective opinion, a human choice?

The other major approach to probability theory is Bayesianism [Bernardo and Smith, 1994], and it is the point of view taken in the works that this dissertation gathers. The idea, roughly speaking, is that probabilities represent *degrees of belief*, and thus are intrinsically connected to an *agent*, that is the individual who makes probability assignments to events. A probability is, then, a state of knowledge: it summarizes what the agent does and does not know about a given situation, i.e., it is an evaluation of his uncertainty. Its numerical value represents a measure of the willingness of the agent to make a bet in favor of the event in question. In a more formal fashion, the probability of a certain hypothesis H, given some background information S, is defined as the plausibility $P(H|S)$ that the agent gives to H. It verifies the properties

$$0 \leqslant P(H|S) \leqslant 1\,, \tag{2.1}$$

$$P(H|S) + P(\neg H|S) = 1\,, \tag{2.2}$$

where $\neg H$ means the negation of H. The plausibility $P(H|S)$ receives the more common name of *prior*. In the acquisition of new evidence E, the prior is updated according to Bayes' rule

$$P(H|E,S) = \frac{P(H|S)P(E|H,S)}{P(E|S)}\,. \tag{2.3}$$

[3] See e.g. [Howson and Urbach, 2006] for a critique of frequentism in statistics.

Now, on a more ontologic note, there are also theories that confer *being*—additionally to *meaning*—to these notions of probability, both in the frequentist and the Bayesian perspectives. The common goal is to answer the second part of the question posed at the beginning of this Section: what does a probability tell us about reality? From the frequentist side, an attempt to explain the emergence of stable relative frequencies in nature can be found, for instance, in Karl Popper's *propensity theory*[4]. This theory establishes that probabilities (frequencies) are to be understood as objective tendencies of experimental situations to produce some outcomes over others. Knowledge of such "physical properties"[5] of systems is then accessible only through multiple repetitions of the experiment. This way of thinking would make sense of single-case probability attributions, which can be very appealing for solving the pressing need of an objectivistic approach to statistics—specially in intrinsically indeterministic theories like quantum mechanics—, but it is a somewhat ad hoc way of giving frequencies a scent of physical reality that is not even falsifiable, to put it in Popper's own terms, not to mention it carries the difficulties and critiques of the frequentist approach.

The Bayesian approach, as presented before, is strongly grounded in subjectivism. It is an exclusively epistemological approach, with no ontological endeavors. To consider probabilities plainly as degrees of belief of a decision making agent, and operating from this starting point on a logical base, together with Bayes' rule, receives the name of subjective (or personalist) Bayesianism[6]. This posture situates probabilities in the agent's mind, while leaving not a tiny bit of separated, objective essence in whatever the probabilities refer to. As a consequence, assuming one or another prior probability distribution is up to the agent's taste and consideration, in the sense that there is no "right" choice (of course, there may still be "unreasonable" choices. But, again, according to other's judgement. Not all that objective). Opposing this view there is objective Bayesianism [Jaynes, 2003], which supports that there is a unique rational probability that one ought to assign for any uncertain event[7]. It is the hope of this standpoint that a way could be found to elucidate these "right" probabilities, sustained by logical analysis alone. But, as for now, it is generally acknowledged that no one has succeeded in

[4] See [Popper, 1982] or, for a more recent version of the theory, [Gillies, 2000].

[5] Be an example of to which extent probability was regarded as a physical feature in pre-Bayesian theories the case of Richard E. von Mises, who even refers to probability theory as a field of theoretical physics, much as like classical mechanics or optics.

[6] Subjective Bayesianism was born with the works of philosophers [de Finetti, 1931] and [Ramsey, 1931]. For an accessible introduction, see [Jeffrey, 2004].

[7] The discussion about true or right values for Bayesian probabilities originates with David Lewis' principal principle, and his notion of *objective chance* [Lewis, 1980].

such enterprise.

From these lines onwards I will assume the subjective Bayesian viewpoint on probabilities. Therefore, no ontologic forethought will be made but, instead, a purely information-theoretic one.

2.2 The quantum state

With all this said about probabilities, I will simply identify the states of quantum systems with Bayesian probability distributions. That is to say, a quantum state is nothing more than the mathematical object we use to represent our degree of uncertainty about a particular quantum system.

Say we are given a quantum system prepared in a certain state. We know nothing about the preparation procedure, but we are said the state of the system is either $|\psi_1\rangle\langle\psi_1|$ or $|\psi_2\rangle\langle\psi_2|$ with probabilities η_1 and $\eta_2 = 1 - \eta_1$, respectively. For *us*, the state of the system, that is our state of knowledge, is then represented by the weighted superposition of the two possibilities $\rho = \eta_1 |\psi_1\rangle\langle\psi_1| + \eta_2 |\psi_2\rangle\langle\psi_2|$. In general, ρ is called a *density operator* and, as such, it stands for a quantum state. A density operator acts on the d-dimensional Hilbert space of the system, where d is finite, and it fulfils the properties

$$\begin{aligned} \rho &= \rho^\dagger \\ \rho &\geqslant 0 \\ \operatorname{tr}\rho &= 1\,, \end{aligned} \tag{2.4}$$

i.e., its matrix representation ought to be Hermitian, have nonnegative eigenvalues and be normalized. If the density operator is a one-dimensional projector, i.e., it is of the form $\rho = |\psi\rangle\langle\psi|$, the state is said to be *pure.* Otherwise, higher-rank density operators are said to be *mixed* states. The density operator is also commonly known as *density matrix.* I will use both terms interchangeably.

Pure states correspond to states of maximal knowledge, whereas mixed states correspond to less than maximal knowledge [Blum, 1996; Fuchs, 1996]. This assertion is evident in the above example, in which we end up with a mixed state because the lack of knowledge about the preparation procedure forces a probabilistic description of the state of the system. This also arises when one has maximal knowledge of a bipartite system, that is when one describes its state with a pure state $|\psi\rangle\langle\psi|$ on some tensor-product Hilbert space $\mathcal{H}_1 \otimes \mathcal{H}_2$. Quantum mechanics then establishes that one's knowledge of a subsystem shall be less than maximal. Indeed, the state of subsystem

1 is obtained through a partial trace operation over $\mathcal{H}_2$. Let $\{|u_i\rangle|v_j\rangle\}$ be a basis for $\mathcal{H}_1 \otimes \mathcal{H}_2$; then

$$|\psi\rangle = \sum_{i,j} c_{ij}\, |u_i\rangle|v_j\rangle \ , \tag{2.5}$$

and the state of subsystem 1 is

$$\rho = \mathrm{tr}_2\, |\psi\rangle\!\langle\psi| = \sum_k \langle v_k|\psi\rangle\!\langle\psi|v_k\rangle = \sum_{i,j} c_{ij} c_{ij}^*\, |u_i\rangle\!\langle u_i| \ , \tag{2.6}$$

i.e., a mixed state. One obtains a similar result for the state of subsystem 2.

In general, a density matrix admits infinitely many decompositions as a combination of pure states. Two ensembles of pure states $\sum_i \eta_i\, |\psi_i\rangle\!\langle\psi_i|$ and $\sum_i \kappa_i\, |\varphi_i\rangle\!\langle\varphi_i|$ represent the same density matrix if its elements are connected by a unitary transformation U, such that

$$\sqrt{\eta_i}\, |\psi_i\rangle = \sum_j U_{ij} \sqrt{\kappa_j}\, |\varphi_i\rangle \ . \tag{2.7}$$

These representations of mixed states in terms of ensembles of pure states do not immediately give an idea of how much "less than maximal" is the knowledge that they represent. Being able to compare mixed states in regards to their "mixedness" is of fundamental importance for many applications in quantum information. For $d = 2$ there is a simple and useful way of expressing a mixed state that tells us explicitly how much mixed it is. Certainly, any mixed state ρ can be expressed as

$$\rho = r\, |\psi\rangle\!\langle\psi| + \frac{1-r}{2}\mathbb{1} \, , \tag{2.8}$$

i.e., a weighted combination of $\mathbb{1}$, the identity operator on the two-dimensional Hilbert space of the system, and some pure state $|\psi\rangle\!\langle\psi|$. The weight r is referred to as the *purity* of ρ, in the sense that it signifies the degree of mixture between an object of maximal knowledge—the pure state—and the complete absence of it—the identity operator.[8]

Now that pure and mixed states have been defined, a clarification is in order. Maximal knowledge shall not be misinterpreted as deterministic

[8] The parameter r gives an idea of how close is ρ to a pure state. This type decomposition exists for two-dimensional systems because there are only two possible ranks for ρ: it is either rank 1 (pure) or full rank (mixed), hence every mixed state can be expressed as Eq (2.8) dictates. For $d > 2$, mixed states with intermediate ranks are possible and the measure of "mixedness" turns subtler. In general, the answer to the question of whether a certain state ρ_1 is more mixed than another state ρ_2 is provided by the majorization relation between the eigenvalue sequences of ρ_1 and ρ_2.

knowledge. The fact that I know with certainty that the state of a system is $|\psi\rangle\langle\psi|$ does not mean that I would get a deterministic result—some prefixed value—if I measure it. As it will become clear in Section 2.3, the measurement outcomes would still be probabilistic. The "maximal" in maximal knowledge means "to the extent that we are allowed by quantum mechanics". And then, one can rise the following question: even though intrinsically probabilistic, if a pure state is the maximal state of knowledge of a quantum system we can aim for, should not we identify it with a *property* of the system itself? Should not we attribute physical reality to the mathematical object $|\psi\rangle$? This question is as old as the quantum theory. Without entering into much detail, let me just say that, as it happens with probability theories, there is no definite answer and an alluring debate around what someone has referred to as ψ-ontology keeps going on. Extensions of subjective Bayesianism (see Section 2.1) into the quantum realm are, for instance, the Deutsch-Wallace variant of the many-worlds interpretation of quantum mechanics [Deutsch, 1999; Wallace, 2007], and "Quantum Bayesianism" [Caves *et al.*, 2002; Fuchs, 2010], an interpretation of quantum theory that is cautious enough to not relate quantum states to physical properties at all. Perhaps the most extreme version of the information-theoretic approach to this matter was worded by John Wheeler in his "it from bit" thesis [Wheeler, 1990]:

> It from bit symbolizes the idea that every item of the physical world has at bottom—at a very deep bottom, in most instances—an immaterial source and explanation; that what we call reality arises in the last analysis from the posing of yes-no questions and the registering of equipment-evoked responses; in short, that all things physical are information-theoretic in origin and this is a participatory universe.

Of course, one can also find arguments in favor of the opposed school, that is the idea of pure states being, indeed, physical properties of systems [Pusey *et al.*, 2012]. The discussion is all but settled.

The Bloch sphere

Quantum states of two-dimensional systems, a.k.a. qubits, find a particularly useful geometrical representation in the so called *Bloch sphere* picture. This representation will be used extensively in the remaining chapters of the dissertation.

Any pure state $|\psi\rangle$ of a two-dimensional system can be written in the computational basis, that is the basis formed by the orthogonal vectors $|0\rangle$ and $|1\rangle$, as $|\psi\rangle = \alpha\,|0\rangle + \beta\,|1\rangle$, where α and β are complex numbers. Since only

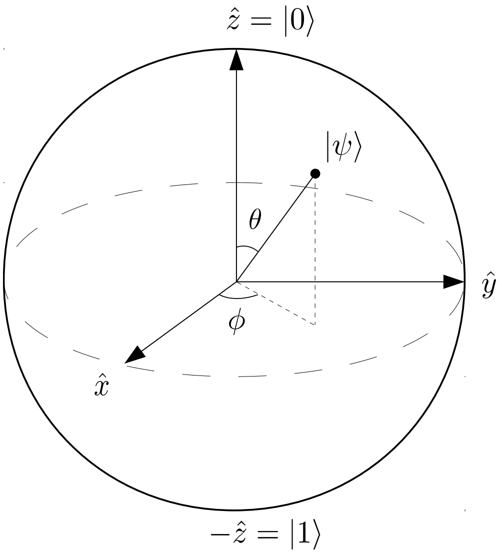


Figure 2.1. The Bloch sphere.

the relative phase between α and β has any physical meaning, α can be taken to be real. The normalization condition $\langle\psi|\psi\rangle = 1$ leaves two free parameters to specify the state. In particular, one can choose the parametrization to be

$$|\psi\rangle = \cos\frac{\theta}{2}\,|0\rangle + e^{i\phi}\sin\frac{\theta}{2}\,|1\rangle \ , \tag{2.9}$$

where $0 \leqslant \theta < \pi$ and $0 \leqslant \phi < 2\pi$. The pair of angles $\{\theta, \phi\}$ fully determines the state $|\psi\rangle$, and, interpreted as spherical coordinates, specifies a point $\boldsymbol{v} = (\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta)$ in the surface of a unit 2-sphere (see Fig. 2.1). Thus, this surface represents the set of all pure states for a qubit.

In a general way, any qubit density matrix ρ can be written in the compact form

$$\rho = \frac{\mathbb{1} + r\,\boldsymbol{v}\cdot\boldsymbol{\sigma}}{2}\,, \tag{2.10}$$

where $\boldsymbol{v}$ is the so-called *Bloch vector* of the state ($|\boldsymbol{v}| = 1$), r is its purity, and $\boldsymbol{\sigma} = (\sigma_x, \sigma_y, \sigma_z)$ is the vector of the Hermitian, traceless Pauli matrices. As it could already be seen from Eq. (2.8), taking the value $r = 0$ in Eq. (2.10) yields the completely mixed state $\mathbb{1}/2$, whereas $r = 1$ leaves us with a rank-1 density matrix, i.e., a pure state. In the Bloch sphere picture, mixed states correspond to interior points of the sphere at a distance $r < 1$ from the origin of coordinates.

2.3 The quantum measurement

The measurement process in quantum mechanics has been a controversial subject of study since the very origins of the theory. Two facts justify the difficulty: quantum indeterminism reveals itself upon measuring quantum systems, and, moreover, the state of the system appears to change abruptly right after the process, an experimental observation that is captured by the wave function collapse postulate of quantum mechanics. The way measurement theory is presented in standard quantum mechanics textbooks is as follows:

- Physical quantities that can be measured are formally represented by self-adjoint operators acting on the state Hilbert space called *observables*. Upon measuring some observable A, only its eigenvalues can be observed as measurement outcomes. Say A has the spectral decomposition $A = \sum_i \lambda_i |a_i\rangle\langle a_i|$. Then, the average value of A when measured in some state ρ is

$$\langle A\rangle = \sum_i \lambda_i p(\lambda_i|\rho) = \sum_i \lambda_i \mathrm{tr}\left(|a_i\rangle\langle a_i|\,\rho\right), \tag{2.11}$$

  where $p(\lambda_i|\rho)$ is the probability of obtaining the eigenvalue λ_i as the outcome of the measurement over ρ. Eq. (2.11) is just the weighted sum of the outcomes with their probabilities of occurrence, which result from projection operations on the state ρ. Hence the measurement of A can be completely characterized as a *projective measurement*, specified by a set of orthogonal projectors $\{\Pi_i \equiv |a_i\rangle\langle a_i|\}$, and its associated set of outcomes $\{\lambda_i\}$.

- After outcome λ_i has been obtained, the state of the system—instantly!—becomes

$$\rho_i = \frac{\Pi_i\rho\Pi_i}{\mathrm{tr}\left(\Pi_i\rho\right)} = |a_i\rangle\langle a_i|\ , \tag{2.12}$$

  where the last equality holds in this case because Π_i is a rank-1 projector. This is the wave function collapse postulate. As it is evident from Eq. (2.12), if the same measurement $\{\Pi_i\}$ is applied to the posterior state ρ_i, the same ith outcome will be obtained. This repeatability is a feature of projective measurements, and its experimental verification is what caused the need to include this rather hard-to-swallow postulate in the earlier formulations of quantum theory.[9]

[9] No physicist is comfortable with abrupt phenomena. Some modern formulations as, for

A projective measurement as the one described also receives the name of *Projection-Valued Measure* (PVM), or *von Neumann measurement*. Generically, this type of measurement includes any complete set of orthogonal, not necessarily rank-1, projectors $\{\Pi_i\}$ over the state space. However, quantum mechanics allows for a more general measurement framework. Dropping the requirements for the elements of the measurement to be orthogonal and to be projectors, one is left with a set of *positive semidefinite* operators $\{E_i\}$, i.e., self-adjoint operators with nonnegative eigenvalues, usually denoted

$$E_i \geqslant 0\,, \tag{2.13}$$

that fulfil a completeness relation of the sort

$$\sum_i E_i = \mathbb{1}\,. \tag{2.14}$$

A set of operators that verify these two conditions is called a *Positive Operator-Valued Measure*, or POVM [Helstrom, 1976]. In such a description of a measurement, the outcomes are not necessarily related to an eigenvalue of some observable but are just a label, one for each element of the set $\{E_i\}$. A picture that may resemble a POVM is that of a machine with a pilot light for each possible outcome. The machine accepts a quantum state ρ as input, measures it, and blinks one of the lights. The probability of obtaining the outcome i, also referred to as the signalling of the element E_i, is given by

$$p(i) = \mathrm{tr}\,(E_i\rho)\,. \tag{2.15}$$

Conditions (2.13) and (2.14) guarantee $p(i) \geqslant 0, \forall i$ and $\sum_i p(i) = 1$, respectively, rendering $p(i)$ a proper probability distribution of the outcomes. In contrast to PVMs, the POVM elements E_i need not commute with each other. Also, POVMs are not repeatable.

The POVM framework is particularly useful in situations in which all that matters is the measurement device itself, i.e., both when the state of the system after the measurement is irrelevant[10], and—as said—when there is no interest in measuring a physical quantity but in the occurrence of certain outcomes. In other words, when the only thing one cares about is the probability distribution of the outcomes. Moreover, there are questions for which PVMs simply do not provide the best answer[11].

instance, Quantum Bayesianism, consider this "spooky" collapse simply as an update of the measurer's knowledge about the state of the system, nothing to do with a physical process.

[10] The post-measurement state will depend on the particular implementation of the POVM, for which there is no unique procedure.

[11] A clear example will be presented in Section 3.3.2: the optimal measurement needed for unambiguous discrimination of two qubits needs three outcomes, despite the Hilbert space of the states is two-dimensional.

But this is a mathematical framework, and the measurements performed in a laboratory are physical after all! Some observable has to be observed, because that is the only thing we can observe. A very relevant result in the field is Neumark's dilation theorem[12] [Peres, 1990], which states that every POVM can be realized as a PVM over an auxiliary system—or *ancilla* [Helstrom, 1976]—correlated with the original system. Specifically, a d-dimensional system can be measured with a POVM with $n > d$ outcomes by performing a repeatable—projective—measurement over an n-dimensional ancilla. This result allows us to set up the optimization problems considered here, in which we optimize some figure of merit over all possible quantum measurements, by focusing solely on sets of operators $\{E_i\}$ fulfilling the POVM conditions (2.13) and (2.14).

[12] Alternatively spelled as Naimark's dilation theorem.

CHAPTER 3

Discrimination of quantum states

> "En todas las ficciones, cada vez que un hombre se enfrenta con diversas alternativas, opta por una y elimina las otras; en la del casi inextricable Ts'ui Pên, opta—simultáneamente—por todas."
>
> —*Jorge Luis Borges*
>
> El jardín de senderos que se bifurcan

Quantum Information is all about the processing of information that is encoded in the state of a quantum system [Nielsen and Chuang, 2000]. But then, after the processing part has taken place, the information has to be read out, or, in other words, the state of the system has to be determined in some sense[1]. There exists a variety of ways to do so, highly dependent on what type of information one is interested in and what one knows already about the state. In particular, when the state is determined by selecting one state among a number of hypotheses, one refers to the task as *quantum state discrimination*. Orthogonal states are relatively straightforward to discriminate. If one counts with the knowledge of the various hypotheses, one can in principle discriminate perfectly among them. This is not so when the possible states are nonorthogonal. In such a case, errors will be unavoidable and the discrimination protocol shall be designed to satisfy some other optimal-

[1] Maybe not necessarily be *completely* determined, depending on the task at hand. In any case, some attribute of it has to be extracted through a measurement.

ity criteria. Designing such protocols has proven to be highly nontrivial and case-specific, the reason for which such a basic decision problem has received great attention by the quantum information community in the last decades[2].

The subsequent chapters of this dissertation (with the exception of Chapter 7) start from quantum state discrimination problems arising in various settings, with the common denominator of the lack of classical information about the hypotheses. It is the purpose of this Chapter to provide general background and definitions for the task of discriminating between *known* quantum states, and set a basis upon which to build the more specific cases treated next.

3.1 The unknown quantum state

Chapter 2 presented quantum states as probability distributions, and probability distributions as states of knowledge of an agent about some physical system. Also, it was said that measurements over the system may provide the agent with new evidence, and his state of knowledge be hence updated via Bayes' rule. Generically, every information processing task can be depicted in an scenario involving *two* agents: the first agent follows a certain processing protocol and prepares some quantum state, which is then sent to the second agent, who has to determine it through a measurement. The first agent may be referred to as *sender*, *preparator*, or even just *source*. The second agent would be the *receiver*, *measurer*, or, very often, *us*. In state determination problems the preparation step has been already carried out, hence the role of the second agent, that is the measurement process, is the central object of analysis.

One may think that the fact of whether there *is* or there *is not* an actual agent sending the state is of no importance as far as the measurer is concerned, for the only thing he should care about is the arrival of the state. However, under the Bayesian framework, the presence of a sender resolves in some way what it may look as a mere linguistic conundrum—but it is actually more than that[3]: what do physicists refer to with the ubiquitous concept of an *unknown quantum state* that the measurer shall unravel? Indeed, if quantum states are, in the end, states of knowledge of an agent, then how can

[2] The fundamentals of quantum state discrimination were pioneered in [Helstrom, 1976]. For a historical review on the topic, see [Chefles, 2000]. For a more recent review, see [Bergou *et al.*, 2004].

[3] Besides its rightful epistemologic relevance in regards to the consistency of the Bayesian view of probabilities, the conundrum has led to mathematical theorems of paramount importance such as the quantum version of the de Finetti theorem.

there be an *unknown* quantum state at all? Its very existence implies that it should be known, if not by the measurer, by someone else! Incorporating a sender to the scene sorts out this apparent contradiction in the sense that we, as measurers, may simply assume that he knows the preparation procedure and, therefore, the state we are commissioned to determine. In short, we are just accessing the state of knowledge of the sender through measurements on the system. This assumption may look somewhat artificial in some settings, for instance in quantum state tomography[4]. For the time being, however, let this simple picture help to sketch the type of state determination tasks that this treatise addresses.

Let me begin with a simple binary decision problem. Imagine that the sender prepares a quantum system in some state and sends it to us, the receivers. The sender does not tell us which of two possible preparation procedures has been carried out, only that it has been selected by tossing a fair coin. Heads corresponds to the first preparation procedure, which yields the quantum state ρ_1, whereas tails corresponds to the second procedure, which outputs some other quantum state ρ_2 (the descriptions ρ_1 and ρ_2 are known). Now our task begins, that is to decide which procedure has taken place. With the piece of information that the sender has provided, our state of knowledge regarding the system has become

$$\rho = \frac{1}{2}\rho_1 + \frac{1}{2}\rho_2 \,. \tag{3.1}$$

To aid in our decision we perform a measurement on the system with two outcomes, 1 and 2. The information gained in the measurement process is then used to make a guess: if the outcome 1 is obtained we will say that the first procedure was selected, hence that the prepared state for the sender was ρ_1, and equivalently for the outcome 2 and the state ρ_2. In general, there exists the possibility of making a wrong guess[5], and we want to engineer the measurement to minimize that chance as much as possible using the information we have available, that is the hypothetical states ρ_1 and ρ_2 together with the fact that the coin is fair.

The described task is a particular instance of *quantum state discrimination.* Now, two remarks are in order:

[4] The objective of this task is to determine an unknown state ρ that some source is believed to be repeatedly preparing, which in turn characterizes it. The concept of a man-in-the-box that owns the state of knowledge ρ, placed inside the source, seems ridiculous. Fortunately, such an elaboration is not necessary at all. The problem and its solution are well posed in [Fuchs and Schack, 2004].

[5] I will extensively comment on this in Section 3.3. As for now, it is enough to consider that an erroneous guess may happen.

The first remark is that no reference to either true or false states has been made whatsoever. Both ρ_1 and ρ_2 are states of knowledge owned by the sender, and ρ is the state of knowledge owned by us before the measurement takes place (hence, at that time, two different descriptions of the same system coexist). After we measure the system we bet for one of the two preparation procedures, i.e., we bet on a past—deterministic—event: the sender's choice. It is in this sense that we may make a mistake[6]. The focus here is completely upon our subjective expectation for this mistake happening.

The second remark is that the prior information we count on greatly influences the task itself. As it is obvious, the form of the state ρ in Eq. (3.1) is a direct product of both the fairness of the coin and the two given hypotheses. If any of this information were different, ρ would change and so would our measurement strategy. But there is more:

The nature of the prior information even determines the questions we might expect to answer by measuring the system. As an example, for a number of hypotheses greater than two, we may end up with a problem with no explicit optimal solution. Such settings fall under the category of *multihypothesis quantum state discrimination* problems, in which only special cases are solvable. Taking this to the limit, in the case of complete absence of prior information we are forced to assume that the received system can be in *any* state of its Hilbert space. Under these circumstances the set of possible states is infinite, and there is no realistic measurement with infinite outcomes to associate with each possibility, hence discrimination becomes nonsensical. We might then expect to answer a different question, that is which state most closely resembles the actual state. This task receives the name of *quantum state estimation* and takes a completely different approach. Lastly, imagine a variation of the setting in which the sender prepares two states, and tells us that they are either equal or different to each other. In such case the task is referred to as *quantum state comparison*.

Starting from the scheme of two agents just exposed, I will cover in the next sections the specifics of quantum state discrimination, for which I begin with its classical analogue: discrimination of probability distributions.

3.2 Discrimination of probability distributions

One of the most fundamental problems in statistical decision theory is that of choosing between two possible explanations or models. It is called *hypothesis testing* [Hoel *et al.*, 1971]. Say a medical test is designed to determine if a pa-

[6] And if there were no sender, no one would be able to tell us that we are wrong!

tient is healthy (hypothesis H_1) or it has contracted some disease (hypothesis H_2). The decision is made in view of the data obtained by the test, which produces a binary result ($i = 1, 2$). There are two types of errors involved: the rejection of a true H_1 and the acceptance of a false H_1, happening with probabilities $p(2|H_1) \equiv p_1(2)$ and $p(1|H_2) \equiv p_2(1)$, respectively. In general these two types of errors do not have to be treated on equal footing, since diagnosing the disease to a healthy patient may not have the same consequences as failing to detect a true disease. It would be desirable to design a test that minimizes both errors, but this is typically not possible since a reduction of one of them is tied to an increase of the other. The Bayesian-like approach to the problem consists in minimizing the average of the errors

$$\eta_1 \, p(2|H_1) + \eta_2 \, p(1|H_2) \,, \tag{3.2}$$

with respect to some prior state of knowledge (encapsulated in the distribution $\{\eta_1, \eta_2\}$ for the *a priori* probabilities of occurrence of each hypothesis). In this context, such approach is known as symmetric hypothesis testing.

Taking this medical example to more abstract grounds, the problem becomes that of discriminating two possible probability distributions $p_1(i)$ and $p_2(i)$, $i = 1, \ldots, n$, by means of *one* sampling. We, the discriminators, must infer the identity of the probability distribution with the smallest probability of error in average, based solely on the drawn sample and the a priori probabilities η_1 and η_2. A reasonable candidate for the best strategy to accomplish this task is to just bet for the distribution that provides the outcome of the sampling with the largest *posterior* probability, i.e., to use the Bayes decision function[7]. Given the outcome i, the posterior probability for the probability distribution $p_1(i)$ to be true is given by Bayes' rule

$$p(1|i) = \frac{\eta_1 p_1(i)}{p(i)} = \frac{\eta_1 p_1(i)}{\eta_1 p_1(i) + \eta_2 p_2(i)} \,, \tag{3.3}$$

and equivalently for $p(2|i)$, where $p(i)$ is the total probability for the outcome i to come up in the sampling. The Bayes decision function simply becomes

$$\delta(i) = \begin{cases} 1 & \text{if} \quad \eta_1 p_1(i) > \eta_2 p_2(i) \\ 2 & \text{if} \quad \eta_1 p_1(i) < \eta_2 p_2(i) \,, \\ \text{anything} & \text{if} \quad \eta_1 p_1(i) = \eta_2 p_2(i) \end{cases} \tag{3.4}$$

where the value of $\delta(i)$ indicates the bet in an obvious way. With this strategy, the probability of a wrong guess i is given by the minimum of the

[7] It is not only reasonable but also optimal, in the sense that any other decision function provides a greater probability of error in average. A simple proof can be found, for instance, in [Fuchs, 1996].

conditional probabilities, i.e. $\min\{p(1|i), p(2|i)\}$. This allows to concisely write the average probability of error according to Bayes decision function—hereafter simply called the minimum probability of error—as

$$\begin{aligned} P_e &= \sum_{i=1}^{n} p(i)\min\{p(1|i), p(2|i)\} \\ &= \sum_{i=1}^{n} \min\{\eta_1 p_1(i), \eta_2 p_2(i)\}\,. \end{aligned} \tag{3.5}$$

Note that P_e explicitly depends not only on the distributions to be discriminated, but also on our subjective prior state of knowledge $\{\eta_1, \eta_2\}$. As it was pointed out in Section 2.1, prior-dependence is neither a shortage nor a strength, but a hard-coded characteristic of Bayesian statistics. One only needs to take this dependence into account when drawing conclusions from Bayesian analysis.

The value of P_e is intuitively related to how distinguishable $p_1(i)$ is from $p_2(i)$. Obviously, the more distinguishable, the less errors we make in identifying them. Unfortunately, although P_e has a clear operational interpretation and it is easily computable, it fails at quantifying the distinguishability of probability distributions. The reason for this is that it is not monotonous under the increase of the number of samplings. Indeed, Eq. (3.5) was derived for one sampling, but nothing prevented us in principle from sampling the distribution more times before making our guess. And if so, it may happen that a pair of probability distributions provides a smaller P_e than another pair when sampling once, while being the other way around if we allow the decision to be based on two samples[8].

It is desirable to overcome this limitation, i.e., to find a function that does not depend explicitly on the number of samplings. A reason to do so is that such function will yield a proper distinguishability measure for probability distributions in the context of decision problems. In addition, such figure will build a notion of *distance* between probability distributions. The answer gets revealed in taking a closer look to the multiple sampling case.

3.2.1 The Chernoff bound

Let us now sample the distribution N times before making a guess. The set of possible outcomes (the sample space) is the N-fold Cartesian product of $\{1, 2, \ldots, n\}$. Denote a particular set of N outcomes as

$$i^{(N)} = (i_1, i_2, \ldots, i_N) \in \{1, 2, \ldots, n\}^{\times N}\,. \tag{3.6}$$

[8] Examples that illustrate such situation can be found in [Cover and Thomas, 2006].

The two probability distributions for a given sequence $i^{(N)}$ are

$$p_1\left(i^{(N)}\right) = p_1(i_1)p_1(i_2)\cdots p_1(i_N)\,, \tag{3.7}$$

and

$$p_2\left(i^{(N)}\right) = p_2(i_1)p_2(i_2)\cdots p_2(i_N)\,. \tag{3.8}$$

Now, using the inequality

$$\min\{a,b\} \leqslant a^s b^{1-s}\,, \quad s \in [0,1]\,, \tag{3.9}$$

that holds for any two positive numbers a and b, the probability of error can be written as

$$\begin{aligned} P_e(N) &= \sum_{i^{(N)}} \min\left\{\eta_1 p_1\left(i^{(N)}\right), \eta_2 p_2\left(i^{(N)}\right)\right\} \\ &\leqslant \eta_1^s \eta_2^{1-s} \sum_{i^{(N)}} \left(\prod_{k=1}^{N} p_1(i_k)^s p_2(i_k)^{1-s}\right) \\ &= \eta_1^s \eta_2^{1-s} \prod_{k=1}^{N} \left(\sum_{i_k=1}^{n} p_1(i_k)^s p_2(i_k)^{1-s}\right) \\ &= \eta_1^s \eta_2^{1-s} \left(\sum_{i=1}^{n} p_1(i)^s p_2(i)^{1-s}\right)^N . \end{aligned} \tag{3.10}$$

The bound becomes even tighter when taking the minimum over s, that is

$$P_e(N) \leqslant \min_{s\in[0,1]} \eta_1^s \eta_2^{1-s} \left(\sum_{i=1}^{n} p_1(i)^s p_2(i)^{1-s}\right)^N \tag{3.11}$$

This is the *Chernoff bound* [Chernoff, 1952].

This is a specially remarkable upper bound for the optimal $P_e(N)$ because it is actually attained in the asymptotic limit $N \to \infty$[9]. At an intuitive level, it is clear that the probability of error goes to zero as N increases. It turns out that the shape of this decrease asymptotically approaches an exponential function, and, moreover, the exact rate exponent is fixed through Eq. (3.11), i.e.,

$$P_e(N \to \infty) \sim e^{-NC(p_1,p_2)}\,, \tag{3.12}$$

with

$$C(p_1,p_2) \equiv -\log \min_{s\in[0,1]} \sum_{i=1}^{n} p_1(i)^s p_2(i)^{1-s}\,. \tag{3.13}$$

[9] The proof for the attainability of the Chernoff bound is more involved and shall not be reproduced here. It can be found in [Cover and Thomas, 2006].

The exponent $C(p_1, p_2)$ is known as the *Chernoff distance*. For the special case of measurements with two outcomes, that is $n = 2$, the meaning of the Chernoff distance can be easily pinned down. This is the case of a biased coin tossed N times, with two possible probability distributions for the outcomes, $p_1 = \{p, 1-p\}$ and $p_2 = \{q, 1-q\}$. A result of N_0 "heads" out of N tosses, according to p_1, occurs with probability

$$P_1(N_0) = \binom{N}{N_0} p^{N_0}(1-p)^{N-N_0}\,, \tag{3.14}$$

whereas, according to p_2, occurs with probability $P_2(N_0)$, defined as $P_1(N_0)$ but with p replaced by q. In the limit of large N these distributions approach Gaussians centred at pN and qN, respectively. Let ξ be the fraction of "heads" above which one must decide in favor of p_1. That is, if $N_0 \geqslant \xi N$ one accepts the distribution p_1, whereas if $N_0 < \xi N$ one accepts p_2. The main contribution to the error probability in the asymptotic regime is due to cases in which $N_0 = \xi N$, i.e., by events that occur with the same probability for both hypotheses (see Fig. 3.1). It can be proven that

$$-\lim_{N\to\infty} \frac{\log P_1(\xi N)}{N} = C(p_1, p_2) \tag{3.15}$$

(the same limit holds for P_2). That is to say, the Chernoff distance, defined as in (3.13) for the case of $n = 2$, is exactly the exponent of the asymptotic probability of such events, and thus of the asymptotic error probability.

The Chernoff distance thus allows to properly compare pairs of probability distributions in regards to their distinguishability, in the sense of the error probability inherent to the task of discriminating among them[10]. Going to the asymptotic limit $N \to \infty$ is the way to get rid of N-dependent results, obtaining a quantity that depends solely on the pair of probability distributions, thus related to some relative property of them. Furthermore, note that even the prior dependence has disappeared in Eq. (3.13). All these nice properties will hold in the quantum version of the Chernoff bound (see Section 3.4.2), plus additional benefits of a purely quantum nature.

[10] The error probability is just one way to define a notion of distinguishability, in this case through a decision problem. There is a variety of figures to assess how much distinguishable are two probability distributions, namely the mutual information, the statistical overlap or fidelity, or the Kullback-Leibler information, although none of them as clearly defined in an operational sense as the probability of error. For a compendium of distinguishability measures, both classical and quantum, see [Fuchs, 1996].

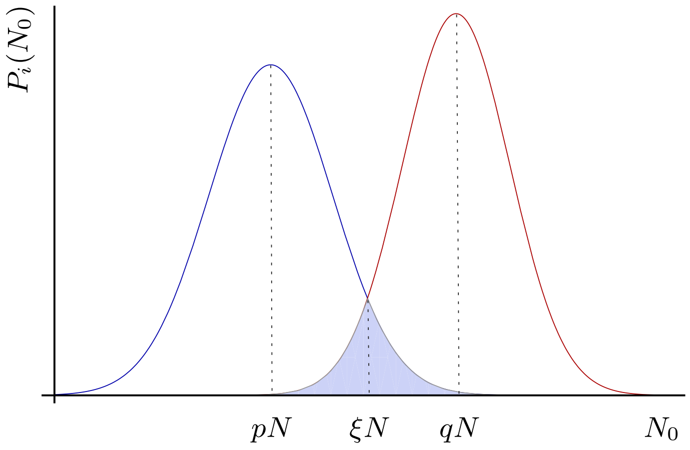


Figure 3.1. The probability distribution of a result of N_0 “heads” is represented for a biased coin that can be of types, 1 or 2. When N is large, the curves approach Gaussians centred at pN and qN, respectively, where p (q) is the bias of coin 1 (2). The filled area corresponds to the error probability in distinguishing the two distributions.

3.3 Discrimination of quantum states

One can think of the classical probability distributions in the previous Section as arising from some kind of fixed quantum measurement $\mathcal{E} = \{E_i\}$ performed over a quantum system which state is either ρ_1 or ρ_2, i.e.,

$$p_1(i) = \operatorname{tr}(E_i\rho_1)\,, \quad p_2(i) = \operatorname{tr}(E_i\rho_2)\,. \tag{3.16}$$

The problem of discriminating quantum states is essentially different to that of discriminating probability distributions in that, in the latter case, the measurement procedure is fixed. That is to say, the process of sampling the probability distributions is not under discussion, since it just consists in randomly picking a value of i (e.g. tossing a coin or rolling a die). Then, given the outcome, one optimizes the guessing part, i.e., one chooses optimally the Bayes decision function to indicate a guess. In the quantum analogue, outcomes are generated by applying a—not predetermined—measurement $\mathcal{E}$. The particular $\mathcal{E}$ used is up to the measurer's choice, and it will directly influence the probabilities of the observed outcomes. Hence it is an extra freedom of the problem. Given a set of quantum states among which one has to discriminate, one needs to optimize the two parts of the process: the measurement *plus* the guess. This combination is summed up neatly by the POVM formalism (recall Section 2.3). Generically, a POVM, that is a set of semidefinite positive operators $\{E_i\}$ such that $\sum_i E_i = \mathbb{1}$, will have as many elements as possible answers the observer may give. In other words, the occurrence of every outcome is directly associated with a particular answer (a different one, in principle) regarding the identity of the unknown state. Hence the optimization of the "measurement + guess" process boils down conveniently to optimize over all possible POVMs $\mathcal{E}$.

The other genuinely quantum feature that makes the task of discriminating quantum states both challenging and interesting is the fact that two nonorthogonal quantum states cannot be discriminated perfectly [Nielsen and Chuang, 2000]. The proof is very simple. Suppose that two pure states $|\psi_1\rangle$ and $|\psi_2\rangle$ are nonorthogonal, and that there is a measurement $\mathcal{E} = \{E_1, E_2\}$ that distinguishes them perfectly. This is mathematically represented by

$$p_1(1) = \operatorname{tr}(E_1 |\psi_1\rangle\langle\psi_1|) = 1\,, \tag{3.17}$$

$$p_2(2) = \operatorname{tr}(E_2 |\psi_2\rangle\langle\psi_2|) = 1\,. \tag{3.18}$$

Because $\mathcal{E}$ is a POVM, the completeness relation $E_1 + E_2 = \mathbb{1}$ holds. This guarantees that the probabilities add up to one, namely $p_1(1) + p_1(2) = 1$ and $p_2(1) + p_2(2) = 1$. Due to Eq. (3.17), it must happen that $p_1(2) = \operatorname{tr}(E_2 |\psi_1\rangle\langle\psi_1|) = 0$. Now, suppose the decomposition

$$|\psi_2\rangle = \alpha\,|\psi_1\rangle + \beta\,|\varphi\rangle\ , \tag{3.19}$$

where $\langle\psi_1|\varphi\rangle = 0$, $|\alpha|^2 + |\beta|^2 = 1$ by normalization, and $|\beta| < 1$ since $\langle\psi_1|\psi_2\rangle > 0$. This means that

$$p_2(2) = \operatorname{tr}(E_2 |\psi_2\rangle\langle\psi_2|) = |\beta|^2 \langle\varphi|E_2|\varphi\rangle \leqslant |\beta|^2 < 1\,, \tag{3.20}$$

which contradicts Eq. (3.18). The second last inequality follows from

$$\langle\varphi|E_2|\varphi\rangle \leqslant \sum_i \langle\varphi|E_i|\varphi\rangle = \langle\varphi|\varphi\rangle = 1\,. \tag{3.21}$$

Quantum indeterminism places in this way its footprint onto the discrimination problem. In other words, if the states to be discriminated are nonorthogonal, even if they are pure, errors will be unavoidable. Now, we may deal with these errors in different ways. The beauty of quantum discrimination resides in that we may tune the measurement $\mathcal{E}$ to meet different requisites in a certain discrimination task, existing essentially two types of approaches in what errors are concerned: *minimum error discrimination* and *unambiguous discrimination*. In a nutshell, the former allows for errors while enforces a guess after each measurement, whereas the latter sets a zero-error condition in guesses by allowing for some chance of abstaining to make a guess. Additionally, there exists a third approach that interpolates between the two extremes: *discrimination with error margins*. The following three sections review the basics of each approach, with the focus placed over the discrimination between two hypotheses.

3.3.1 Minimum-error discrimination

Carl W. Helstrom pioneered the study of discrimination problems in quantum mechanics in [Helstrom, 1976] within the context of hypothesis testing, introduced in Section 3.2, but applied to quantum states. The scenario is a particular instance of the paradigm of two agents outlined in Section 3.1. A sender prepares a quantum system in either the state ρ_1 or the state ρ_2 (pure or mixed), with a priori probabilities η_1 and η_2, and sends it to us. Our task is to identify the state of the system by applying some measurement $\mathcal{E} = \{E_1, E_2\}$ and making a guess according to the obtained outcome: we shall guess that the state was ρ_1 if the outcome 1 is obtained, whereas the outcome 2 would indicate us to guess ρ_2. The problem consists in finding the optimal strategy, that is the optimal binary-valued POVM $\mathcal{E}$, that accomplishes the task while minimizing the average probability of error.

For an arbitrary $\mathcal{E}$, the average probability of error is

$$P_e(\mathcal{E}) = \eta_1 \mathrm{tr}\,(E_2\rho_1) + \eta_2 \mathrm{tr}\,(E_1\rho_2)\,, \tag{3.22}$$

that is the probability of obtaining the outcome 2 when the state was ρ_1 times its a priori probability, plus a similar term for ρ_2. Using the fact that $E_2 = \mathbb{1} - E_1$, Eq. (3.22) becomes

$$\begin{aligned} P_e(\mathcal{E}) &= \eta_1 \mathrm{tr}\,[(\mathbb{1} - E_1)\,\rho_1] + \eta_2 \mathrm{tr}\,(E_1\rho_2) \\ &= \eta_1 + \eta_2 \mathrm{tr}\,(E_1\rho_2) - \eta_1 \mathrm{tr}\,(E_1\rho_1) \\ &= \eta_1 + \mathrm{tr}\,(E_1\Gamma)\,, \end{aligned} \tag{3.23}$$

where

$$\Gamma = \eta_2\rho_2 - \eta_1\rho_1 \tag{3.24}$$

is the so-called *Helstrom matrix.* Note that, if $E_1 = \mathbb{1} - E_2$ is used instead, one obtains the similar expression

$$P_e(\mathcal{E}) = \eta_2 - \mathrm{tr}\,(E_2\Gamma)\,. \tag{3.25}$$

The minimum error probability is just

$$P_e \equiv P_e(\mathcal{E}^*) = \min_{\mathcal{E}} P_e(\mathcal{E})\,, \tag{3.26}$$

and the (optimal) POVM $\mathcal{E}^* = \{E_1^*, E_2^*\}$ that accomplishes it receives the name of *Helstrom measurement.* The explicit expression for P_e was originally derived in [Helstrom, 1976], although a simpler and more insightful method can be found, e.g., in [Bergou *et al.*, 2004]. It works as follows. First, note that Γ can have, in general, positive as well as negative and zero eigenvalues. Let its spectral decomposition be

$$\Gamma = \sum_{k=1}^{d} \gamma_k \, |\varphi_k\rangle\langle\varphi_k| \ , \tag{3.27}$$

where d is the dimension of the Hilbert space of the system. Without loss of generality one can order the eigenvalues γ_k as

$$\begin{aligned} \gamma_k < 0 \quad &\text{for} \quad 1 \leqslant k < k_0 \, , \\ \gamma_k > 0 \quad &\text{for} \quad k_0 \leqslant k \leqslant D \, , \\ \gamma_k = 0 \quad &\text{for} \quad D < k \leqslant d \, . \end{aligned} \tag{3.28}$$

Plugging Eq. (3.27) into Eq. (3.23) one has

$$P_e(\mathcal{E}) = \eta_1 + \sum_{k=1}^{d} \gamma_k \, \langle\varphi_k|E_1|\varphi_k\rangle \ . \tag{3.29}$$

The constraint $0 \leqslant \langle\varphi_k|E_1|\varphi_k\rangle \leqslant 1$ holds, since $\mathrm{tr}\,(E_1\rho)$ must be a probability for any ρ. It immediately follows that the optimal POVM element E_1^*, that is the one that minimizes Eq. (3.29), must verify $\langle\varphi_k|E_1^*|\varphi_k\rangle = 1$ when $\gamma_k < 0$, and $\langle\varphi_k|E_1^*|\varphi_k\rangle = 0$ when $\gamma_k > 0$. Hence the elements of $\mathcal{E}^*$ can be written as

$$E_1^* = \sum_{k=1}^{k_0-1} |\varphi_k\rangle\langle\varphi_k| \ , \quad E_2^* = \mathbb{1} - E_1^* = \sum_{k=k_0}^{d} |\varphi_k\rangle\langle\varphi_k| \ . \tag{3.30}$$

The projectors onto the eigenstates of Γ associated with the eigenvalues $\gamma_k = 0$ appear in E_2^* to complete the identity operator, but this is an arbitrary choice. They may be shared in any way between E_1^* and E_2^*, for it has no effect on the value of P_e.

Summing up, the optimal measurement operators E_1^* and E_2^* are projectors onto the orthogonal subspaces of negative and positive eigenvalues of the Helstrom matrix Γ, respectively. The projector onto the subspace of zero eigenvalues of Γ, needed to fulfil the completeness relation $E_1^* + E_2^* = \mathbb{1}$, may be chosen in any way. Interestingly, if there are no negative eigenvalues, the measurement operators turn to be $E_1^* = 0$ and $E_2^* = \mathbb{1}$. This situation corresponds to the optimal strategy being to always guess that the state is ρ_2, i.e., there is no need to measure the system at all (an equivalent situation arises when there are no positive eigenvalues). Plugging Eq. (3.30) into Eqs. (3.23) and (3.25), one finds

$$P_e = \eta_1 - \sum_{k=1}^{k_0-1} |\gamma_k| = \eta_2 - \sum_{k=k_0}^{D} |\gamma_k| \, . \tag{3.31}$$

Taking the sum of these two alternative forms of P_e and using $\eta_1 + \eta_2 = 1$ leads to

$$P_e = \frac{1}{2}\left(1 - \sum_k |\gamma_k|\right) = \frac{1}{2}\left(1 - \operatorname{tr}|\Gamma|\right) . \tag{3.32}$$

This is the well-known Helstrom formula for the minimum error probability in discriminating ρ_1 and ρ_2, more commonly written as

$$P_e = \frac{1}{2}\left(1 - \|\eta_1\rho_1 - \eta_2\rho_2\|_1\right) , \tag{3.33}$$

where $\|A\|_1 = \operatorname{tr}|A| = \sqrt{A^\dagger A}$ is the *trace norm* operation.

The form of Eq. (3.33) becomes much simpler in the special case of pure states, that is when $\rho_1 = |\psi_1\rangle\langle\psi_1|$ and $\rho_2 = |\psi_2\rangle\langle\psi_2|$:

$$P_e = \frac{1}{2}\left(1 - \sqrt{1 - 4\eta_1\eta_2|\langle\psi_1|\psi_2\rangle|^2}\right) . \tag{3.34}$$

When the states are orthogonal, that is $\langle\psi_1|\psi_2\rangle = 0$, the discrimination can be done perfectly and $P_e = 0$. In contrast, if $\langle\psi_1|\psi_2\rangle = 1$, that is the case of indistinguishable states, the error probability depends only on the a priori knowledge contained in the probabilities η_1 and η_2. When $\eta_1 = \eta_2 = 1/2$, one has $P_e = 1/2$ since one can do no more than guessing randomly one of the states.

It is worth mentioning that, for pure states, the matrix Γ has rank 2 and, consequently, it has only one positive and one negative eigenvalue. Thus everything can be considered to happen in the two-dimensional subspace $\mathcal{S} = \operatorname{span}\{|\psi_1\rangle, |\psi_2\rangle\}$, just as if $|\psi_1\rangle$ and $|\psi_2\rangle$ were qubit states. This simplification allows for the simple and useful geometrical representation of the states and the POVM elements as vectors in a plane. Let $\{|0\rangle, |1\rangle\}$ be an orthonormal basis of $\mathcal{S}$. Then, we can always write the states as

$$|\psi_i\rangle = \cos\frac{\theta}{2}|0\rangle - (-1)^i \sin\frac{\theta}{2}|1\rangle , \quad i = 1, 2 , \tag{3.35}$$

where $0 \leqslant \theta < \pi/2$ and $|\langle\psi_1|\psi_2\rangle| = \cos\theta$. Similarly, since $\mathcal{S}$ is two-dimensional, the POVM elements E_i need to be one-dimensional orthogonal projectors, i.e., $E_i = |\varphi_i\rangle\langle\varphi_i|$ for $i = 1, 2$, with

$$|\varphi_1\rangle = \cos\frac{\phi}{2}|0\rangle + \sin\frac{\phi}{2}|1\rangle , \tag{3.36}$$

$$|\varphi_2\rangle = \cos\frac{\pi - \phi}{2}|0\rangle - \sin\frac{\pi - \phi}{2}|1\rangle , \tag{3.37}$$

and $\theta \leqslant \phi \leqslant \pi - \theta$. The optimization procedure consists in finding the optimal orientation of the pair of orthogonal vectors $|\varphi_i\rangle$, i.e., the optimal

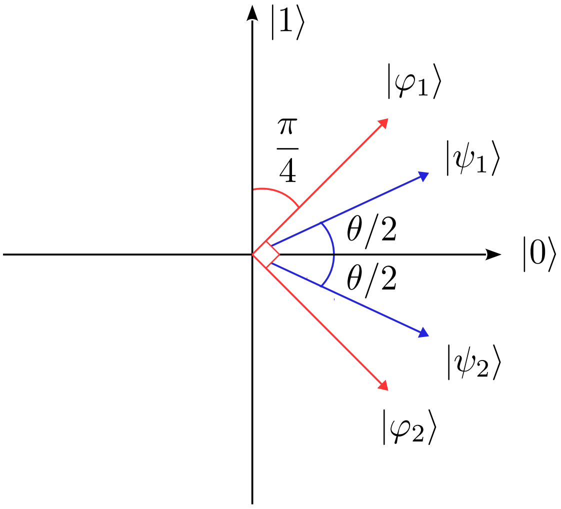


Figure 3.2. Optimal orientation of the POVM vectors $|\varphi_i\rangle$ with respect to the states $|\psi_i\rangle$ for minimum-error discrimination.

angle ϕ, such that $P_e(\mathcal{E})$, as defined in Eq. (3.22), is minimized. When the a priori probabilities are equal, the optimal orientation is symmetric with respect to the states $|\psi_i\rangle$, that is an angle $\phi = \pi/4$ (see Fig. 3.2). When $\eta_1 > \eta_2$, one just has to rotate the pair of vectors $|\varphi_i\rangle$ clockwise such that the overlap $\langle\varphi_1|\psi_1\rangle$ increases (and $\langle\varphi_2|\psi_2\rangle$ decreases accordingly). Such an increase translates into a greater probability of detection $\operatorname{tr}(E_1\rho_1)$. The reverse situation occurs when $\eta_1 < \eta_2$. The optimal angles in these asymmetrical cases are trivially obtained from Eqs. (3.22), (3.35), (3.36) and (3.37), and the corresponding minimum error probability is given by Eq. (3.34).

3.3.2 Unambiguous discrimination

The minimum error approach to the discrimination problem considered in the previous Section assumes by default a nonzero chance for erroneous guesses if the states to discriminate are nonorthogonal. There, a solution is considered optimal if this chance is minimized. However, there might be cases in which errors cannot be tolerated under any circumstances. Can one still say something about the identity of the unknown quantum state under such restriction? This question was first addressed by Ivanovic for the case of discriminating between two possible pure states $\rho_1 = |\psi_1\rangle\langle\psi_1|$ and $\rho_2 = |\psi_2\rangle\langle\psi_2|$

[Ivanovic, 1987][11].

The type of measurements described for minimum error discrimination can be used to produce an outcome with no errors associated. Take a projective two-outcome measurement with elements $E_i = |\varphi_i\rangle\langle\varphi_i|$ $(i = 1, 2)$ defined through Eqs. (3.36) and (3.37), and set the extreme angle $\phi = \pi - \theta$. This angle makes the vector $|\varphi_1\rangle$ orthogonal to $|\psi_2\rangle$. The consequence is that the operator E_1 never "clicks" whenever the state is ρ_2, i.e., $\mathrm{tr}\,(E_1\rho_2) = 0$. Thus, if the outcome 1 is obtained, one can guess *with certainty* that the state was ρ_1. Unfortunately, for this value of ϕ it also happens that $|\varphi_2\rangle$ is parallel to $|\psi_2\rangle$ and hence nonorthogonal to $|\psi_1\rangle$. This means that the outcome 2 is not as reliable as the outcome 1, for it will occur with some nonzero probability both if the state is ρ_1 and if it is ρ_2. The other extreme angle $\phi = \theta$ yields the reverse situation, in which outcome 2 is error free and outcome 1 is uncertain. Ivanovic proposed that, provided several copies of the unknown state, a series of these two measurements can be used to give conclusive guesses, at the expense of not making any guess if an unreliable outcome is obtained. A year later Dieks unified this sequence of measurements under a single POVM realization [Dieks, 1988], and Peres proved that such POVM is optimal in the sense that it provides a minimum probability of inconclusive outcomes [Peres, 1988].

The optimal solution for unambiguously discriminating two equally probable arbitrary pure states is known as the Ivanovic-Dieks-Peres (IDP) result. It invokes the use of a POVM with *three*[12] elements $\mathcal{E} = \{E_1, E_2, E_0\}$. The element E_1 should identify with certainty the state as ρ_1, the element E_2 should identify it as ρ_2 also with certainty, and the element E_0 completes the POVM and represents an inconclusive outcome. This is to say, if such outcome is obtained, the measurer learns nothing about the identity of the state and he abstains from giving an answer. The unambiguous guessing requirement is mathematically represented by the condition

$$\mathrm{tr}\,(E_1\rho_2) = \mathrm{tr}\,(E_2\rho_1) = 0\,. \tag{3.38}$$

[11] Historically, unambiguous discrimination was introduced first for pure states, and only recently some results for mixed states have appeared. Just the opposite as minimum error discrimination, that started from the general case of two mixed states, and from which the pure states cases are derived.

[12] Unambiguous discrimination is one example of a task which optimal solution requires the more general POVM formulation of quantum measurements, for it needs to overcome the limitation that von Neumann measurements impose to the number of outcomes—that of being equal to the dimension of the Hilbert space spanned by the states.

This condition enforces the POVM elements to be of the form

$$E_1 = \mu_1|\psi_2^\perp\rangle\langle\psi_2^\perp|\,, \tag{3.39}$$

$$E_2 = \mu_2|\psi_1^\perp\rangle\langle\psi_1^\perp|\,, \tag{3.40}$$

$$E_0 = \mathbb{1} - E_1 - E_2\,, \tag{3.41}$$

where μ_1 and μ_2 are two coefficients yet to be determined by optimality, and $|\psi_i^\perp\rangle$ stands for a vector orthogonal to $|\psi_i\rangle$. One must now realize two facts. On the one hand, for equally probable states the probabilities of outcomes 1 and 2 should be equal by symmetry, hence one can safely assume that $\mu_1 = \mu_2 = \mu$. On the other hand, since $\mathcal{E}$ is a POVM its elements must be semidefinite positive, i.e., the conditions $\mu \geqslant 0$ and $E_0 \geqslant 0$ must hold. The latter can be assured by using the decomposition $|\psi_2^\perp\rangle = c|\psi_1^\perp\rangle + \sqrt{1-c^2}\,|\psi_1\rangle$, where $c = |\langle\psi_1|\psi_2\rangle|$, to diagonalize E_0 and impose positivity for its eigenvalues. This leads to the condition

$$\mu \leqslant \frac{1}{1+c}\,. \tag{3.42}$$

The probability of obtaining an inconclusive outcome is defined as

$$Q = \mathrm{tr}\,(E_0\rho) = \mathbb{1} - \mathrm{tr}\,(E_1\rho) - \mathrm{tr}\,(E_2\rho) = 1 - \mu(1-c^2)\,, \tag{3.43}$$

where $\rho = \rho_1/2 + \rho_2/2$. Note that $\mathcal{E}$ is fully determined by the parameter μ. The only thing left to do is to choose μ such that Q is minimized. This happens for the maximum value $\mu = 1/(1+c)$, and yields the minimum probability of inconclusive results

$$Q = c\,, \tag{3.44}$$

and consequently the maximum probability of successful unambiguous discrimination

$$P_s = \mathrm{tr}\,(E_1\rho_1) + \mathrm{tr}\,(E_2\rho_2) = 1 - c\,. \tag{3.45}$$

Eqs. (3.44) and (3.45) comprise the IDP result. The graphical representation of the optimal $\mathcal{E}$ for unambiguous discrimination of equally-probable pure states is depicted in Fig. 3.3.

This result was generalized by Jaeger and Shimony to the case of arbitrary prior probabilities η_1 and η_2 [Jaeger and Shimony, 1995]. The bottom line of their argument is that a three-outcome POVM as described by Eqs. (3.39-3.41) is optimal for unambiguous discrimination, but only when it exists, and it does not so in the whole range of values for the prior probabilities. The

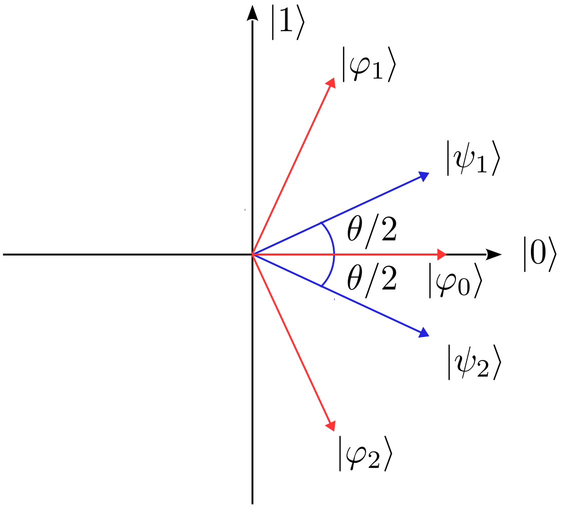


Figure 3.3. Optimal orientation of the POVM with respect to the states $|\psi_i\rangle$ for unambiguous discrimination. The vectors associated to the POVM elements E_1 and E_0 are $|\varphi_1\rangle = |\psi_2^\perp\rangle$ and $|\varphi_2\rangle = |\psi_1^\perp\rangle$, respectively. The inconclusive element is $E_0 \sim |\varphi_0\rangle\langle\varphi_0|$.

existence of such POVM is determined by whether the detection probabilities that it generates are valued between 0 and 1. Using $\eta_1 + \eta_2 = 1$, it is not difficult to show that the POVM exists in the range

$$\frac{c^2}{1+c^2} \leqslant \eta_1 \leqslant \frac{1}{1+c^2} \,. \tag{3.46}$$

Above this range, the optimal POVM turns out to be the first two-outcome projective measurement described at the beginning of this Section, which elements $\{E_1, E_0\}$ either identify with certainty the state ρ_1 or produce an inconclusive answer. Below this range, the optimal POVM is the second one described, with elements $\{E_2, E_0\}$. The general solution for arbitrary η_1 and η_2 renders the optimal inconclusive probability

$$Q = \begin{cases} \eta_1 + \eta_2 c^2 & \text{if} \quad \eta_1 < \frac{c^2}{1+c^2} \\ 2\sqrt{\eta_1\eta_2}\,c & \text{if} \quad \frac{c^2}{1+c^2} \leqslant \eta_1 \leqslant \frac{1}{1+c^2} \\ \eta_1 c^2 + \eta_2 & \text{if} \quad \frac{1}{1+c^2} < \eta_1 \end{cases} \,. \tag{3.47}$$

The IDP result obtained in the late 80's, in addition to Jaeger and Shimony's generalization in 1995, completely solve the problem of unambiguous discrimination of two pure states. Results related to mixed states appeared much later. A reason for this delay may be attributed to the following common statement, phrased, e.g., by Fiurášek and Ježek as: "[...] it is known that one cannot unambiguously discriminate mixed states (the reason is that the IDP scheme does not work for linearly dependent states)." [Fiurášek and Ježek, 2003]. Indeed, the IDP method cannot be straightforwardly generalized, or, more precisely, it does not apply to general *full-rank* mixed states. This is so because in such case both hypotheses have the same support[13], hence a measurement operator cannot project onto a subspace that is orthogonal to the support of only one hypothesis, which is the trick that allows to conclusively say that the true hypothesis is the other one when the corresponding outcome is obtained. It is possible, however, to unambiguously discriminate mixed states which do not have the same support. Along this line are, for instance, the tasks of unambiguous discrimination between sets of states or *unambiguous filtering* [Sun *et al.*, 2002], *state comparison* [Barnett *et al.*, 2003; Jex *et al.*, 2004] and unambiguous programmable state discrimination, also known as *unambiguous identification* (see Chapter 4). While these tasks have case-specific solutions, results of a more general nature can be found in [Rudolph *et al.*, 2003; Herzog and Bergou, 2005; Raynal, 2006].

3.3.3 Discrimination with an error margin

Unambiguous and minimum-error discrimination are the two extremes of a more general scheme. Intuitively, if the unambiguous scheme is relaxed by tolerating some error rate, the success probability can be increased. Likewise, by allowing some rate of inconclusive answers in the minimum-error scheme, the reliability of the answers can also be increased. These relaxations of the zero-error condition (unambiguous scheme) and the always-guess condition (minimum-error scheme) yield two different parametrizations of the same unified approach to the problem. In the former case, the discrimination protocol is optimized for a *fixed rate of inconclusive outcomes* Q[14]. In the latter, the optimal protocol is derived for a given *error margin* r that the probability of error must not exceed[15]. In both cases the optimization is carried out by maximizing the probability of success, and both are equivalent

[13] The support of a state, described by a density matrix, is defined as the subspace spanned by its eigenvectors associated to nonzero eigenvalues.

[14] Analytical solutions for simple cases, numerical solutions and useful bounds were derived in [Chefles and Barnett, 1998b; Zhang *et al.*, 1999; Fiurášek and Ježek, 2003; Eldar, 2003], and a general method for converting the problem into a standard minimum-error discrimination between some stochastically transformed states was recently obtained in [Bagan *et al.*, 2012]. The techniques derived there were also successfully applied to quantum state estimation with postprocessing in [Gendra *et al.*, 2012, 2013].

[15] This scheme was first considered in [Touzel *et al.*, 2007] for projective measurements. The solution for pure states allowing generalized measurements was derived in [Hayashi *et al.*, 2008; Sugimoto *et al.*, 2009].

ways to connect smoothly the unambiguous and the minimum-error extremes.

These general scenarios cover many practical situations, in which only a limited rate of inconclusive answers is affordable, or a certain low error rate is tolerable. Also, cases of linearly dependent states or full rank mixed states, where unambiguous discrimination is not possible, are in principle tractable under this general scheme, providing a way to increase the success probability over that provided by minimum-error discrimination.

In this Section, I describe the unified scheme for pure states in terms of an error margin. The results that follow were first obtained in [Hayashi *et al.*, 2008; Sugimoto *et al.*, 2009], but I present them here in a simpler way[16].

Consider two pure nonorthogonal states $\rho_1 = |\psi_1\rangle\langle\psi_1|$, $\rho_2 = |\psi_2\rangle\langle\psi_2|$ as hypotheses of a standard two-state discrimination problem, where for simplicity we assign equal prior probabilities to each state. The discrimination with an error margin protocol can be thought of as a generalized measurement on the system, described by the POVM $\mathcal{E} = \{E_1, E_2, E_0\}$, where, as in Section 3.3.2, the operator E_1 (E_2) is associated to the statement "the measured state is ρ_1 (ρ_2)", whereas E_0 is associated to the inconclusive answer or abstention. The overall success, error and inconclusive probabilities are

$$\begin{aligned} P_\mathrm{s} &= \frac{1}{2}\left[\mathrm{tr}\,(E_1\rho_1) + \mathrm{tr}\,(E_2\rho_2)\right] , && (3.48)\\ P_\mathrm{e} &= \frac{1}{2}\left[\mathrm{tr}\,(E_2\rho_1) + \mathrm{tr}\,(E_1\rho_2)\right] , && (3.49)\\ Q &= \frac{1}{2}\left[\mathrm{tr}\,(E_0\rho_1) + \mathrm{tr}\,(E_0\rho_2)\right] , && (3.50) \end{aligned}$$

respectively. The relation $P_\mathrm{s} + P_\mathrm{e} + Q = 1$ is guaranteed by the POVM condition $E_0 + E_1 + E_2 = \mathbb{1}$. The optimal discrimination with an error margin protocol is obtained by maximizing the success probability P_s over any possible POVM $\mathcal{E}$ that satisfies that certain errors occur with a probability not exceeding the given margin. Generically, these conditions imply a non vanishing value of the inconclusive probability Q.

We consider two error margin conditions: *weak* and *strong*. The weak condition states that the *average* error probability cannot exceed a margin, i.e.,

$$P_\mathrm{e} = \frac{1}{2}\left[\mathrm{tr}\,(E_2\rho_1) + \mathrm{tr}\,(E_1\rho_2)\right] \leqslant r\,. \tag{3.51}$$

The strong condition imposes a margin on the probabilities of misidentifying

[16] The remaining of this Section follows closely the first part of [Sentís *et al.*, 2013].

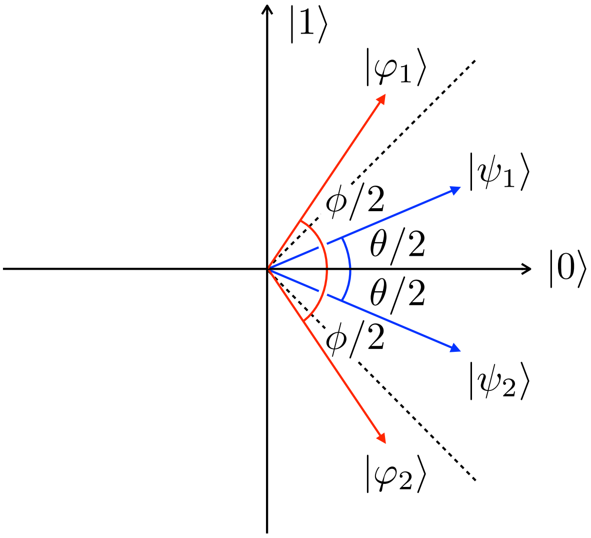


Figure 3.4. Parametrization of the states $|\psi_1\rangle$, $|\psi_2\rangle$, $|\varphi_1\rangle$ and $|\varphi_2\rangle$ as in Eqs. (3.54) and (3.55). The dashed lines, at an angle of $\pi/4$ with respect to the horizontal axis, represent the limit of minimum-error discrimination.

each possible state, i.e.,

$$p(\rho_2|E_1) = \frac{\text{tr}\,(E_1\rho_2)}{\text{tr}\,(E_1\rho_1)+\text{tr}\,(E_1\rho_2)} \leqslant r\,, \tag{3.52}$$

$$p(\rho_1|E_2) = \frac{\text{tr}\,(E_2\rho_1)}{\text{tr}\,(E_2\rho_1)+\text{tr}\,(E_2\rho_2)} \leqslant r\,, \tag{3.53}$$

where $p(\rho_2|E_1)$ and $p(\rho_1|E_2)$ are the probabilities that the state identified as ρ_1 is actually ρ_2 and the other way around, respectively. The strong condition is obviously more restrictive, as it sets a margin on both types of errors separately. However, as we will see, the two conditions are directly related: the strong one just corresponds to the weak one with a tighter error margin [Sugimoto *et al.*, 2009]. Note that both error margin schemes have the unambiguous (when $r = 0$) and the minimum-error schemes (when r is large enough) as extremal cases. We will denote by r_c the critical margin above which the success probability does not increase and thus coincides with that of (the unrestricted) minimum-error discrimination.

For the weak condition, it is straightforward to obtain the maximum success probability by taking into account that the corresponding error probability must saturate the margin condition (3.51) for $r \leqslant r_c$, namely $P_\text{e} = r$. Furthermore, the symmetry of the problem dictates that $\text{tr}\,(E_1\rho_1) = \text{tr}\,(E_2\rho_2) = P_\text{s}$ and $\text{tr}\,(E_1\rho_2) = \text{tr}\,(E_2\rho_1) = P_\text{e}$. Without loss of generality (see Fig. 3.4) and as in Section 3.3.1, we can use the parametrization (3.35) in terms of a single angle for the input states, i.e.,

$$|\psi_i\rangle = \cos\frac{\theta}{2}\ |0\rangle - (-1)^i \sin\frac{\theta}{2}\ |1\rangle\ , \quad i = 1,2\,, \tag{3.54}$$

where $0 \leqslant \theta < \pi/2$. The POVM elements can be as well written as $E_i = \mu\,|\varphi_i\rangle\langle\varphi_i|$ for $i = 1, 2$, with

$$|\varphi_i\rangle = \cos\frac{\phi}{2}\ |0\rangle - (-1)^i \sin\frac{\phi}{2}\ |1\rangle\ , \quad \frac{\pi}{2} \leqslant \phi < \pi \tag{3.55}$$

(in contrast to Eqs. (3.36) and (3.37), E_1 and E_2 need not be orthogonal, since in this case there is a third POVM element). The POVM condition implies $E_0 = \mathbb{1} - E_1 - E_2$, and the optimal value of μ is fixed by the extremal value of the inequality $E_0 \geqslant 0$. One obtains $\mu = 1/(1-\cos\phi) \leqslant 1$ and finally the symmetry conditions fix ϕ to be

$$\tan\frac{\phi}{2} = \begin{cases} \dfrac{\sqrt{1+c}}{\sqrt{1-c}+2\sqrt{r}} & \text{if} \quad 0 \leqslant r \leqslant r_c\,, \\ 1 & \text{if} \quad r_c \leqslant r \leqslant 1\,, \end{cases} \tag{3.56}$$

where $c = |\langle\psi_1|\psi_2\rangle| = \cos\theta$ is the overlap of the states $|\psi_1\rangle$ and $|\psi_2\rangle$. Note that in the unambiguous limit, $r = 0$, the POVM elements E_1 and E_2 are orthogonal to the states $|\psi_2\rangle$ and $|\psi_1\rangle$, respectively. In the other extreme case, when the error margin coincides with, or is larger than, the minimum error, $r \geqslant r_c$, one has $E_0 = 0$ (no abstention) and E_1 becomes orthogonal to E_2, i.e., $\phi = \pi/2$. In this range the measurement becomes of von Neumann type and the first case in Eq. (3.56) implies

$$r_c = \frac{1}{2}\left(1 - \sqrt{1-c^2}\right). \tag{3.57}$$

Taking into account Eq. (3.56), the optimal success probability reads

$$P_{\rm s}^{W}(r) = \begin{cases} \left(\sqrt{r} + \sqrt{1-c}\right)^2 & \text{if} \quad 0 \leqslant r \leqslant r_c\,, \\ \frac{1}{2}\left(1 + \sqrt{1-c^2}\right) & \text{if} \quad r_c \leqslant r \leqslant 1\,. \end{cases} \tag{3.58}$$

This result was derived in [Hayashi *et al.*, 2008] and its generalization to arbitrary prior probabilities in [Sugimoto *et al.*, 2009] (also in [Bagan *et al.*, 2012], by fixing an inconclusive rate Q instead of an error margin). Note that the POVM $\mathcal{E}$ is fully determined by the angle ϕ, which in turn is fully determined by the margin r through Eq. (3.56).

The optimal success probability under the strong condition can be obtained along the same lines of the weak case, but it will prove more convenient to use the connection between both conditions to derive it directly from (3.58). Let us denote by r^S (r^W) the error margin of the strong (weak) condition. From the symmetry of the problem, Eqs. (3.52) and (3.53) can be written in the form of a weak condition with a margin r^W as

$$P_\mathrm{e} \leqslant r^S(P_\mathrm{e}+P_\mathrm{s}) \equiv r^W. \tag{3.59}$$

Hence, if $\mathcal{E}$ is the optimal POVM for a strong margin r^S, it is also optimal for the weak margin r^W, where $P_\mathrm{e} = r^W$ and $P_\mathrm{s} = P_\mathrm{s}^W(r^W)$ is given by Eq. (3.58). In terms of the success probability, the relation between r^W and r^S reads

$$r^S = \frac{r^W}{P_\mathrm{s}^W(r^W)+r^W}\,. \tag{3.60}$$

By solving for r^W and substituting into Eq. (3.58) one derives the success probability for a given r^S, which we denote by $P_\mathrm{s}^S(r^S)$. For the function P_s^S one readily obtains

$$P_\mathrm{s}^S(r) = \begin{cases} \left(\dfrac{\sqrt{1-r}}{\sqrt{r}-\sqrt{1-r}}\right)^2(1-c) & \text{if} \quad 0 \leqslant r \leqslant r_c\,, \\ \frac{1}{2}\left(1+\sqrt{1-c^2}\right) & \text{if} \quad r_c \leqslant r \leqslant 1\,, \end{cases} \tag{3.61}$$

in agreement with [Hayashi *et al.*, 2008]. Note that the critical margin is the same for both the weak and the strong conditions, i.e., $r_c^W = r_c^S = r_c$. Indeed, beyond the critical point inconclusive results are excluded by optimality ($Q = 0$ and $P_\mathrm{s} + P_\mathrm{e} = 1$) and thus there is no difference between the two types of conditions. As in the weak case, there is a correspondence between the angle ϕ and r^S, thus $\mathcal{E}$ can also be parametrized in terms of the strong margin:

$$\tan\frac{\phi}{2} = \begin{cases} \dfrac{\sqrt{1-r^S}-\sqrt{r^S}}{\sqrt{1-r^S}+\sqrt{r^S}}\,\dfrac{\sqrt{1+c}}{\sqrt{1-c}} & \text{if} \quad 0 \leqslant r \leqslant r_c\,, \\ 1 & \text{if} \quad r_c \leqslant r \leqslant 1\,. \end{cases} \tag{3.62}$$

Note that an ambiguity arises for $c = 1$, as $\phi = \pi$ and then E_1 and E_2 become proportional to one another, independently of the value of r^S. Note also that for $r^S = 0$ and $r^S = r_c$ the values of ϕ for both, weak and strong conditions, coincide.

3.4 The many copies paradigm

As decisions in classical hypothesis testing may be based on more than one sampling of the unknown probability distribution (see Section 3.2), the discrimination of quantum states may be supported by more than one measurement of the unknown state. However, after the first measurement the state of a quantum system changes irremediably, hence a second measurement over the same system—if the first was optimal—would give no aid in the identification of the original state[17]. This is why a number of *copies* of the system, all prepared in the same unknown quantum state, is typically considered as resources in quantum state discrimination tasks.

Formally, one considers that N independent and identically-distributed (i.i.d.) states are provided. Such an ensemble of systems is described by a big d^N-dimensional Hilbert space $\mathcal{H}^{\otimes N}$, where $\mathcal{H}$ is the d-dimensional Hilbert space of each individual system. If the state of each copy is either ρ_1 or ρ_2, then one just has to discriminate the global states $\rho_i \otimes \rho_i \otimes \ldots \otimes \rho_i \equiv \rho_i^{\otimes N}$, $i = 1, 2$, where $\otimes$ is the direct Kronecker product of the density matrices.

It is in the possible measurements that are at one's disposal where quantum discrimination differs the most from its classical counterpart, for quantum mechanics allows for sophisticated measurements on all N systems at once. Such *collective* measurements typically outperform any strategy based on individual measurements [Peres and Wootters, 1991], although there are cases in which they give no advantage. The question of whether a collective measurement strategy is necessary to achieve optimal performance represents the crux of many works in quantum state discrimination. A paradigmatic example for which this is true can be found in the context of unambiguous discrimination [Chefles, 2001]: a set of linearly dependent states—thus not unambiguously distinguishable—can be made linearly independent if enough copies of the states are provided; one can then unambiguously determine the collective state of the set of systems through a collective measurement. On the other hand, in binary minimum-error discrimination, the optimal performance is achievable through *local operations and classical communication*[18] (LOCC) if the states are pure [Acín *et al.*, 2005], but not if they are mixed [Calsamiglia *et al.*, 2010; Higgins *et al.*, 2011].

The POVM formalism covers all possible measurements, thus any mea-

[17] Although a second observer, with no knowledge about the result of the first measurement, could still "scavenge" information about the state that was previously measured [Rapčan *et al.*, 2011].

[18] This denomination stands for any strategy consisting of sequential adaptive measurements performed on each system: the result of measuring the first system determines the measurement to be used in the second, and so on.

surement for discriminating $\rho_1^{\otimes N}$ and $\rho_2^{\otimes N}$ can still be characterized by a two- or a three-outcome POVM just as in Sections 3.3.1, 3.3.2 and 3.3.3, but which elements E_i now operate over the total Hilbert space $\mathcal{H}^{\otimes N}$. It is then straightforward to generalize the Helstrom formula for single-copy minimum-error discrimination, that is Eq. (3.33), to the N-copy case: following identical steps, one simply obtains

$$P_e(N) = \frac{1}{2}\left(1 - \left\|\eta_1\rho_1^{\otimes N} - \eta_2\rho_2^{\otimes N}\right\|_1\right) . \tag{3.63}$$

Note that the derivation of this formula imposes no additional constraints over the operators E_i (apart from the POVM conditions), hence the measurement that achieves the limit (3.63) is, in principle, a collective one. Although the problem is formally solved, the computational cost of the trace norm grows exponentially with N. General analytical results for arbitrary N and arbitrary states are scarce, existing only bounds for $P_e(N)$ [Audenaert *et al.*, 2012]. The remaining of the Section is devoted to present two results that enable tractable analytical expressions of $P_e(N)$ in special cases. The first is a mathematical tool that will prove useful in Chapters 4 and 5 for obtaining analytical results when the number of copies is kept finite. The second concerns the asymptotic expression $P_e(N \to \infty)$.

3.4.1 Irreducible representations and block decomposition.

The purpose of this Section is to present a particular decomposition of density operators of multicopy systems. It was introduced in [Vidal *et al.*, 1999; Cirac *et al.*, 1999] within the context of estimation and purification of qubits, respectively, and later applied to the full estimation of qubit mixed states in [Bagan *et al.*, 2006]. Although here I will focus on qubit systems ($d = 2$), it is straightforward to extend the decomposition to systems of dimension $d > 2$ by including the irreducible representations of SU(d) in the formalism.

A set of N qubit systems in the state ρ is represented by the density operator $\rho^{\otimes N}$. This operator is invariant under the permutation of any pair of qubits, thus invariant under the action of the symmetric group S_N. One may use the group S_N to write $\rho^{\otimes N}$ in the basis of the SU(2) invariant subspaces of $\left(\frac{\mathbf{1}}{\mathbf{2}}\right)^{\otimes N}$ [bold characters stand for the irreducible representations of SU(2)], in a similar way as it is used to obtain the Clebsch-Gordan decomposition in SU(2). The relation between the tensor-product (decoupled) representation and that of the invariant subspaces (coupled) is

$$\left(\frac{\mathbf{1}}{\mathbf{2}}\right)^{\otimes N} = \bigoplus_{j,\alpha} \mathbf{j}^{(\alpha)} , \tag{3.64}$$

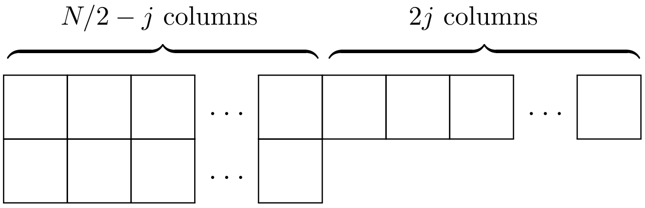


Figure 3.5. A generic Young diagram with N boxes.

where $j = 0\,(1/2), \ldots, J = N/2$ for even (odd) N, and α labels the different equivalent irreducible representations $\mathbf{j}$, i.e., $\alpha = 1, \ldots, \nu_j$, where ν_j is the multiplicity of $\mathbf{j}$. The density operator $\rho^{\otimes N}$, written in the invariant subspaces basis, has the block-diagonal form

$$\rho^{\otimes N} = \bigoplus_{j,\alpha} \rho_j^{(\alpha)} \,, \tag{3.65}$$

where $\rho_j^{(\alpha)}$ represents the block associated to the subspace $\mathbf{j}^{(\alpha)}$.

The explicit form of the blocks can be easily obtained by analyzing the Young diagrams that can be constructed with N boxes, one for each qubit. There will be as many different $\mathbf{j}$ as Young diagrams[19]. A particular $\mathbf{j}$ corresponds to a diagram with $N/2 - j$ double-box and $2j$ single-box columns (see Fig. 3.5), where each of the former is associated to a fully-antisymmetric two-qubit state or *singlet*, and the remaining to a fully-symmetric state of $2j$ qubits. This means that the matrix $\rho_j^{(\alpha)}$ has dimension $2j + 1$, and each singlet contributes a multiplicative factor $\det \rho$ to it.

Let r be the purity of the state ρ and $\vec{v}$ its Bloch vector. Let $\{|j, m, \alpha\rangle\}$ be a basis of the subspace $\mathbf{j}^{(\alpha)}$ (in analogy to the angular momentum basis), constructed from the computational basis $\{|0\rangle, |1\rangle\}$ of a single qubit. If $\vec{v} = \hat{z}$, the matrix $\rho_j^{(\alpha)}$ is diagonal in the basis $\{|j, m, \alpha\rangle\}$ and its expression is easily deduced. Since

$$\det \rho = \frac{1 - r^2}{4} \,, \tag{3.66}$$

one can write $\rho_j^{(\alpha)}$ as

$$\rho_j^{(\alpha)} = \left(\frac{1 - r^2}{4}\right)^{N/2 - j} \sum_{m=-j}^{j} \left(\frac{1 - r}{2}\right)^{j - m} \left(\frac{1 + r}{2}\right)^{j + m} |j, m, \alpha\rangle\langle j, m, \alpha| \;. \tag{3.67}$$

[19] Given a Young diagram, the value of the associated label α corresponds to a specific Young tableau for that diagram (see below). As the explicit form of $\rho_j^{(\alpha)}$ does not depend on α, one only needs to focus on Young diagrams for now.

For an arbitrary direction $\vec{v}$, it suffices to rotate the basis elements $|j,m,\alpha\rangle$ by means of the Wigner matrices $D(\vec{v})$ [Edmonds, 1960]. From the standard definition, one has

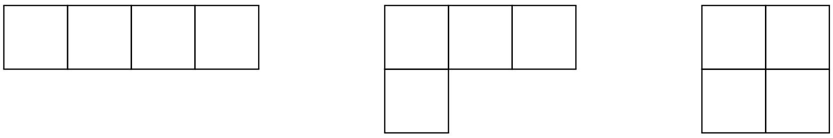
$$|j,m,\alpha\rangle_{\vec{v}} = U_{\vec{v}}^{\otimes n}\,|j,m,\alpha\rangle = \sum_{m'} \mathscr{D}^{j}_{m',m}(\vec{v})\,|j,m',\alpha\rangle \ , \tag{3.68}$$

where $U_{\vec{v}} \in \mathrm{SU}(2)$ is a rotation on a single copy, and the matrix elements are $\mathscr{D}^{j}_{m',m} = \langle j,m',\alpha|\,D(\vec{v})\,|j,m,\alpha\rangle$. Hence $\rho_j^{(\alpha)}$ takes the general form

$$\begin{aligned}\rho_j^{(\alpha)} = \left(\frac{1-r^2}{4}\right)^{N/2-j} \sum_{m=-j}^{j} \left(\frac{1-r}{2}\right)^{j-m} \left(\frac{1+r}{2}\right)^{j+m} \\ \otimes\, D(\vec{v})\,|j,m,\alpha\rangle\!\langle j,m,\alpha|\,D^{\dagger}(\vec{v})\,, \end{aligned} \tag{3.69}$$

which is the same for all the equivalent irreducible representations (i.e., its coefficients do not depend on the label α). Note that for pure states $\rho^{\otimes N}$ has projection only in the symmetric $(N+1)$-dimensional subspace $\mathbf{J} = \mathbf{N/2}$, whereas for mixed states it has components in all subspaces, including equivalent representations, $\mathbf{j}^{(\alpha)}$.

The only thing left to do is to determine how many equivalent irreducible representations are for each $\mathbf{j}$, that is the multiplicity ν_j. It reads off from simple combinatorics. The value of j associated to a subspace $\mathbf{j}^{(\alpha)}$ is determined by the shape of its Young diagram, that is the particular partition of N boxes in two rows such that the length of the second row is equal or shorter than that of the first. The different values that α can take correspond to all the possible *standard* Young tableaux that can be built with that diagram. Given a diagram, a Young tableau is obtained by filling the boxes with integer numbers, from 0 to N; it is called standard if the following rules are fulfilled: (i) the entries in each row are in increasing order, from left to right, and (ii) the entries in each column are in increasing order, from top to bottom. For example, for $N = 4$ the possible Young diagrams are



They are associated to the subspaces $\mathbf{j} = \mathbf{2}$, $\mathbf{j} = \mathbf{1}$, and $\mathbf{j} = \mathbf{0}$, respectively. With the first diagram only one standard Young tableaux can be constructed: $\begin{array}{|c|c|c|c|}\hline 1&2&3&4\\\hline\end{array}$. With the second, $\begin{array}{|c|c|c|}\hline 1&2&3\\\hline 4\\\cline{1-1}\end{array}$, $\begin{array}{|c|c|c|}\hline 1&3&4\\\hline 2\\\cline{1-1}\end{array}$, and $\begin{array}{|c|c|c|}\hline 1&2&4\\\hline 3\\\cline{1-1}\end{array}$. Finally, for the third diagram one finds $\begin{array}{|c|c|}\hline 1&2\\\hline 3&4\\\hline\end{array}$ and $\begin{array}{|c|c|}\hline 1&3\\\hline 2&4\\\hline\end{array}$. This means that, in the representation of invariant subspaces, the fully-symmetric subspace **1** occurs one time, the subspace **2** occurs three times and **0** occurs two times[20]. It is easy to convince oneself that the multiplicity ν_j of an arbitrary subspace **j** can be written as

$$\nu_j = \binom{N}{N/2-j}\frac{2j+1}{N/2+j+1}\,. \tag{3.70}$$

The block-decomposition of $\rho^{\otimes N}$, comprised by Eqs. (3.65), (3.69) and (3.70), turns out to be very useful in the computation of Eq. (3.63). Since the trace norm operation is base independent, one can write the states $\rho_1^{\otimes N}$ and $\rho_2^{\otimes N}$ in the basis that block-diagonalizes them to split the trace norm over the global states into a sum of trace norms over each orthogonal subspace (hence reducing drastically the dimension of the matrices involved in the computation), i.e.,

$$\left\|\eta_1\rho_1^{\otimes N} - \eta_2\rho_2^{\otimes N}\right\|_1 = \sum_{j=0,1/2}^{N/2} \nu_j \left\|\eta_1\rho_{1,j} - \eta_2\rho_{2,j}\right\|_1 \,. \tag{3.71}$$

Furthermore, for each j, the contribution of all the equivalent representations $\mathbf{j}^{(\alpha)}$ boils down to a multiplicative factor (its multiplicity ν_j), since $\rho_j^{(\alpha)}$ is the same matrix for all values of α.

3.4.2 Infinitely many copies: the quantum Chernoff bound

In the same spirit as Section 3.2.1, it is interesting to study the behaviour of the minimum-error probability in the asymptotic limit of infinite copies. As it happens with the minimum-error probability for distinguishing classical probability distributions, the trace norm, as a distance measure between quantum states, lacks monotonicity under the increase of the tensor powers of its arguments. That is to say, it is not difficult to find two pairs of states ρ_1, ρ_2 and σ_1, σ_2 for which $\|\rho_1-\rho_2\|_1 < \|\sigma_1-\sigma_2\|_1$, but $\left\|\rho_1^{\otimes 2}-\rho_2^{\otimes 2}\right\|_1 > \left\|\sigma_1^{\otimes 2}-\sigma_2^{\otimes 2}\right\|_1$. It is thus desirable to count with a distance measure that does not explicitly depend on the provided number of copies N. In an analogous way to the Chernoff bound (3.11), the minimum-error probability for distinguishing two quantum states, defined in Eq. (3.63), is upper-bounded by the *quantum Chernoff bound* [Audenaert *et al.*, 2007]

$$P_e(N) \leqslant \min_{s\in[0,1]} \eta_1^s\eta_2^{1-s}\,\mathrm{tr}\,\rho_1^s\rho_2^{1-s}\,, \tag{3.72}$$

[20] Recalling that a subspace **j** has dimension $2j+1$, one can check at this stage that the dimension of the total state $\rho^{\otimes 4}$ in this representation is indeed correct: $5\times 1+3\times 3+1\times 2 = 2^4 = 16$.

which is tight in the asymptotic limit $N \to \infty$[21]. Furthermore, the error probability decreases exponentially with the number N of copies as N goes to infinity [Cover and Thomas, 2006], and the rate exponent is determined by the quantum Chernoff bound. That is

$$P_e(N \to \infty) \sim e^{-ND(\rho_1,\rho_2)}\,, \tag{3.73}$$

where

$$D(\rho_1,\rho_2) = -\min_{s\in[0,1]} \log \operatorname{tr} \rho_1^s \rho_2^{1-s} \tag{3.74}$$

is known as the *quantum Chernoff distance.*

As the classical Chernoff distance, defined in Eq. (3.13), its quantum counterpart gives a proper measure of distinguishability between quantum states [Calsamiglia *et al.*, 2008]. Most importantly, although it is operationally based in a discrimination protocol (and consequently in a measurement procedure), this measure defines the optimal error rate in a device-independent way. The quantity $D(\rho_1,\rho_2)$ thus provides a nice tool for benchmarking particular strategies. In contrast to the classical case, in quantum discrimination one has to optimize the strategy, and if there are restrictions over the available measurements this can be a tedious process. A quick test to see if a particular strategy is optimal is to compare the error rate that it gives with $D(\rho_1,\rho_2)$: if both match, then optimality is guaranteed.

An additional feature of the quantum Chernoff distance is that it induces a physically motivated metric to the space of quantum states, thus endowing it with a geometrical structure [Petz, 1996]. This enables a relation between geometrical concepts (e.g., distance, volume, curvature) to physical ones (e.g., state discrimination and estimation). The metric is obtained, roughly speaking, by defining a line element between the infinitesimally close states ρ and $\rho - d\rho$ through the distinguishability measure $D(\rho,\rho-d\rho)$. In particular, the so-called Chernoff metric [Audenaert *et al.*, 2007; Calsamiglia *et al.*, 2008] provides an operationally defined volume element $d\rho^{\mathrm{Ch}}$, that for qubits with purity r reads

$$d\rho^{\mathrm{Ch}} = \frac{1}{\pi-2}\frac{\left(\sqrt{1+r}-\sqrt{1-r}\right)^2}{\sqrt{1-r^2}}dr\frac{d\Omega}{4\pi}\,, \tag{3.75}$$

where $d\Omega/4\pi$ is the invariant measure on the two-sphere. Alternatively, there exist other metrics that are based on different criteria, such as the Bures

[21] The upper bound is a direct application of the relation $\operatorname{tr}(A^sB^{1-s}) \geqslant \operatorname{tr}(A+B-|A-B|)/2$, that holds for any two positive operators A and B and for all $0 \leqslant s \leqslant 1$. A lower bound for $P_e(N)$ was found in [Nussbaum and Szkoła, 2009] that coincides with the upper bound introduced in [Audenaert *et al.*, 2007] when $N \to \infty$, thus proving attainability.

metric, induced by the fidelity distance [Życzkowski and Sommers, 2005] (see Section 4.3 for more details on different metrics for the qubits state space).

CHAPTER 4

Programmable quantum state discrimination

The standard theory of quantum state discrimination, covered in Chapter 3, is built on the premise of a measurer agent receiving both *quantum* and *classical* information, namely a quantum system in an unknown state, and a description of the possible states of the system and their *a priori* probabilities. The agent uses all this available information to devise the discrimination machine that best determines the state of the system. As a consequence, the machine is specifically oriented to that particular discrimination instance: the given hypotheses are hard-coded into its design, and the machine is mistrusted in facing any other set of hypotheses.

It is then natural to wonder whether a device for discriminating arbitrary pairs of states—a universal (multipurpose) quantum-measurement apparatus so to say—, can be constructed. Such a "quantum multimeter" can be understood at an abstract level as a programmable quantum processor [Bužek *et al.*, 2006], that is a device with a *data* port and a *program* port, where the input at the program port determines the operation to be performed on the input at the data port[1]. The usual discrimination task between known states would correspond to a processor specifically programmed by a set of instructions—the *classical* description of the possible states—to determine the state of a system in the data port, very much as programming a computer to perform a task by setting dials or switches to particular positions,

[1] Programmable quantum processors were first considered by Nielsen and Chuang as gate arrays [Nielsen and Chuang, 1997]. They restricted their study to the case in which a unitary operation, rather than a measurement or a more general completely positive linear map, is performed on the state in the data port.

each task requiring a different configuration. Programmable quantum processors admit a much more general approach, that is to consider that the programming is carried out, not by a human agent manipulating switches, but directly by raw information in a *quantum* form, i.e., information stored in the state of some quantum system. In the state discrimination context, this means that the information about the possible states of the system at the data port is not provided as a set of instructions, but instead as quantum systems in particular states entering the program port of the processor. A quantum processor programmed in this way would be able to read this quantum information by itself and adjust accordingly a discrimination measurement performed on the data system without human intervention. It could even take advantage of quantum correlated joint measurements over both the program and data systems to carry out the task more efficiently. In short, supplied with the correct programs, this machine would be capable of discriminating between any pair of quantum states.

Programmable quantum state discrimination machines have been extensively analysed in the literature. A programmable device that uses projective measurements to discriminate the state of a qubit, the basis of the projection being specified by the program, was discussed in [Fiurášek *et al.*, 2002; Fiurášek and Dušek, 2004]. In [Dušek and Bužek, 2002] the case of distinguishing two equatorial qubits with generalized measurements was considered. The separation angle between the states, which specifies the POVM, was encoded in a single-qubit program, yielding a good but suboptimal performance. Later, Bergou and collaborators proposed a different encoding system: their machine has two program ports, each of them fed with a system in one of the possible states, and a data port, fed with the state to be identified. The authors obtained the optimal solution in both the unambiguous and the minimum-error schemes for general pure qubit states [Bergou and Hillery, 2005; Bergou *et al.*, 2006a], which works by exploiting the difference between the permutation symmetry of the global state of the three ports in the two alternatives. This last approach benefits from not requiring beforehand any classical information about the hypotheses in order to prepare a specific encoding, as copies of the possible states—whatever they are—are just plugged into the program ports, perhaps coming out from some other quantum information processing device. Several other works, as well as the contents of this Chapter, extend further this idea[2].

Interestingly, these devices can also be regarded as *learning machines*:

[2] See, e.g., [Hayashi *et al.*, 2005, 2006; Bergou *et al.*, 2006b; Zhang *et al.*, 2006; He and Bergou, 2007; Ishida *et al.*, 2008; Herzog and Bergou, 2008a,b; Sedlák *et al.*, 2007, 2009; Bartůšková *et al.*, 2008; Zhou, 2011, 2014; Colin, 2012].

the device is instructed, or trained, through the program ports about different states, and, based on the acquired knowledge, it associates the state in the data port with one of the states belonging to the training set. This view implies that the discrimination task is carried out by two separate operations, an initial training step and a subsequent identification step, i.e., it considers a particular type of process happening inside the quantum processor (Chapter 5 is devoted entirely to make clear this distinction). Furthermore, programmable discrimination machines are mathematically equivalent to a change-point problem [Akimoto and Hayashi, 2011]: a source produces states of one type and, either at time t_1 or at time t_2, it starts producing states of a different type; the change-point problem consists in identifying whether the time at which the change occurs is t_1 or t_2.

In this Chapter we consider the programmable discrimination of two general qubit states, although most of our results can be generalized to higher dimensional systems. For simplicity we assume that the prior occurrence probability of each state is identical and compute the unambiguous and minimum-error rates for optimal programmable devices when an arbitrary number of copies of the states is provided at every port. We first study the performance of such devices for pure states. Some of these results are already available in the literature, but the way we formalize the problem here is crucial to treat the more general mixed state case. In addition, we obtain analytical expressions that enable us to present the results and study limiting cases in a unified way. In particular, when the program ports are loaded with an infinitely large number of copies of the states we recover the usual state discrimination problem for known states (Section 3.3.1), since, clearly, then one has as much information as the classical description of the states entering the program ports[3]. On the other hand, when the number of copies at the data port is infinitely large, while the number of copies at the program ports are kept finite, we recover the state comparison problem [Barnett *et al.*, 2003; Sedlák *et al.*, 2008].

We extend the previous pure state study to the case of mixed input states. In this scenario we only compute the minimum-error probability, as no unambiguous answers can be given if the states have the same support[4]. The performance of the device for a given purity of the input states allows to quantify how the discrimination power is degraded in the presence of noise. The expressions here are much more involved, however one can still exploit

[3] An infinite number of copies of an unknown state permits perfect quantum state tomography. As a result, the classical description of the unknown state, that is its density matrix, is obtained.

[4] See Section 3.3.2 for details. As we will see, this is indeed the case here, since the global states entering the machine are full-rank matrices.

the permutation symmetry of the input states to write the problem in a block-diagonal form, as shown in Section 3.4.1. We then obtain closed expressions for the probability of error that can be computed analytically for small number of copies and numerically evaluated for a fairly large number of copies. We are also able to obtain analytical expressions for some asymptotic rates. Again, the leading term, as in the pure state case, is seen to coincide with the average minimum error for known states.

We also analyze the fully universal discrimination machine, i.e., a device that works optimally for completely unknown input states. In this case one has to assume a uniform distribution for the purity. In contrast to the pure state distribution, there is no unique choice [Petz and Sudár, 1996], and different reasonable assumptions lead to different uniform priors. Here we consider the hard-sphere, Bures, and Chernoff priors.

4.1 Pure states

Let us start by fixing the notation and conventions that we use. We label the two program ports by A and C. These will be loaded with states $|\psi_1\rangle$ and $|\psi_2\rangle$, respectively. The data port, B, is the middle one and will be loaded with the states we wish to identify as of type 1 or type 2. We also use the short hand notation $[\psi]$ to denote $|\psi\rangle\langle\psi|$, and similarly $[\psi\phi\ldots] = [\psi]\otimes[\phi]\otimes\cdots = |\psi\rangle\langle\psi|\otimes|\phi\rangle\langle\phi|\otimes\cdots$. We may also omit the subscripts A, B and C when no confusion arises. We assume that the program ports are fed with n copies of each state and the data port with n' copies of the unknown state. This is a rather general case for which closed expressions of the error probabilities can be given. The case with arbitrary n_A, n_B, and n_C copies at each port is discussed in Appendix A.2. The expressions are more involved but the techniques are a straightforward extension of the ones presented here.

When the state at the data port is $|\psi_1\rangle^{\otimes n'}$ or $|\psi_2\rangle^{\otimes n'}$, the effective states entering the machine are given by the averages

$$\begin{aligned}
\sigma_1 &= \int d\psi_1 d\psi_2 [\psi_1^{\otimes n}]_A [\psi_1^{\otimes n'}]_B [\psi_2^{\otimes n}]_C \,, \\
\sigma_2 &= \int d\psi_1 d\psi_2 [\psi_1^{\otimes n}]_A [\psi_2^{\otimes n'}]_B [\psi_2^{\otimes n}]_C \,,
\end{aligned} \tag{4.1}$$

respectively. Note that, by taking the average over all possible input states, σ_1 and σ_2 summarize our absolute lack of knowledge about $|\psi_1\rangle$ and $|\psi_2\rangle$ in a Bayesian way, much as like it was emphasized in Section 3.1. Note also that this allows us to assess the performance of the machine in a state-independent way, in turn characterizing a machine that works for any $|\psi_1\rangle$

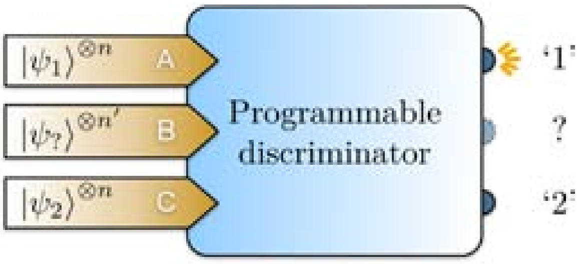

$|\psi_1\rangle^{\otimes n}$
A
$|\psi_?\rangle^{\otimes n'}$
B
$|\psi_2\rangle^{\otimes n}$
C
Programmable discriminator
'1'
?
'2'

Figure 4.1. A programmable discrimination machine with n copies of the possible states entering the program ports A and C, and n' copies of the state to be identified entering the data port B. The machine has two possible outcomes if the discrimination is done within the minimum-error approach. If the unambiguous approach is used instead, a third (inconclusive) outcome has to be considered.

and $|\psi_2\rangle$. The integrals in Eq. (4.1) can be easily computed using the Schur's lemma $\int d\phi [\phi]_X = \mathbb{1}_X/d_X$, where d_X is the dimension of the Hilbert space spanned by $\{|\phi\rangle\}$ and $\mathbb{1}_X$ is the projector onto this space. Hence

$$\begin{aligned} \sigma_1 &= \frac{1}{d_{AB}d_C}\mathbb{1}_{AB}\otimes\mathbb{1}_C\,, \\ \sigma_2 &= \frac{1}{d_A d_{BC}}\mathbb{1}_A\otimes\mathbb{1}_{BC}\,, \end{aligned} \tag{4.2}$$

where $\mathbb{1}_{XY}$ is the projector onto the completely symmetric subspace of $\mathcal{H}_X\otimes\mathcal{H}_Y$, and $d_{XY} = \operatorname{tr}\mathbb{1}_{XY}$ is its dimension. For qubits we have $d_A = d_C = n+1$ and $d_{AB} = d_{BC} = n+n'+1$.

The structure of the states (4.2) suggests the use of the angular momentum basis: $|j_A, j_B(j_{AB}), j_C; JM\rangle$ for σ_1, and $|j_A, j_B, j_C(j_{BC}); JM\rangle$ for σ_2. The quantum numbers $j_{AB} = j_A + j_B$ and $j_{BC} = j_B + j_C$ recall the way the three spins are coupled to give the total angular momentum J. Here the angular momenta have a fixed value determined by the number of copies at the ports, $j_A = j_C = n/2$ and $j_B = n'/2$, hence we can very much ease the notation by only writing explicitly the labels j_{AB} and j_{BC}. We would like to stress, however, that, in general, one needs to keep track of all the quantum numbers, specially when dealing with mixed states as in Section 4.2.

In σ_1 the first $n+n'$ spins are coupled in a symmetric way, while in σ_2 the symmetrized spins are the last $n+n'$, thus $j_{AB} = (n+n')/2 = j_{BC}$. The states are diagonal in the angular momentum bases discussed previously, and we have

$$\begin{aligned}\sigma_1 &= \frac{1}{d_{AB}d_C}\sum_{J=0,1/2}^{n'/2+n}\sum_{M=-J}^{J}[j_{AB};JM]\,,\\ \sigma_2 &= \frac{1}{d_A d_{BC}}\sum_{J=0,1/2}^{n'/2+n}\sum_{M=-J}^{J}[j_{BC};JM]\,,\end{aligned} \tag{4.3}$$

where the lower limit of the first summation takes the value 0 (1/2) for n' even (odd). Note that the spectrum of both matrices is identical and that the basis elements of their support differ only in the way the three spins are coupled. Further, the key feature of the total angular momentum bases is the orthogonality relation

$$\langle j_{AB};JM|j_{BC};J'M'\rangle = 0\,,\quad \forall J\neq J' \text{ or } M\neq M'\,. \tag{4.4}$$

Bases of this type are known as Jordan bases of subspaces [Bergou *et al.*, 2006b]. Since a state of the first basis (labeled by j_{AB}) has overlap with only one state of the second basis (labeled by j_{BC}), the problem is reduced to a series of discrimination instances between pairs of pure states. Then, the total error probability is simply the sum of the contributions of each pair.

In the unambiguous approach, the minimum probability of an inconclusive result for a pair of states $|\phi_1\rangle\,,|\phi_2\rangle$ with equal priors is simply given by Eq. (3.44) as $Q(|\phi_1\rangle\,,|\phi_2\rangle)=|\langle\phi_1|\phi_2\rangle|$, hence

$$Q=\frac{1}{d_{AB}d_C}\sum_{JM}|\langle j_{AB};JM|j_{BC};JM\rangle|\,. \tag{4.5}$$

These overlaps can be computed in terms of Wigner's 6j-symbols (see Appendix A.1):

$$\begin{aligned}&\langle j_{AB};JM|j_{BC};JM\rangle\\ &\qquad=(-1)^{j_A+j_B+j_C+J}\sqrt{(2j_{AB}+1)(2j_{BC}+1)}\begin{Bmatrix} j_A & j_B & j_{AB}\\ j_C & J & j_{BC}\end{Bmatrix}\,.\end{aligned} \tag{4.6}$$

Note that the 6j-symbols are independent of M, therefore in what follows we omit writing the quantum number M, and we perform the sum over M in Eq. (4.5) trivially by adding the multiplicative factor $2J+1$. Substituting the value of the 6j-symbols for $j_A=j_C=n/2$, $j_B=n'/2$, $j_{AB}=j_{BC}=(n+n')/2$, and setting $J=n'/2+k$, we obtain

$$\langle j_{AB};J|j_{BC};J\rangle=\binom{n}{k}\binom{n+n'}{n-k}^{-1}\,, \tag{4.7}$$

with $k = 0, 1, \ldots, n$ (observe that J takes values from $J = n + n'/2$ of the totally symmetric space down to $J = n'/2$).

Plugging the overlaps in Eq. (4.7) into Eq. (4.5), we obtain

$$Q = \sum_{k=0}^{n} \frac{n' + 2k + 1}{(n + n' + 1)(n + 1)} \frac{(n' + k)!n!}{(n' + n)!k!} = 1 - \frac{nn'}{(n + 1)(n' + 2)} , \tag{4.8}$$

where the dimension of the subspace of total angular momentum J is $n' + 2k + 1$, and in the second equality we have used the binomial sums

$$\begin{aligned} \sum_{k=0}^{n} \binom{n' + k}{n'} &= \binom{n + n' + 1}{n' + 1} , \\ \sum_{k=0}^{n} k \binom{n' + k}{n'} &= \binom{n + n' + 1}{n' + 1} \frac{n(n' + 1)}{n' + 2} . \end{aligned} \tag{4.9}$$

In the minimum-error approach no inconclusive results are allowed, but the machine is permitted to give wrong answers with some probability that one tries to minimize. This minimum-error probability can be computed along the same lines as in the previous case. Recall that the error probability $P_{\rm e}$ for two pure states $|\phi_1\rangle , |\phi_2\rangle$ and equal *a priori* probabilities is given by Eq (3.34), i.e.,

$$P_{\rm e}(|\phi_1\rangle , |\phi_2\rangle) = \frac{1}{2} \left(1 - \sqrt{1 - |\langle \phi_1 | \phi_2 \rangle|^2} \right) . \tag{4.10}$$

The total error probability is just the sum of the contribution of each pair of states with the same quantum numbers JM, $\{|j_{AB}; JM\rangle , |j_{BC}; JM\rangle\}$,

$$P_{\rm e} = \frac{1}{2} \left(1 - \sum_{k=0}^{n} \frac{n' + 2k + 1}{(n + 1)(n + n' + 1)} \sqrt{1 - \left(\frac{(n' + k)!n!}{(n' + n)!k!} \right)^2} \right) . \tag{4.11}$$

It is instructive to obtain the well-known results when the ports are loaded with just one copy of each state [Bergou and Hillery, 2005] (i.e., $n = n' = 1$). The inconclusive probability in the unambiguous approach reads

$$Q = \frac{1}{6} \sum_{J=1/2}^{3/2} (2J + 1) |\langle j_{AB} = 1; J | j_{BC} = 1; J\rangle| = \frac{5}{6} ; \tag{4.12}$$

in average, five out of six times the machine gives an inconclusive result and only 1/6 of the times it identifies the state without error. Note that the overlaps for $J = 3/2$ are one. This must be so since $J = 3/2$ corresponds to the totally symmetric subspace, which is independent of the way the spins are coupled. That is, this subspace is identical for σ_1 and σ_2. This is the main contribution to Q as it supplies $4/6 = 4/6 \times 1$ out of the total $5/6$ probability of inconclusive results. The remaining $1/6 = 2/6 \times 1/2$ is the contribution of the $J = 1/2$ subspace, where the $2/6$ is the probability of having an outcome on this subspace and $1/2$ is the overlap between the states [cf. Eq. (4.7)].

The minimum-error probability in the one copy case reads

$$P_{\mathrm{e}} = \frac{1}{2}\left(1 - \frac{1}{6}\sum_{J=1/2}^{3/2}(2J+1)\sqrt{1 - |\langle j_{AB} = 1; J|j_{BC} = 1; J\rangle|^2}\right) , \tag{4.13}$$

which, by using either Eq. (4.7) or directly Eq. (4.11), gives

$$P_{\mathrm{e}} = \frac{1}{2}\left(1 - \frac{1}{2\sqrt{3}}\right) \simeq 0.356 \; . \tag{4.14}$$

That is, approximately $1/3$ of the times the outcome of the machine will be incorrect.

The error probability in both minimum-error and unambiguous approaches will, of course, decrease when using more copies of the states at the ports of the discrimination machine. Equations (4.8) and (4.11) give the unambiguous and minimum-error probability for arbitrary values of n and n'. They enable us to study the behaviour of the machine for a large number of copies in the program and the data ports, which is what we next discuss.

4.1.1 Asymptotic limits for pure states

Let us start by considering the case of an asymptotically large number of copies at the program ports ($n \to \infty$) while keeping finite the number of copies n' at the data port. For unambiguous discrimination, from Eq. (4.8) one obtains

$$\lim_{n\to\infty} Q = \frac{2}{n'+2} \, . \tag{4.15}$$

We wish to show that, in this limit, the programmable machine has a performance that is equivalent to a protocol consisting in first estimating the states at the program ports and then performing a discrimination of *known* states over the data port. The average of the inconclusive probability of such protocol over all input states should coincide with Eq. (4.15). Recall that, for known $|\psi_1\rangle$ and $|\psi_2\rangle$, when a number n' of copies of the unknown state is given, this probability reads

$$Q(\psi_1, \psi_2) = |\langle\psi_1|\psi_2\rangle|^{n'} \; . \tag{4.16}$$

One can do an explicit calculation of the average

$$\langle Q(\psi_1,\psi_2)\rangle = \frac{1}{2}\int_0^\pi \sin\theta\cos^{n'}\frac{\theta}{2}d\theta\,, \tag{4.17}$$

but it is interesting to obtain it in a very simple way from the Schur's lemma:

$$\begin{aligned}\int d\psi_2\left(|\langle\psi_1|\psi_2\rangle|^2\right)^{\frac{n'}{2}} &= \langle\psi_1|^{\otimes\frac{n'}{2}}\left(\int d\psi_2[\psi_2]^{\otimes\frac{n'}{2}}\right)|\psi_1\rangle^{\otimes\frac{n'}{2}} \\ &= \frac{1}{d_{n'/2}} = \frac{1}{n'/2+1}\,, \end{aligned}\tag{4.18}$$

where $d_{n'/2}$ is the dimension of the symmetric space of $n'/2$ qubits (note that *sensu stricto* this procedure is only valid for n' even). Plugging this average into Eq. (4.16) one immediately recovers Eq. (4.15).

Now we turn our attention to the minimum-error probability. Taking $n\to\infty$ and using the Stirling approximation $z!\approx z^z\mathrm{e}^{-z}\sqrt{2\pi z}$ in Eq. (4.11), one obtains

$$\begin{aligned}\lim_{n\to\infty} P_\mathrm{e} &= \frac{1}{2}\left[1-2\int_0^1 dx\,x\sqrt{1-x^{2n'}}\right] \\ &= \frac{1}{2}\left[1-\frac{\sqrt{\pi}}{2}\frac{\Gamma(1+1/n')}{\Gamma(3/2+1/n')}\right]\,, \end{aligned}\tag{4.19}$$

where we have defined $x=k/n$ and used the Euler-McLaurin summation formula at leading order $\sum_{k=0}^n f(k)\simeq n\int_0^1 dx f(nx)$.

This result could be easily anticipated from the minimum-error probability with classical knowledge of the pure states. Recall that the minimum-error probability given n' identical copies is

$$P_\mathrm{e}(\psi_1,\psi_2) = \frac{1}{2}\left(1-\sqrt{1-|\langle\psi_1|\psi_2\rangle|^{2n'}}\right)\,, \tag{4.20}$$

so we just have to compute the average for all pairs of states of the above expression. Using $|\langle\psi_1|\psi_2\rangle|^2=\cos^2\theta/2$, where θ is the relative angle between the Bloch vectors of the two states, one has

$$\langle P_\mathrm{e}(\psi_1,\psi_2)\rangle = \frac{1}{2}\left[1-\frac{1}{2}\int_0^\pi d\theta\sin\theta\sqrt{1-\cos^{2n'}(\theta/2)}\right]\,, \tag{4.21}$$

and, performing the change of variables $x=\sin\theta/2$, this equation is cast exactly in the form of (4.19).

What cannot be anticipated is the next order $O(1/n)$, which gives very relevant information on how fast the protocol reaches the asymptotic value (4.19).

A lengthy, but rather straightforward, calculation yields the remarkable result that this term has a coefficient which coincides with the value of the integral $\int_0^1 dx\, x\sqrt{1-x^{2n'}}$. At this order we therefore can write

$$P_{\mathrm{e}} = \frac{1}{2} - \frac{\sqrt{\pi}}{4}\frac{\Gamma(1+1/n')}{\Gamma(3/2+1/n')}\left(1-\frac{1}{n}\right) . \tag{4.22}$$

We now analyze the complementary case, that is, when the number of copies at the data port is infinitely large ($n' \to \infty$) while the number n of copies at the program ports is kept finite. In this limit we have perfect knowledge of the data state $|\psi\rangle$, but we do not know to which program port it should be associated. Observe that this situation is very much the same as state comparison [Barnett *et al.*, 2003].

In this scenario, the inconclusive probability in the unambiguous approach reads from Eq. (4.8) as

$$\lim_{n'\to\infty} Q = \frac{1}{n+1} . \tag{4.23}$$

Let us see that this agrees with the average performance of a standard state comparison protocol. If the data state is the same as the program state in the upper or lower port, the effective states to be discriminated are

$$\begin{aligned} \sigma_1 &= \frac{1}{d_n}[\psi^{\otimes n}] \otimes \mathbb{1}_n , \\ \sigma_2 &= \frac{1}{d_n}\mathbb{1}_n \otimes [\psi^{\otimes n}] , \end{aligned} \tag{4.24}$$

respectively, where $d_n = n+1$ is the dimension of the symmetric space of n-qubits and $\mathbb{1}_n$ is the projector onto this subspace. The minimal inconclusive probability for these two states can be obtained with a POVM with elements $E_1 = [\psi^{\otimes n}] \otimes [\psi^{\otimes n}]^{\perp}, E_2 = [\psi^{\otimes n}]^{\perp} \otimes [\psi^{\otimes n}]$, both representing conclusive answers, and $E_0 = \mathbb{1} \otimes \mathbb{1} - E_1 - E_2$, which represents the inconclusive one. In these expressions $[\psi^{\otimes n}]^{\perp} = \mathbb{1}_n - [\psi^{\otimes n}]$. Note that this POVM checks whether the state in each register is $|\psi\rangle$ or not. The probability of obtaining the inconclusive answer reads

$$Q(\psi) = \frac{1}{2}\left(\mathrm{tr}\, E_0\sigma_1 + \mathrm{tr}\, E_0\sigma_2\right) = \frac{1}{n+1} \tag{4.25}$$

independently of the state $|\psi\rangle$.

The minimum-error probability in this limit can be tackled in a similar fashion. The asymptotic expression of Eq. (4.11), though not as direct as in the unambiguous case, is rather straightforward to obtain. Note that the dominant factor in the term containing factorials inside the square root is

$n'^{\,-2(n-k)}$. Hence we can effectively replace the square root term by 1, for all $k < n$. Taking into account that for $k = n$ the square root vanishes, we have

$$\lim_{n'\to\infty} P_{\rm e} = \frac{1}{2}\left(1 - \frac{n}{n+1}\right) = \frac{1}{2(n+1)}\,. \tag{4.26}$$

The minimum-error probability of a strategy that first estimates perfectly the data states and then tries to associate the correct label to them is given by the Helstrom formula (3.33) for σ_1 and σ_2, that is

$$P_{\rm e} = \frac{1}{2}\left(1 - \frac{1}{2}\left\|\sigma_1 - \sigma_2\right\|_1\right)\,. \tag{4.27}$$

Substituting the expression of the states (4.24) we obtain

$$\begin{aligned} P_{\rm e} &= \frac{1}{2}\left(1 - \frac{1}{2(n+1)}\left\|[\psi^{\otimes n}]\otimes[\psi^{\otimes n}]^{\perp} - [\psi^{\otimes n}]^{\perp}\otimes[\psi^{\otimes n}]\right\|_1\right) \\ &= \frac{1}{2}\left(1 - \frac{2}{2(n+1)}\left\|[\psi^{\otimes n}]\otimes[\psi^{\otimes n}]^{\perp}\right\|_1\right) \\ &= \frac{1}{2}\left(1 - \frac{n}{n+1}\right) = \frac{1}{2(n+1)}\,, \end{aligned} \tag{4.28}$$

where in the first equality we have subtracted the common term $[\psi^{\otimes n}]\otimes[\psi^{\otimes n}]$ from both states, in the second we have used the orthogonality of the operators and in the last equality we have taken into account that $\mathrm{tr}\,[\psi^{\otimes n}]^{\perp} = \mathrm{tr}\,(\mathbb{1}_n - [\psi^{\otimes n}]) = n$ (i.e., one unit less than the dimension of the corresponding symmetric space). As expected, the result is again independent of $|\psi\rangle$.

To end this section we compute the asymptotic error probabilities for the symmetric case, that is, when all the ports are loaded with the same $n' = n$ (and large) number of copies.

In the unambiguous approach, when $n = n' \to \infty$ the first nonvanishing order of (4.8) reads

$$Q = \frac{3}{n} + \ldots \tag{4.29}$$

To compute the minimum-error probability, it is convenient to write Eq. (4.11) for $n = n'$ as

$$P_{\rm e} = \frac{1}{2}\sum_{k=0}^{n} p_k\left(1 - \sqrt{1 - c_k^2}\right)\,, \tag{4.30}$$

where

$$p_k = \frac{n+1+2k}{(2n+1)(n+1)}\,, \tag{4.31}$$

and

$$c_k = \binom{n+k}{n}\binom{2n}{n}^{-1} . \tag{4.32}$$

We first observe that c_k is a monotonically increasing function and hence it takes its maximum value at $k = n$. Second, we note that around this point

$$\begin{aligned} \binom{n+k}{n} &\simeq 2^{(n+k)H(\frac{n}{n+k})} \\ &\simeq 2^{(n+k)H(1/2)} = 2^{n+k} , \end{aligned} \tag{4.33}$$

where $H(x) = -x \ln x - (1-x)\ln(1-x)$ is the Shannon entropy of a binary random variable, and we have used that $k \approx n$ and $H(1/2) = 1$. Similarly, one has

$$\binom{2n}{n} \simeq 2^{2nH(1/2)} = 2^{2n} , \tag{4.34}$$

and hence $c_k \simeq 2^{-(n-k)}$. With this, the probability of error in this limit reads

$$P_{\rm e} = \frac{1}{2}\sum_{k=0}^{\infty} p_k \left(1 - \sqrt{1 - \left(\frac{1}{4}\right)^{n-k}}\right) . \tag{4.35}$$

Finally, we perform the change of variables $k \to n-k$ and use that in Eq. (4.31) $p_{n-k} \simeq 3/(2n)$ for $k \simeq 0$ to obtain

$$P_{\rm e} = \frac{3}{4n}\zeta(1/4) \approx \frac{0.882}{n} , \tag{4.36}$$

where we have defined the function

$$\zeta(x) = \sum_{k=0}^{\infty}\left(1 - \sqrt{1 - x^k}\right) , \tag{4.37}$$

which converges very quickly to its exact value (the first four terms already give a value that differ in less than 10^{-3} from the exact value).

4.2 Mixed states

We now move to the case when the program and data ports are loaded with mixed states. This situation arises for instance when there are imperfections in the preparation or noise in the transmission of the states. It is reasonable to suppose that these imperfections have the same effect on all states (i.e.

to consider that the states have all the same purity r). The input states are then tensor products of

$$\rho_i = \frac{\mathbb{1} + r\,\boldsymbol{n}_i\,\boldsymbol{\sigma}}{2}\,, \tag{4.38}$$

where $\boldsymbol{n}_i$ is a unitary vector and $\boldsymbol{\sigma} = (\sigma_x, \sigma_y, \sigma_z)$ are the usual Pauli matrices. In what follows we assume that only the purity is known, i.e., one knows the characteristics of the noise affecting the states, but nothing else. This means that the averages will be performed over the isotropic Haar measure of the $\mathbb{S}^2$ sphere, in the same manner as for pure states. At the end of this section we also analyze the performance of a fully universal discrimination machine, that is, when not even the purity is considered to be known.

Note that mixed states can only be unambiguously discriminated if they have different supports (see Section 3.3.2), which is not the case when the ports are loaded with copies of the states (4.38) as they are full-rank matrices. Therefore, only the minimum-error discrimination approach will be analyzed here. It is worth stressing that the computation of the optimal error probability in the multicopy case is highly nontrivial, even for known qubit mixed states. Only recently have feasible methods for computing the minimum-error probability for a rather large number of copies been developed [Calsamiglia *et al.*, 2010], and the asymptotic expression of such probability obtained[5]. The main difficulty can be traced back to the computation of the trace norm [see Eq.(4.27)] of large matrices. The dimension of the matrices grows exponentially with the total number of copies entering the machine, and for a relative small number of them the problem becomes unmanageable. However, as it will be clear, it is possible to exploit the permutation symmetry of the input states to write them in the block-diagonal form given in Eq. (3.65), crucially reducing the complexity of the problem.

The two effective states we have to discriminate are

$$\begin{aligned}
\sigma_1 &= \int dn_1 dn_2 \rho_{1\,A}^{\otimes n} \otimes \rho_{1\,B}^{\otimes n'} \otimes \rho_{2\,C}^{\otimes n}\,, \\
\sigma_2 &= \int dn_1 dn_2 \rho_{1\,A}^{\otimes n} \otimes \rho_{2\,B}^{\otimes n'} \otimes \rho_{2\,C}^{\otimes n}\,,
\end{aligned} \tag{4.39}$$

where $dn_i = d\Omega_i/(4\pi)$ is the invariant measure on the two-sphere. Recall that, as discussed in Section 3.4.1, any state having permutation invariance (e.g., $\rho^{\otimes n}$) can be written in a block-diagonal form using the irreducible representations of the symmetric group S_n. Each block is specified by the total angular momentum j and a label α that distinguishes the different

[5] This is achieved via attainability of the quantum Chernoff bound. See Section 3.4.2 for details.

equivalent representations for a given j

$$\rho^{\otimes n} = \bigoplus_{j,\alpha} \rho_j^{(\alpha)} \,. \tag{4.40}$$

The angular momentum takes values $j = n/2, n/2-1, \ldots, 1/2\,(0)$ for odd (even) n, and the number of equivalent representations for each j is [cf. Eq. (3.70)]

$$\nu_j^n = \binom{n}{n/2-j} \frac{2j+1}{n/2+j+1} \,, \tag{4.41}$$

that is $\alpha = 1, \ldots, \nu_j^n$. For each block we have

$$\operatorname{tr} \rho_j^{(\alpha)} = \left(\frac{1-r^2}{4}\right)^{n/2-j} \sum_{k=-j}^{j} \left(\frac{1-r}{2}\right)^{j-k} \left(\frac{1+r}{2}\right)^{j+k} \equiv (2j+1) C_j^n \,, \tag{4.42}$$

which, of course, is the same for all equivalent irreducible representations (i.e., independent on the label α). The origin of the factors appearing in Eq. (4.42) was outlined in deducing Eq. (3.67), but let us briefly remember it here[6]. The first factor comes from the contribution from the $n/2-j$ singlets present in a representation j made up of n spin-1/2 states. The summation term is the trace of the projection of the remaining states in the symmetric subspace with total angular momentum j, where we can use the rotational invariance of the trace to write each state in the form $\operatorname{diag}\left(\frac{1+r}{2}, \frac{1-r}{2}\right)$. This term simply reads

$$t_j = \sum_{k=-j}^{j} \left(\frac{1-r}{2}\right)^{j-k} \left(\frac{1+r}{2}\right)^{j+k} = \frac{1}{r}\left[\left(\frac{1+r}{2}\right)^{2j+1} - \left(\frac{1-r}{2}\right)^{2j+1}\right] \,, \tag{4.43}$$

and hence

$$C_j^n = \frac{1}{2j+1} \left(\frac{1-r^2}{4}\right)^{n/2-j} t_j \,. \tag{4.44}$$

Very much in the same way as it happened in previous sections, the only difference between the diagonal basis of σ_1 and σ_2 is the ordering of the angular momenta couplings. In σ_1 we first couple subspaces A and B and obtain

$$\rho_{AB} = \int dn_1 \rho_{1\,A}^{\otimes n} \otimes \rho_{1\,B}^{\otimes n'} = \sum_{\xi_{AB}} C_{j_{AB}}^{n+n'} \mathbb{1}_{\xi_{AB}} \,, \tag{4.45}$$

where

$$\mathbb{1}_{\xi_{AB}} = \sum_{M_{AB}} |\xi_{AB} M_{AB}\rangle\langle \xi_{AB} M_{AB}| \tag{4.46}$$

[6] Also, full details can be found in [Bagan *et al.*, 2006].

is the projector onto the subspace with associated quantum numbers $\xi_{AB} = \{j_A, \alpha_A, j_B, \alpha_B, j_{AB}\}$, and $C^{n+n'}_{j_{AB}}$ is defined in Eq. (4.42). Note that $C^{n+n'}_{j_{AB}}$ depends only on the purity of the state and on the total angular momentum j_{AB}. Note also that the tensor product of a mixed state has projections in all subspaces and the blocks are not uniquely determined by the value of j_{AB}, i.e., one has to keep track of the labels j_A and j_B as well. Of course, subspaces with different quantum numbers ξ_{AB} are orthogonal, i.e., $\mathrm{tr}\,[\mathbb{1}_\xi \mathbb{1}_{\xi'}] = \delta_{\xi\xi'} \mathrm{tr}\,\mathbb{1}_\xi$. When coupling the third system one plainly adds the quantum numbers $\xi_C = \{j_C, \alpha_C\}$.

The diagonal bases of σ_1 and σ_2 are written as $\mathcal{B}_1 = \{|\xi_{AB}\xi_C; JM\rangle\}$ and $\mathcal{B}_2 = \{|\xi_A\xi_{BC}; JM\rangle\}$, respectively. Obviously, each set contains $2^{2n+n'}$ orthonormal states and Eq. (4.39) reads

$$\begin{aligned} \sigma_1 &= \sum_{\xi_{AB}\xi_C} \sum_{JM} C^{n+n'}_{j_{AB}} C^n_{j_C} [\xi_{AB}\xi_C; JM]\,, \\ \sigma_2 &= \sum_{\xi_A\xi_{BC}} \sum_{JM} C^n_{j_A} C^{n+n'}_{j_{BC}} [\xi_A\xi_{BC}; JM]\,. \end{aligned} \tag{4.47}$$

We just have to compute the minimum-error probability from the Helstrom formula (4.27) for these two states. It is convenient to define the trace norm term

$$T = \|\sigma_1 - \sigma_2\|_1\,, \tag{4.48}$$

so that

$$P_\mathrm{e} = \frac{1}{2}\left(1 - \frac{T}{2}\right)\,. \tag{4.49}$$

To compute T we need to know the unitary matrix Λ that transforms $\mathcal{B}_2$ into $\mathcal{B}_1$ or vice versa. The elements of this unitary are given by the overlaps between the elements of both basis $\langle\xi_{AB}\xi_C; JM|\xi'_A\xi'_{BC}; J'M'\rangle$. We observe that these overlaps are nonvanishing only if $j_X = j'_X$, $\alpha_X = \alpha'_X$ $(X = A, B, C)$ and $J = J', M = M'$. Furthermore, as mentioned previously, their value does not depend on M or α_X, thus sums over these quantum numbers simply amount to introduce the corresponding multiplicative factors. Therefore, it is useful to introduce a label containing the quantum numbers that determine the orthogonal blocks in $\mathcal{B}_1$ and $\mathcal{B}_2$ that may have non vanishing overlaps, $\xi = \{j_A, j_B, j_C, J\}$, and the corresponding multiplicative factor

$$\gamma_\xi = \nu^n_{j_A} \nu^{n'}_{j_B} \nu^n_{j_C} (2J+1)\,, \tag{4.50}$$

where ν^n_j is given in Eq. (4.41). Eq. (4.48) then reads

$$T = \sum_\xi \gamma_\xi T^\xi = \sum_\xi \gamma_\xi \left\| \sigma_1^{(\xi)} - \Lambda^{(\xi)} \sigma_2^{(\xi)} {\Lambda^{(\xi)}}^T \right\|_1\,, \tag{4.51}$$

where the explicit expressions of the matrix elements are

$$\begin{aligned}\left[\sigma_1^{(\xi)}\right]_{j_{AB}j'_{AB}} &= \delta_{j_{AB}j'_{AB}} C^{n+n'}_{j_{AB}} C^{n}_{j_C}\,, \\ \left[\sigma_2^{(\xi)}\right]_{j_{BC}j'_{BC}} &= \delta_{j_{BC}j'_{BC}} C^{n}_{j_A} C^{n+n'}_{j_{BC}}\,,\end{aligned} \tag{4.52}$$

and

$$\Lambda^{(\xi)}_{j_{AB},j_{BC}} = \langle \xi, j_{AB} | \xi, j_{BC} \rangle\,. \tag{4.53}$$

Recall that the overlap (4.53) is independent of the quantum number labelling the equivalent representations (recall also that it is independent of M), and therefore is given by Eq. (4.6).

The computation of the minimum-error probability reduces to a sum of trace norms of small-size Helstrom matrices that have dimensions of the allowed values of j_{AB} and j_{BC} for given $\xi = \{j_A, j_B, j_C, J\}$. Hence

$$P_{\rm e} = \frac{1}{2}\left(1 - \frac{1}{2}\sum_{\xi} \gamma_\xi T^\xi\right)\,, \tag{4.54}$$

and this computation can be done very efficiently.

We would like to show the analytical results for the simplest case of having just one state at each port, i.e., when $n = n' = 1$. In this situation we have fixed values $j_A = j_B = j_C = 1/2$, so the total angular momentum can be $J = 3/2, 1/2$, and $j_{AB} = 1, 0$ (and similarly for j_{BC}). Here there is no degeneracy, the number of equivalent representations defined in Eq. (4.41) is 1, and, therefore, the multiplicative factor (4.50) simply reads $\gamma_\xi = 2J + 1$. The only relevant quantum number in this case is $\xi = J$, as all the others are fixed, and we do not need to write them explicitly. The minimum-error probability is then

$$P_{\rm e} = \frac{1}{2}\left[1 - \frac{1}{2}\sum_{J=1/2}^{3/2} (2J+1) \left\| \sigma_1^{(J)} - \Lambda^{(J)} \sigma_2^{(J)} \Lambda^{(J)T} \right\|_1 \right]\,. \tag{4.55}$$

The term of the sum corresponding to $J = 3/2$ vanishes since it corresponds to the projection of $\sigma_{1,2}$ onto the completely symmetric subspace, which is identical for both states. Indeed, in this subspace $\sigma_1^{(3/2)} = \sigma_2^{(3/2)} = C^2_1 C^1_{1/2} = (3 + r^2)/24$, where we have used Eq. (4.44), and from Eq. (4.53) we obtain $\Lambda^{(3/2)} = 1$. In the subspace $J = 1/2$ we have

$$\sigma_1^{(1/2)} = \sigma_2^{(1/2)} = \begin{pmatrix} C^2_1 C^1_{1/2} & 0 \\ 0 & C^2_0 C^1_{1/2} \end{pmatrix} = \begin{pmatrix} \frac{1}{24}(3 + r^2) & 0 \\ 0 & \frac{1}{8}(1 - r^2) \end{pmatrix}\,, \tag{4.56}$$

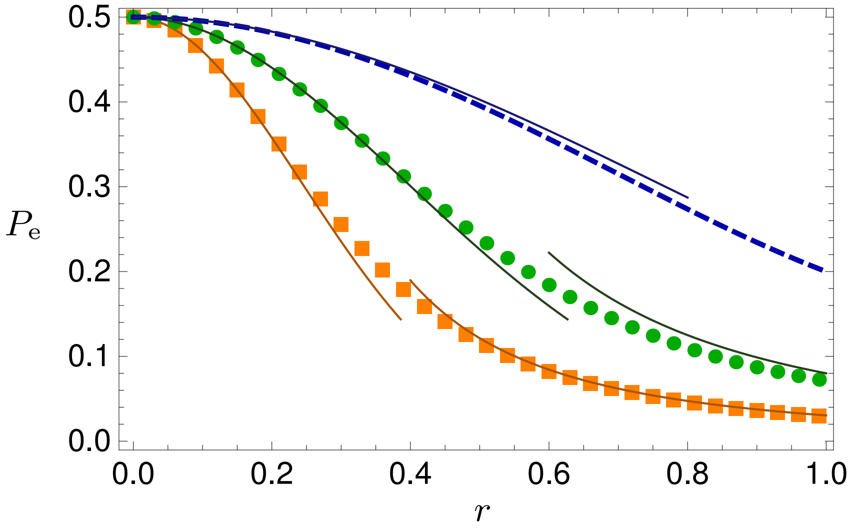


Figure 4.2. Error probability P_e for $n = n' = 3$ (blue dashed line), 11 (green circles) and 29 (yellow squares) versus purity. The fit $P_\mathrm{e} \simeq 0.882/(nr^2)$ in the regime of high purities for $n = 11$ and $n = 29$ and the Gaussian approximation $P_\mathrm{e} \simeq 1/2 \exp[-nr^2/(2\sqrt{3})]$ in the regime of low purities for all cases is represented (solid lines).

and

$$\Lambda^{(1/2)} = \begin{pmatrix} \frac{1}{2} & \frac{\sqrt{3}}{2} \\ \frac{\sqrt{3}}{2} & -\frac{1}{2} \end{pmatrix} . \tag{4.57}$$

Plugging these expressions into Eq. (4.55) we obtain the minimum-error probability for the one-copy state case

$$P_\mathrm{e} = \frac{1}{2}\left(1 - \frac{r^2}{2\sqrt{3}}\right) . \tag{4.58}$$

As expected, when $r \to 1$ we recover the pure state value (4.14).

Numerical results of the minimum-error probability as a function of the purity of the input states for the symmetric case $n = n'$ are depicted in Fig. 4.2. One sees that, for low values of n ($n \lesssim 3$), the dependence on the purity is not very marked: the curves are concave almost in the whole range of the purity. For larger n, however, there is an interval of purities where the behavior changes quite significantly. For instance, for $n = 29$, the inflection point occurs at $r \approx 0.3$. At very large values of n one expects a step-like shape with an inflection point approaching $r = 0$ because the probability of error remains very small for $r \neq 0$ and is strictly $1/2$ at $r = 0$. The shape of the curves is explained by the existence of two distinct regimes. For high purities

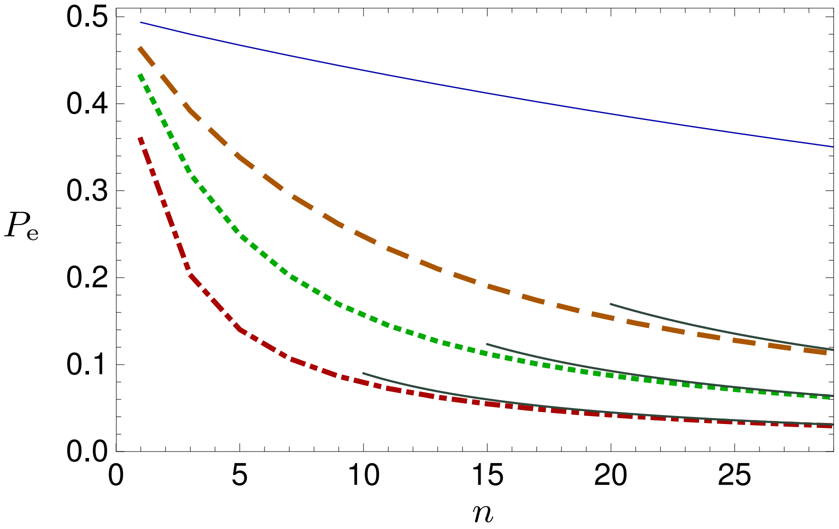


Figure 4.3. Error probability $P_{\rm e}$ for $r = 0.2$ (blue thin solid line), $r = 0.5$ (brown dashed line), $r = 0.7$ (green dotted line), and $r = 1$ (red dot-dashed line) versus n ($n = n'$ is assumed). Numerical points have been joined for an easier visualization. The approximation $0.882/(nr^2)$ is represented (thin solid lines).

the probability of error is well fitted by a linear function in the inverse of the number of copies. We get $P_{\rm e} \simeq 0.88/(nr^2)$, where the value 0.88 coincides with the analytical value computed for pure states in Eq. (4.36). Of course, this approximation cannot be valid for low purities. In the low-purity regime, the minimum-error probability is very well approximated by the Gaussian function $P_{\rm e} \simeq 1/2 \exp[-nr^2/(2\sqrt{3})]$, where we have taken the argument of the exponential from the exponentiation of the error probability for the exact $1 \times 1 \times 1$ case, given in Eq. (4.58). This approximation works for purities in the interval of the width of the Gaussian, i.e., up to $\sim 1/\sqrt{n}$. Therefore, as n increases the asymptotic approximation $P_{\rm e} \propto 1/(nr^2)$ extends its validity to almost the whole range of purities, and the expected jump discontinuity develops in $r = 0$ as $n \to \infty$. Similar information is depicted in Fig. 4.3, where the error probability is plotted as function of the number of copies n for different purities. We have superimposed the asymptotic result, which is seen to yield a very good approximation to the exact error probability already for $n \gtrsim 20$.

4.2.1 Asymptotic $n \times 1 \times n$

As in previous sections, it is interesting to study the performance of the machine in the asymptotic regimes. A particularly important instance where it is possible to obtain closed expressions is the case when the number of copies at the program ports is asymptotically large and there is one state at the data port. We show how to compute the leading order and sketch the generalizations needed to obtain the subleading term.

Observe first that j_{AB} can only take the values $j_{AB} = j_A \pm 1/2$, and similarly for j_{BC}. Therefore $\sigma^{(\xi)}_{1,2}$ are 2×2 matrices (except in the extremal case of $J = j_A + j_C + 1/2$, in which are one-dimensional). It is useful to write

$$\sigma(j) = C^n_{j_A} C^n_{j_C} \begin{pmatrix} R_+(j) & 0 \\ 0 & R_-(j) \end{pmatrix} , \tag{4.59}$$

with

$$R_\pm(j) = \frac{C^{n+1}_{j\pm 1/2}}{C^n_j} . \tag{4.60}$$

With this definition one simply has [see Eq.(4.52)]

$$\sigma^{(\xi)}_1 = \sigma(j_A) \qquad \text{and} \qquad \sigma^{(\xi)}_2 = \sigma(j_C). \tag{4.61}$$

We further note that for large n

$$\nu^n_j C^n_j \approx \frac{1}{n/2 + j + 1} \frac{1+r}{2r} \sqrt{\frac{2}{n\pi(1-r^2)}} \exp\left[-n \frac{(2j/n - r)^2}{2(1-r^2)} \right] . \tag{4.62}$$

Defining $y = 2j/n$ and using the Euler-Maclaurin summation formula, we have for a generic function $f(j)$

$$\sum_j \nu^n_j C^n_j f(j) \approx \frac{1+r}{2r} \int_{-\infty}^{\infty} \frac{dy\, G_n(y)}{n/2 + ny/2 + 1} f(\frac{ny}{2}) , \tag{4.63}$$

where we have extended limits of integration from (0,1) to $(-\infty, \infty)$, which is legitimate for large n, and defined

$$G_n(y) = \sqrt{\frac{n}{2\pi(1-r^2)}} \exp\left[-n \frac{(y-r)^2}{2(1-r^2)} \right] , \tag{4.64}$$

i.e., a Gaussian distribution centered at $y = r$ with variance $\sigma^2 = (1-r^2)/n$. Note that, at leading order in the limit $n \to \infty$, $G_\infty \approx \delta(y - r)$, and hence

$$\sum_j \nu^n_j C^n_j f(j) \approx \frac{1}{nr} f(\frac{nr}{2}) . \tag{4.65}$$

Note also that at this order

$$R_\pm(j) \approx R_\pm(\frac{nr}{2}) = \frac{1 \pm r}{2} . \tag{4.66}$$

There only remains to compute the unitary matrix Eq. (4.53). Observe that the total angular momentum takes values $J = |j_A - j_C| + 1/2 + k$, with $k = 0, 1, \ldots, 2\min\{j_A, j_C\}$. The leading order is rather easy to write (the subleading term, although straightforward, is far more involved and we will not show it here). At this order we have $J = 1/2 + k$ and $k = 0, 1, \ldots, nr$, and the matrix elements computed from Eq. (4.6) yield

$$\Lambda^{(\xi)} = \frac{1}{nr}\begin{pmatrix} k & \sqrt{(nr)^2 - k^2} \\ \sqrt{(nr)^2 - k^2} & -k \end{pmatrix} . \tag{4.67}$$

Plugging Eqs. (4.59-4.67) into Eq. (4.51) one gets

$$T \simeq \sum_{k=0}^{nr} 2k \frac{2}{n^3 r^2} \sqrt{(nr)^2 - k^2}\,, \tag{4.68}$$

where the sum over j_A and j_C has been trivially performed by substituting their central value $nr/2$ in the summand, and the only remaining multiplicative of γ_ξ [cf. Eq. (4.50)] is $2J + 1 \simeq 2k$. Finally, defining $x \equiv k/nr$ and using the Euler Mac-Laurin approximation as in Eq. (4.19) we obtain

$$T \simeq 4r \int_0^1 dx\, x\sqrt{1 - x^2} = \frac{4r}{3}\,, \tag{4.69}$$

and hence

$$P_{\rm e} \simeq \frac{1}{2} - \frac{r}{3}\,, \tag{4.70}$$

which obviously coincides with the pure state result Eq. (4.19) for $n' = 1$ and $r \to 1$.

As for the computation of the next-to-leading order, the integrals approximating the sums over j_A and j_C have to incorporate the fluctuations around the central value, that is, one defines $j_A = \frac{n}{2}(r + \eta_A)$ and $j_C = \frac{n}{2}(r + \eta_C)$, where the variables η_X have effective dimension $n^{-1/2}$. Then one can expand the matrix elements of $\sigma_{1,2}$, Λ, and the terms of ν_j^n present in Eq. (4.63), taking into account the effective dimensionality of all the terms [notice that $k \to n(r+\eta)x$, where the integration range of x is $(0, 1)$]. One then performs the sum in k by means of the Euler-Maclaurin summation formula as before. Finally one computes the integration in $j_{A/B}$ taking into account that range of the variables $\eta_{A/B}$ can be taken to be $(-\infty, \infty)$. After a somewhat lengthy calculation we obtain

$$P_{\rm e} \simeq \frac{1}{2} - \frac{r}{3} + \frac{1}{3nr}\,. \tag{4.71}$$

Note that the limit $r = 0$ is singular and not surprisingly the expansion breaks down for purities of order $1/n$. As it should, the error probability (4.71) increases monotonically with the purity.

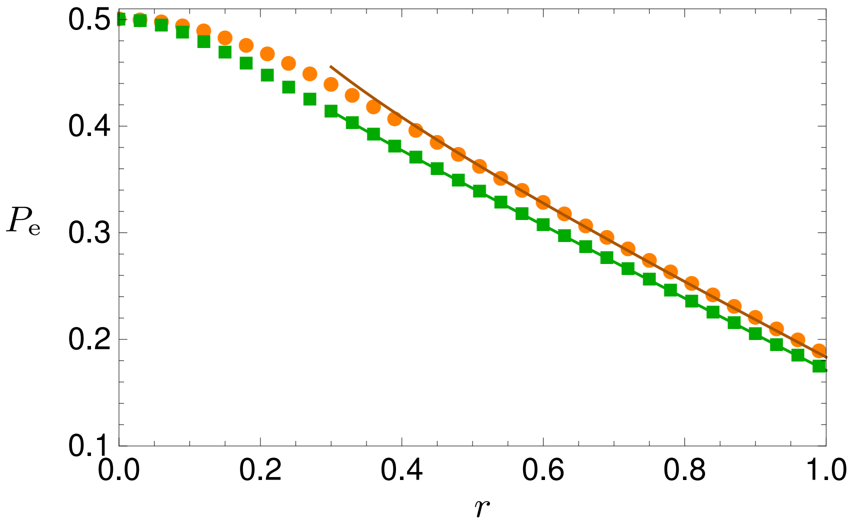


Figure 4.4. Error probability P_e for $n = 20$ (yellow circles) and $n = 79$ (green squares) versus purity. The asymptotic behaviour given by Eq. (4.71) is represented for both cases.

In Fig. 4.4 we plot the error probability as a function of the purity for $n = 20$ and $n = 79$. One sees that the asymptotic expression (4.71) approximates very well the minimum-error probability even for a small number of copies. For larger n (e.g., for $n = 79$) the approximation works extremely well down to values below $r = 0.3$.

We finish this section by showing that the leading term (4.70) coincides with the average error of a device that first estimates the mixed states at the program ports and afterwards does the usual minimum-error discrimination of the data state. From the Helstrom formula (4.27) particularized for mixed qubit states one has

$$P_e = \left\langle \frac{1}{2}\left(1 - \frac{1}{2}|\boldsymbol{r}_1 - \boldsymbol{r}_2|\right)\right\rangle , \tag{4.72}$$

where the average is taken over all possible orientations of the Bloch vectors $\boldsymbol{r_1}$ and $\boldsymbol{r}_2$. For equal purity states it simply reads

$$P_e = \frac{1}{2}\left(1 - \frac{r}{2}\int_0^\pi d\theta \sin\theta \sin\theta/2\right) = \frac{1}{2} - \frac{r}{3} . \tag{4.73}$$

4.3 Universal discrimination

Let us finally address the fully universal discrimination machine, that is a machine that distinguishes states from which nothing is assumed to be known,

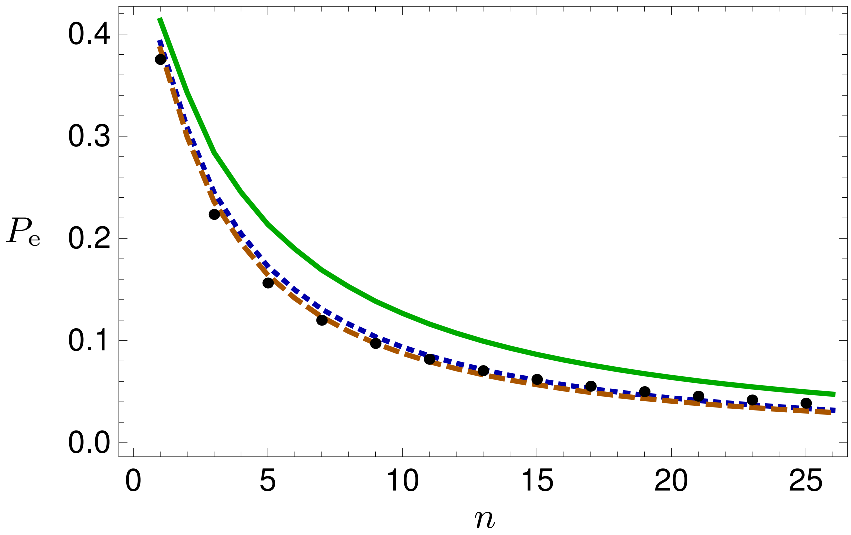


Figure 4.5. Error probability P_{e} for hard-sphere (green solid line), Bures (blue dotted line) and Chernoff (red dashed line) priors versus n ($n = n'$ is assumed). The points correspond to the error probability for a fixed $r = 0.9$; its proximity to the Chernoff curve exposes the fact that this prior gives larger weights to states of high purity.

not even its purity. For this type of machine, we need to specify a prior distribution for the purity. While the isotropy of the angular variables yields a unique uniform distribution for the angular variables, the Haar measure on the two-sphere used in previous sections, the corresponding expression for a fully unbiased distribution of the purity $w(r)$ is not uniquely determined. This is a longstanding issue, and several priors haven been suggested depending on the assumptions made [Petz and Sudár, 1996; Bengtsson and Zyczkowski, 2006]. Here we will not stick to a particular distribution, rather we will show results for three reasonable distributions. The actual values of the probability of error may depend on the chosen prior, but the overall performance is seen to be very similar.

The most straightforward, but perhaps not very well grounded, choice is that of the distribution of a hard-sphere $w(r) \propto r^2$, that is, a normalized integration measure given by

$$d\rho^{\mathrm{HS}} = 3r^2 dr \frac{d\Omega}{4\pi} \,. \tag{4.74}$$

The Bures distribution is far better motivated. It corresponds to the volume element induced by the fidelity distance [Życzkowski and Sommers, 2005]. It is monotonically decreasing under coarse graining [Petz and Sudár, 1996] and it has been argued that it corresponds to maximal randomness of the signal states [Hall, 1998]. In this case one has $w(r) \propto r^2/\sqrt{1-r^2}$. Note that this distribution assigns larger weights to pure states, as their distinguishability in terms of the fidelity is larger than that of mixed states. The integration measure reads

$$d\rho^{\mathrm{Bu}} = \frac{4}{\pi}\frac{r^2}{\sqrt{1-r^2}}dr\frac{d\Omega}{4\pi}\,. \tag{4.75}$$

Lastly, we also consider the Chernoff distribution [Audenaert *et al.*, 2007; Calsamiglia *et al.*, 2008]. It is the prior induced by the Chernoff distance, which has a clear operational meaning in terms of the distinguishability between states (see Section 3.4.2). By construction it is monotonically decreasing under coarse graining. This measure assigns even larger weights to states of high purity and lower to the very mixed ones. This assignment is, again, based on distinguishability properties, but in terms of the asymptotic behavior of the error probability. The measure can be written as [Audenaert *et al.*, 2007; Calsamiglia *et al.*, 2008]

$$d\rho^{\mathrm{Ch}} = \frac{1}{\pi-2}\frac{\left(\sqrt{1+r}-\sqrt{1-r}\right)^2}{\sqrt{1-r^2}}dr\frac{d\Omega}{4\pi}\,. \tag{4.76}$$

The effective states we have to discriminate are

$$\Sigma_k = \int d\rho_1 d\rho_2 \rho_{1\,A}^{\otimes n}\otimes\rho_{k\,B}^{\otimes n'}\otimes\rho_{2\,C}^{\otimes n}\,, \qquad k=1,2\,, \tag{4.77}$$

where $d\rho_k$ takes the expressions of the measures (4.74) through (4.76). Note that the block structure of the states is preserved, as it only depends on the permutation invariance of the input states, which remains untouched. Further, we can use rotational invariance in the same fashion as in Eqs. (4.45) and (4.47). Therefore, here it is only required to compute the average of the coefficients C_j^n in Eq. (4.42) according to priors (4.74) through (4.76). To calculate the minimum-error probability of this fully universal machine one simply uses Eq. (4.54) for the states (4.47) with the averaged coefficients $\langle C_j^n\rangle$ computed in Appendix A.3.

In Fig. 4.5 we present the minimum-error probability of the fully universal machine for the three priors discussed for an equal number of program and data states up to $n = n' = 26$. As anticipated, the smaller average error corresponds to the Chernoff distance, because states with higher purity are assigned a larger weight, and these are easier to discriminate. The probability of error, as somehow expected, is inversely proportional to the number of copies, and attains very similar values than for the discrimination of states with fixed known purity of the order of $r \sim 0.9$.

4.4 Programmable discrimination with an error margin

In this Section we analyze the paradigm of quantum state discrimination with an error margin, presented in Section 3.3.3 for known states, in the context of programmable discrimination machines for pure qubit states, when n copies of the program states and n' copies of the data state are provided. By doing so we connect the results for unambiguous and minimum-error discrimination derived in Section 4.1. We will show that, by relaxing the zero-error condition slightly, the resulting scheme provides an important enhancement in performance over the widely used unambiguous scheme for programmable machines. We discuss the two ways of imposing an error margin to the error probability, i.e., via a *weak* condition and a *strong* condition.

Although so far not much attention has been paid to the POVM that represents the machine, for this Section it is convenient to explicitly refer to it. A programmable discriminator is generically defined by a POVM with three elements $\mathcal{E} = \{E_1, E_2, E_0\}$. Recall that, as a consequence of the orthogonality relation of the Jordan bases (4.4), the averaged global states σ_1 and σ_2 have a block-diagonal structure in the angular momentum basis, each block corresponding to a Jordan subspace with an associated total angular momentum J. Hence the total Hilbert space of the states is of the form $\mathcal{H} = \bigoplus_J \mathcal{H}_J$, and, consequently, the optimal POVM can also be chosen to be of the form $\mathcal{E} = \bigoplus_J \mathcal{E}_J$, where, clearly, $\mathcal{E}_J$ acts on $\mathcal{H}_J$. To ease the notation, rather than labelling the various subspaces $\mathcal{H}_J$ by their total angular momentum J, let us simply enumerate them hereafter by natural numbers, $\alpha = 1, 2, \ldots, n+1$, and sort them by increasing value of J. Hence $J = \alpha + n'/2 - 1$. With a slight abuse of notation, we will accordingly write $\mathcal{H}_\alpha$ and enumerate the corresponding POVMs and overlaps as $\mathcal{E}_\alpha$ and c_α, respectively, where one has [cf. Eq. (4.7)]

$$c_\alpha = \binom{n' + \alpha - 1}{n'} \binom{n + n'}{n'}^{-1} . \tag{4.78}$$

A direct consequence of the block structure of the averaged states and $\mathcal{E}$ is that the overall success probability of a programmable discriminator can be expressed as

$$P_\mathrm{s} = \sum_{\alpha=1}^{n+1} p_\alpha P_{\mathrm{s},\alpha} , \tag{4.79}$$

$$p_\alpha = \mathrm{tr}\,(\sigma_i \mathbb{1}_\alpha) = \frac{2\alpha + n' - 1}{(n+1)(n+n'+1)} , \quad i = 1, 2 , \tag{4.80}$$

where $P_{\mathrm{s},\alpha}$ is the success probability of discrimination in the subspace $\mathcal{H}_\alpha$, and p_α is the probability of σ_1 and σ_2 projecting onto that subspace upon performing the measurement $\{\mathbb{1}_\alpha\}$. Likewise, P_e and Q can be expressed as a convex combination of the form (4.79).

4.4.1 Weak error margin

Let us start by considering the weak condition. If we denote the error margin by R, the weak condition reads $P_\mathrm{e} \leqslant R$. According to the previous paragraph, the optimal strategy and the corresponding success probability P_s are defined through the maximization problem

$$P_\mathrm{s} = \max_{\mathcal{E}} \sum_{\alpha=1}^{n+1} p_\alpha P_{\mathrm{s},\alpha} \quad \text{subject to} \quad \sum_{\alpha=1}^{n+1} p_\alpha P_{\mathrm{e},\alpha} \leqslant R. \tag{4.81}$$

Recall now that the POVMs $\mathcal{E}_\alpha$ are independent, and that each of them is parametrized through Eq. (3.56) by a margin $r = r_\alpha$ which, moreover, satisfies the constraint $P_{\mathrm{e},\alpha} \leqslant r_\alpha$. Therefore, Eq. (4.81) can be cast as

$$P_\mathrm{s} = \max_{\{r_\alpha\}} \sum_{\alpha=1}^{n+1} p_\alpha P^W_{\mathrm{s},\alpha}(r_\alpha) \quad \text{subject to} \quad \sum_{\alpha=1}^{n+1} p_\alpha r_\alpha = R\,, \tag{4.82}$$

where the functions $P^W_{\mathrm{s},\alpha}$ are defined as in Eq. (3.58) with $c = c_\alpha$. In other words, these functions give the success probability of discrimination in the subspaces $\mathcal{H}_\alpha$ with *weak* error margins r_α. The maximization of the success probability translates into finding the optimal set of weak margins $\{r_\alpha\}_{\alpha=1}^{n+1}$ which average, $\sum_{\alpha=1}^{n+1} p_\alpha r_\alpha$, equals a (global) margin R.

Let us start by discussing the extreme cases of this scheme. On the unambiguous side, $R = 0$, the only possible choice is $r_\alpha = 0$ for all values of α, and the success probability is hence $P_\mathrm{s}^\mathrm{UA} = 1 - Q$, where Q is given by Eq. (4.8). At the other end point, if $R \geqslant R_c = \sum_{\alpha=1}^{n+1} p_\alpha r_{c,\alpha}$, where $r_{c,\alpha}$ is the critical margin in the subspace $\mathcal{H}_\alpha$, given by Eq. (3.57) with $c = c_\alpha$, we immediately recover the minimum-error result $P_\mathrm{s}^\mathrm{ME} = 1 - P_\mathrm{e}$, with P_e given by Eq. (4.11). We will refer to R_c as the global critical margin.

An explicit expression for P_s if $0 < R < R_c$ is most easily derived by starting at the unambiguous end and progressively increasing the margin R. For a very small error margin, the Lagrange multiplier method provides the maximum. It occurs at $r_\alpha = r_\alpha^{(1)}$, where

$$r_\alpha^{(1)} = \frac{1 - c_\alpha}{\sum_{\alpha=1}^{n+1} p_\alpha (1 - c_\alpha)} R\,. \tag{4.83}$$

This solution is valid only when all (partial) error margins are below their critical values, $r_\alpha^{(1)} \leqslant r_{c,\alpha}$. If this inequality holds, the maximum success probability is $P_{\mathrm{s}} = \sum_\alpha p_\alpha P^W_{\mathrm{s},\alpha}(r^{(1)}_\alpha)$. The use of the superscript "(1)" will become clear shortly.

If we keep on increasing the global margin R, it will eventually reach a value $R = R_1$ at which the error margin of the first subspace $\mathcal{H}_1$ is saturated, namely, where $r^{(1)}_1 = r_{c,1}$. This is so because the overlaps, given in Eq. (4.78), satisfy $c_1 < c_2 < \ldots < c_{n+1} = 1$. Hence we have $r^{(1)}_1 > r^{(1)}_2 > \ldots > r^{(1)}_{n+1}$ and $r_{c,1} < r_{c,2} < \cdots < r_{c,n+1}$, according to Eqs. (4.83) and (3.57), respectively. The expression for R_1 can be read off from Eq. (4.83):

$$R_1 = \frac{r_{c,1}}{1-c_1} \sum_{\alpha=1}^{n+1} p_\alpha (1-c_\alpha)\,. \tag{4.84}$$

For $R > R_1$, the optimal value of the margin of subspace $\mathcal{H}_1$ is then frozen at the value $r_1 = r_{c,1}$, and the remaining margins are obtained by excluding the fixed contribution of the subspace $\mathcal{H}_1$, i.e., by computing the maximum on the right-hand side of

$$\begin{gathered} P_{\mathrm{s}} - p_1 P^W_{\mathrm{s},1}(r_{c,1}) = \max_{\{r_\alpha\}} \sum_{\alpha=2}^{n+1} p_\alpha P^W_{\mathrm{s},\alpha}(r_\alpha) \\ \text{subject to} \\ \sum_{\alpha=2}^{n+1} p_\alpha r_\alpha = R - p_1 r_{c,1}\,. \end{gathered} \tag{4.85}$$

The location of this maximum, which we denote by $\{r^{(2)}_\alpha\}^{n+1}_{\alpha=2}$, is formally given by Eq. (4.83) with R replaced by $R - p_1 r_{c,1}$ and the sum in the denominator running from $\alpha = 2$ to $n+1$. In this case, we have

$$P_{\mathrm{s}} = p_1 P^W_{\mathrm{s},1}(r_{c,1}) + \sum_{\alpha=2}^{n+1} p_\alpha P^W_{\mathrm{s},\alpha}(r^{(2)}_\alpha). \tag{4.86}$$

Again, this is valid only until R reaches a second saturation point R_2, i.e., provided $R_1 < R < R_2$, and so on. Clearly, the margins r_α saturate in an orderly fashion as we increase R.

Iterating the procedure described above, the optimal error margins in the interval $R_{\beta-1} \leqslant R \leqslant R_\beta$ (throughout the remaining of the Chapter, Greek indexes run from 1 to $n+1$), where $R_0 \equiv 0$ and $R_{n+1} \equiv R_c$, are found to be

$$r^{(\beta)}_\alpha = \frac{1-c_\alpha}{\chi_\beta}\left(R - \xi_\beta\right)\,, \tag{4.87}$$

where

$$R_\beta = \frac{r_{c,\beta}}{1-c_\beta}\chi_\beta + \xi_\beta\,, \tag{4.88}$$

and

$$\xi_\beta = \sum_{\alpha=1}^{\beta-1} p_\alpha r_{c,\alpha}\,, \qquad \chi_\beta = \sum_{\alpha=\beta}^{n+1} p_\alpha (1-c_\alpha)\,. \tag{4.89}$$

The success probability in this interval [analogous to Eq. (4.86)] is

$$P_\mathrm{s} = P_{\mathrm{s},\beta}^{\mathrm{sat}} + \sum_{\alpha=\beta}^{n+1} p_\alpha P_{\mathrm{s},\alpha}^{W}(r_\alpha^{(\beta)}), \tag{4.90}$$

where

$$P_{\mathrm{s},\beta}^{\mathrm{sat}} = \sum_{\alpha=1}^{\beta-1} p_\alpha P_{\mathrm{s},\alpha}(r_{c,\alpha}) = \frac{1}{2}\sum_{\alpha=1}^{\beta-1} p_\alpha \left(1+\sqrt{1-c_\alpha^2}\right) \tag{4.91}$$

is the contribution to the success probability of the subspaces where the error margins are frozen at their critical values. After some algebra, we find that the success probability can be written in a quite compact form as

$$P_\mathrm{s} = P_{\mathrm{s},\beta}^{\mathrm{sat}} + \left(\sqrt{R-\xi_\beta} + \sqrt{\chi_\beta}\right)^2, \quad R_{\beta-1} \leqslant R \leqslant R_\beta. \tag{4.92}$$

Eqs. (4.87) through (4.92) comprise our main result.

4.4.2 Strong error margin

The concept of a strong margin for programmable machines requires a more careful formulation than that of a weak margin since, in principle, there are different conditions one can impose on the various probabilities involved. For instance, one could require the strong conditions (3.52) and (3.53) for *every* possible pair of states fed into the machine, that is, for every given $\{\rho_1 = [\psi_1], \rho_2 = [\psi_2]\}$. This approach is quickly seen to be trivial since the machine, which performance is independent of the states, is required to satisfy the condition in a worst case scenario, in which $|\psi_1\rangle$ and $|\psi_2\rangle$ are arbitrarily close to each other. For any value of the error margin less than 1/2 the inconclusive probability must then approach unity, i.e., $Q \to 1$. This implies that both P_s and P_e vanish. A similar argument leads to the trivial solution $P_\mathrm{s} = P_\mathrm{e} = 1/2$ if the margin is larger than or equal to 1/2.

The task performed by a programmable discriminator can be most naturally viewed as state labelling: the machine attaches the label 1 (2) to the data if its state is identified, by a "clicking" of the operator E_1 (E_2), to be that of the qubits loaded through program port A (C); i.e., the state of the ports has the pattern $[\psi_1^{\otimes n}][\psi_1^{\otimes n'}][\psi_2^{\otimes n}]$ ($[\psi_1^{\otimes n}][\psi_2^{\otimes n'}][\psi_2^{\otimes n}]$). For this task, the relevant error probabilities are $p(2|E_1)$ and $p(1|E_2)$, namely, the probability of wrongly assigning the labels 1 and 2, respectively. It seems, therefore, more suitable for programmable discrimination to impose the strong margin conditions $p(2|E_1) \leqslant R$ and $p(1|E_2) \leqslant R$. In terms of the average states σ_1 and σ_2 in Eq. (4.1) these conditions are

$$p(2|E_1) = \frac{\operatorname{tr} E_1\sigma_2}{\operatorname{tr} E_1\sigma_1 + \operatorname{tr} E_1\sigma_2} \leqslant R\,, \tag{4.93}$$

and likewise for $p(1|E_2)$.

Note that, in contrast to the weak case, here the conditional probabilities are nonlinear functions of the POVM elements, thus the maximization of the success probability under these conditions is *a priori* more involved. To circumvent this problem, we can use the relation (3.60), which for programmable discrimination also holds, and reads

$$R^S = \frac{R^W}{P_{\rm s}(R^W) + R^W} \tag{4.94}$$

to express the (global) weak error margin R^W in terms of the strong one R^S. Then, one simply uses Eqs. (4.87) through (4.92) to obtain the maximum success probability. The inversion of Eq. (4.94) is somewhat lengthy but straightforward. The difficulty arises from the fact that the success probability, Eq. (4.92), is a piecewise function which expression depends specifically on how many margins r_α have reached their critical value $r_{c,\alpha}$ for a given R^S. Thus we need to compute the strong saturation points R^S_β, analogous to (4.88), through the relation (4.94).

4.4.3 Analysis of the results

In Fig. 4.6 we plot the maximum success probabilities for both the weak and the strong conditions as a function of a common (global) margin R, for nine program and two data copies. We also show in Fig. 4.6 the results of a numerical optimization with the strong condition (dots), which exhibit perfect agreement with our analytical solution. We observe that by allowing just a 5% error margin, the success probability increases by more than 50%. This is just an example of a general feature of programmable discrimination with an error margin: the success probability increases sharply for small values of the error margin.

A comment about the effect of the subspace $\mathcal{H}_{n+1}$ on the shape of the plots is in order. This subspace contains the completely symmetric states of the whole system ABC and, hence, it is impossible to tell if the state of the data (B) coincides with that of one program (A) or that of the other (C); more succinctly, $c_{n+1} = 1$. Therefore, half the number of conclusive answers will be correct and half of them will be wrong, and $P^W_{{\rm s},n+1} = r_{n+1}$,

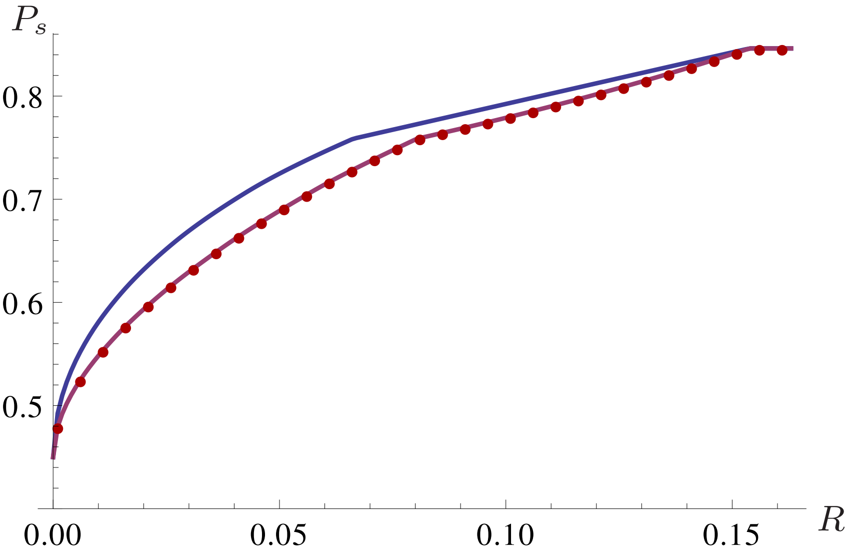


Figure 4.6. P_s vs R for a weak (upper line) and a strong (lower line) condition, for $n = 9$ and $n' = 2$. The global critical margin is $R_c \simeq 0.154$. A numerical maximization of the success probability under the strong condition (4.93) (points) is seen to agree with our analytical solution.

provided $r_{n+1} \leqslant r_{c,n+1} = 1/2$. Increasing the error margin simply allows for an equal increase in the success probability. This is reflected in the linear stretch in the upper curve in Fig. 4.6, right before the (rightmost) flat plateau. For the strong condition, the same situation arises in the interval $R_n^S \leqslant R \leqslant R_c$, but the plot of the success probability is *not* a straight line due to the nonlinear relation (4.94) between the weak and the strong margin.

An alternative (though completely equivalent) way to compute the maximum success probability with a strong margin is based on the observation that the POVMs $\mathcal{E}_\alpha$ are also fully determined by strong margins, r_α^S, through Eq. (3.62), with the exception of $\mathcal{E}_{n+1}$, for which $c = c_{n+1} = 1$ [giving rise to an ambiguity, as discussed after Eq. (3.62)]. In this approach, the success probability becomes a convex combination of $P^S_{\mathrm{s},\alpha}(r^S_\alpha)$, as in Eq. (4.79), where these functions are given in Eq. (3.61) with $c = c_\alpha$. The optimal set $\{r_\alpha^{S\,(\beta)}\}$ can be readily obtained from the weak margins in Eq. (4.87) using the relation (3.60). The strategy in the last subspace $\mathcal{H}_{n+1}$ can be easily seen to consist in abstention with a certain probability, and a random choice of the labels 1 and 2 otherwise.

The bar chart in Fig. 4.7 represents an optimal strategy in terms of the corresponding weak and strong error margins. For this example, we have chosen 11 program and two data copies. For illustration purposes, the (global)

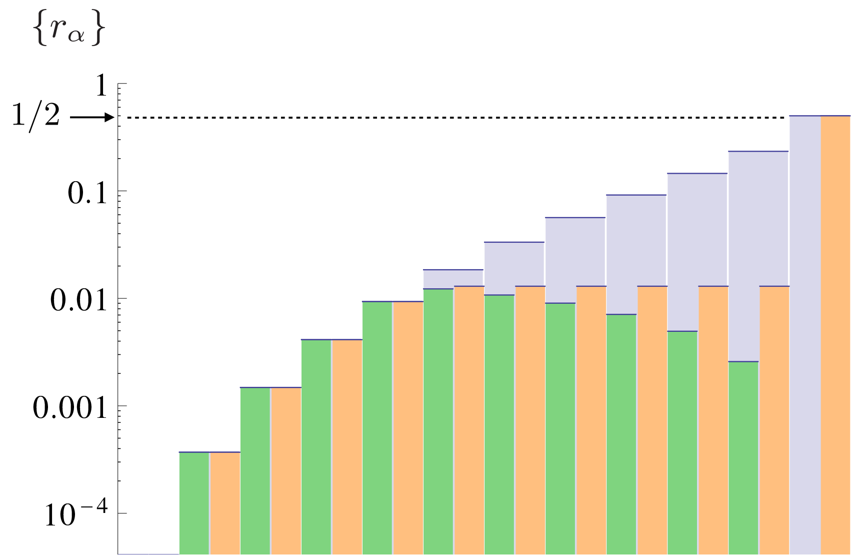


Figure 4.7. The various error margins for $n = 11$, $n' = 2$ and a (global) margin $R = 0.0055$. The full heights of the wide bars in the background (blue) represent the values of the critical margins $r_{c,\alpha}$, starting from $\alpha = 1$ (leftmost) up to $\alpha = 12$ (rightmost). For the same values of α, each pair of narrow bars represents the weak margin r^W_α [left (green)] and the strong margin r^S_α [right (orange)]. We note that the first five error margins have reached their critical value. The values for $\alpha = 1$ are very small, which explains why the corresponding bars do not show up in the chart.

margin is set to a low value of 0.0055. The wide vertical bars in the background depict the critical margins $r_{c,\alpha}$. There are 12 of them, displayed in increasing order of α (the first one is not visible because of the small value of $r_{c,1}$). On their left (right) halves, a narrow green (orange) bar depicts the optimal weak (strong) margin r^W_α (r^S_α) (we attach the subscripts W and S through the rest of the Section to avoid confusion). We note that the first 5 margins ($\alpha \leqslant 5$) have reached their critical value. For $\alpha > 5$, the weak margins decrease monotonically according to Eq. (4.87). For the last one, we have $r^W_{n+1} = r^W_{12} = 0$, which holds for any value of R, provided $R \leqslant R_n$. This must be so, since we recall that the projections of σ_1 and σ_2 onto the subspace with maximum angular momentum are indistinguishable. Clearly, allowing for $r^W_{n+1} > 0$ while there is still room for the other margins to increase cannot be optimal.

Also noticeable in Fig. 4.7 is that the set of strong margins that have not reached their critical value $r_{c,\alpha}$ has a flat profile (this does not apply to r^S_{n+1} that is always frozen to its critical value of 1/2). To provide an explanation for this, we write the equality in Eq. (4.93), which is attained if $R \leqslant R_c$, as $RP_\mathrm{s}-(1-R)P_\mathrm{e}=0$, using once again the symmetry of the problem. We next write the success and error probabilities as a convex sum over α and use the equality in the strong conditions (3.52) and (3.53) for each subspace $\mathcal{H}_\alpha$ to express $P^S_{\mathrm{e},\alpha}$ in terms of $P^S_{\mathrm{s},\alpha}$. We obtain the strong condition

$$\sum_\alpha p_\alpha P^S_{\mathrm{s},\alpha}(r^S_\alpha)\left[R-(1-R)\frac{r^S_\alpha}{1-r^S_\alpha}\right]=0. \tag{4.95}$$

The terms in square brackets can be positive or negative depending on r^S_α being smaller or larger than R, both of which are possible. So, at face value, this equation cannot explain the flat profile of r^S_α and more work is needed. Next, we use the Lagrange multiplier method to maximize $P_\mathrm{s}=\sum_\alpha p_\alpha P^S_{\mathrm{s},\alpha}(r^S_\alpha)$ and note that the dependence of $P^S_{\mathrm{s},\alpha}$ on α (i.e., the term $1-c_\alpha$) factorizes, as can be checked from Eq. (3.61). Without further calculation, we can anticipate that the optimal margins will be determined by $n+1$ equations of the form $p_\alpha(1-c_\alpha)f(r^S_\alpha)=0$, where f can be a function only of R, the Lagrange multiplier and the number of margins below their critical value. Hence, all the (unfrozen) margins will have the same optimal value. For $\beta=1$ (no frozen margins) we have the simple solution $r^{S,(1)}_\alpha=R$ for all α, and the corresponding success probability is

$$P_\mathrm{s}=\left(\frac{\sqrt{1-R}}{\sqrt{R}-\sqrt{1-R}}\right)^2\frac{nn'}{(n+1)(n'+2)} \tag{4.96}$$

for a sufficiently small strong margin R.

4.5 Discussion

In the first part of the Chapter, we have analyzed the problem of programmable discrimination of two unknown general qubit states when multiple copies of the states are provided. For pure states we have obtained the optimal unambiguous discrimination and minimum-error probabilities (Section 4.1). Some results along these lines can be found in [He and Bergou, 2007], however no closed expressions were given there. Knowing the error in the asymptotic regimes is very relevant information, as it allows to assess and compare the performance of devices in a way that is independent on the number of copies. We have obtained analytical expressions for the leading and subleading terms in several cases of interest. As could be anticipated, when the number of copies at the program ports is asymptotically large, at leading order we recover the average of the usual discrimination problem of known states in both unambiguous and minimum-error approaches. When the data port is loaded with an asymptotically large number of copies, we recover the state comparison averaged errors. These cases correspond to estimate-and-discriminate protocols, where the estimation unveils the classical information about the states.

We have also addressed, for the first time, the programmable discrimination of copies of mixed states (Section 4.2). By taking advantage of the block decomposition of permutationally invariant states to crucially reduce the computational complexity of the problem, we have obtained the minimum-error probability when the ports are loaded with copies of qubits of known purity. We have assumed that all states have the same purity. This would correspond to a scenario where all the initially pure data and program states are subject to the same depolarizing noise before entering the machine. Closed analytical results for a small number of copies can be obtained and efficiently computable expressions for a fairly large number of copies are given. The asymptotic analytical results show very good agreement with the numerics. The latter show a characteristic $1/N$ dependence with the number N of available copies—in contrast to the usual exponential decay found in standard (nonuniversal) state discrimination—and provide a very good approximation already for a relatively low number of copies when the states have high purity. For very mixed states the error probability has a drastically different behavior. Logically, in both cases the error probability monotonically decreases with increasing purity r, but in the low-purity regime the dependence is much less pronounced. The range of purities exhibiting this behavior shrinks as the number of copies increases, and the characteristic $1/N$ behavior of the asymptotic regime extends its validity over almost the whole range of purities.

We have analyzed next the fully universal discrimination machine, a device that takes in states of which nothing is known, not even their purity (Section 4.3). We have computed the minimum-error probability for three reasonable prior distributions of the purity: the hard-sphere, Bures, and Chernoff. The latter is seen to give the lowest error probability. This comes as no surprise, since the Chernoff distribution assigns larger weights to pure states (because they are better distinguished). Our results also indicate that the fully universal discrimination machine yields an error probability comparable to the discrimination of states of known purity, being that remarkably large ($r \sim 0.9$).

Finally, we have provided two generalizations of programmable state discrimination that enable control on the rate with which errors inevitably arise because of the very principles of quantum mechanics (Section 4.4). In the first, a margin is set on the average error probability of mislabeling the input data states (weak condition). In the second, a more stringent condition is required that, for each label, the probability of it being wrongly assigned is within a given margin (strong condition). Generically, in both cases, the discrimination protocol may result sometimes in an inconclusive outcome (i.e., in being unable to assign a label to the data). We have shown that there is a one-to-one correspondence between these two margins, so that weak and strong conditions turn out to be the same if their margins are related by a simple equation. These generalizations extend the range of applicability of programmable discriminators to scenarios where some rate of errors and some rate of inconclusive outcomes are both affordable; or, more specifically, to situations where a trade-off between these two rates is acceptable, which depart from the standard unambiguous (zero error) and minimum-error (zero abstention) discrimination scenarios.

Our results include the analytical expression of the success probability for the optimal programmable device as a function of both weak and strong error margins, as well as the characterization of the POVM that specifies such optimal device. From the analysis of these results, we conclude that small error margins can significantly boost the success probability; i.e., a small departure from the unambiguous scheme can translate into an important increase of the success rate while still having very reliable results (very low error rate). We provide an example of this, where a mere error margin value of 5% adds about 50% to the success probability.

Throughout this Chapter we have considered programmable discriminators to be black boxes, as we optimized always over completely general POVMs. It is very relevant to examine restricted measurement schemes compatible with a machine learning scenario, in which the machine first “learns” about the states at the program ports and then assigns a label to the states at the data port, in that particular order. In Chapter 5 we consider this scenario in detail, and we contrast the results with the ones obtained here.

CHAPTER 5

Quantum learning of qubit states

Programmable processors, as pointed out in Chapter 4, are expected to automate information processing tasks, lessening human intervention by adapting their functioning according to some input program. This adjustment, that is, the process of extraction and assimilation of information relevant to perform efficiently some task, is often called *learning*, borrowing a word most naturally linked to living beings. *Machine learning* is a broad research field that seeks to endow machines with this sort of ability, so that they can "learn" from past experience, perform "pattern recognition" or "discover patterns in scrambled data" [MacKay, 2003; Bishop, 2006]. Algorithms featuring learning capabilities have numerous practical applications, including speech and text recognition, image analysis, and data mining. In *supervised* machine learning, a machine is trained using a learning algorithm that takes a dataset as input, namely a *training set* (TS), consisting in some observations on the characteristics of certain objects. Once trained, the machine is expected to recognize these (*classification*) or other (*regression*) characteristics in upcoming new objects. On the other hand, *unsupervised* learning machines try to find structure hidden in unlabeled data.

Whereas conventional machine learning theory implicitly assumes the TS to be fundamentally classical—a set of classical features of classical objects, an array of symbols and numbers—, its quantum variant explores training with quantum objects, and, in doing so, it links the notion of learning in the real—quantum—world with the underlying physical theory on which it is grounded. *Quantum learning* [Aïmeur *et al.*, 2006] has recently raised great attention. Particularly, the use of programmable quantum processors has been investigated to address machine learning tasks such as pattern matching [Sasaki and Carlini, 2002], binary classification [Guţă and Kotłowski, 2010; Neven *et al.*, 2009; Pudenz and Lidar, 2013], feedback-adaptive quantum measurements [Hentschel and Sanders, 2010], learning of unitary transformations [Bisio *et al.*, 2010], Probably Approximately Correct learning [Servedio and Gortler, 2004], and unsupervised clustering [Lloyd *et al.*, 2013]. Quantum learning algorithms not only provide improvements over some classical learning problems, but also have a wider range of applicability. Quantum learning has also strong links with quantum control theory, and is becoming a significant element of the quantum information processing toolbox.

This Chapter is concerned with a simple, yet fundamental instance of quantum state identification, which finds its motivation in learning theory. A source produces two unknown pure qubit states with equal probability. A human expert (who knows the source specifications, for instance) classifies a number of $2n$ states produced by this source into two sets of size roughly n (statistical fluctuations of order $\sqrt{n}$ should be expected) and attaches the labels 0 and 1 to them. We view these $2n$ states as a training sample, and we set ourselves to find a universal machine that uses this sample to assign the right label to a new unknown state produced by the same source with the smallest error rate. We refer to this task as quantum classification for short. Clearly, quantum classification can be understood as a supervised quantum learning problem, as has been noticed by Guta and Kotlowski in their recent work [Guţă and Kotłowski, 2010] (though they use a slightly different setting).

It is worth mentioning that a very similar problem was proposed in [Sasaki and Carlini, 2002] under the name of “universal quantum matching machine”. The task of this machine differs from that of ours in that, rather than identifying the unknown qubit as one of the states in the TS, it determines to which of them is *closest*, thus a fidelity-related figure of merit is used instead of the error probability. The work of Sasaki and Carlini pioneered the view on the quantum classification problem as a learning protocol, and set an inspiration for later works on—the more general—programmable discrimination machines.

Of course, an absolute limit on the minimum error in quantum classification is provided by the optimal programmable discrimination machine (see Chapter 4). In that context, to ensure optimality one assumes that a fully general two-outcome joint measurement is performed on *both* the $2n$ training qubits and the qubit we wish to classify, where the observed outcome determines which of the two labels, 0 or 1, is assigned to the latter qubit. Thus, in principle, this assumption implies that, in a learning scenario, a quantum memory is needed to store the training sample till the very moment we wish to classify the unknown qubit. The issue of whether or not the joint measurement assumption can be relaxed has not yet been addressed. Nor has the issue of how the information left after the joint measurement can be used to classify a second unknown qubit produced by the same source, unless a fresh new TS is provided (which may seem unnatural in a learning context).

The main objective of this Chapter is to show that, for a sizable TS (asymptotically large n), the absolute lower bound on the probability of misclassifying the unknown qubit, set by programmable discrimination, can be attained by first performing a suitable measurement on the TS followed by a Stern-Gerlach type of measurement on the unknown qubit, where forward classical communication is used to control the parameters of the second measurement[1]. The whole protocol can thus be undersood as a learning machine (LM), which requires much less demanding assumptions while still having the same accuracy as the optimal programmable discrimination machine. All the relevant information about the TS needed to control the Stern-Gerlach measurement is kept in a *classical* memory, thus classification can be executed any time after the learning process is completed. Once trained, this machine can be subsequently used an arbitrary number of times to classify states produced by the same source. Moreover, this optimal LM is robust under noise, i.e., it still attains optimal performance if the states produced by the source undergo depolarization to any degree. Interestingly enough, in the ideal scenario where the qubit states are pure and the TS consists in exactly the same number of copies of each of the two types 0/1 (no statistical fluctuations are allowed) this LM attains the optimal programmable discrimination bound for *any* size $2n$ of the TS, not necessarily asymptotically large.

At this point it should be noted that LMs without quantum memory can be naturally assembled from two quantum information primitives: state estimation and state discrimination. We will refer to these specific constructions as "estimate-and-discriminate" (E&D) machines. The protocol they execute is as follows: by performing, e.g., an optimal covariant measurement on the n qubits in the TS labeled 0, their state $|\psi_0\rangle$ is estimated with some accuracy, and likewise the state $|\psi_1\rangle$ of the other n qubits that carry the label 1 is characterized. This classical information is stored and subsequently used to discriminate an unknown qubit state. It will be shown that the excess risk (i.e., excess average error over classification when the states $|\psi_0\rangle$ and $|\psi_1\rangle$ are perfectly known) of this protocol is twice that of the optimal LM. The

[1] Interestingly, this result is the opposite to the one found by Sasaki and Carlini for their universal quantum matching machine, where any strategy of two separate measurements is suboptimal [Sasaki and Carlini, 2002]. Again, their protocol is slightly different to ours.

fact that the E&D machine is suboptimal means that the kind of information retrieved from the TS and stored in the classical memory of the optimal LM is specific to the classification problem at hand, and that the machine itself is more than the mere assemblage of well known protocols.

We will first present our results for the ideal scenario where states are pure and no statistical fluctuation in the number of copies of each type of state is allowed. The effect of these fluctuations and the robustness of the LM optimality against noise will be postponed to the end of the Chapter.

5.1 The learning machine

In this Chapter we use the notation and conventions of Chapter 4. Before presenting our results, let us briefly recall the setting of the problem for programmable machines and its optimal solution. Neglecting statistical fluctuations, the TS of size $2n$ is given by a state pattern of the form $[\psi_0^{\otimes n}]\otimes[\psi_1^{\otimes n}]$, where we have used the shorthand notation $[\,\cdot\,]\equiv|\,\cdot\,\rangle\langle\,\cdot\,|$, and where no knowledge about the actual states $|\psi_0\rangle$ and $|\psi_1\rangle$ is assumed (the figure of merit will be an average over *all* states of this form). The qubit state that we wish to label (the *data* qubit) belongs either to the first group (it is $[\psi_0]$) or to the second one (it is $[\psi_1]$). Thus the optimal machine must discriminate between the two possible states: either $\varrho_0^n=[\psi_0^{\otimes(n+1)}]_{AB}\otimes[\psi_1^{\otimes n}]_C$, in which case it should output the label 0, or $\varrho_1^n=[\psi_0^{\otimes n}]_A\otimes[\psi_1^{\otimes(n+1)}]_{BC}$, in which case the machine should output the label 1. Here and when needed for clarity, we name the three subsystems involved in this problem A, B and C, where AC is the TS and B is the data qubit. In order to discriminate ϱ_0^n from ϱ_1^n, a joined two-outcome measurement, independent of the actual states $|\psi_0\rangle$ and $|\psi_1\rangle$, is performed on all $2n+1$ qubits. Mathematically, it is represented by a two-outcome POVM $\mathscr{E}=\{E_0,E_1=\mathbb{1}-E_0\}$. The minimum average error probability of the quantum classification process is [cf. Eq. (3.33)] $P_{\rm e}=(1-\Delta/2)/2$, where

$$\Delta=2\max_{E_0}\operatorname{tr}\left[\left(\sigma_0^n-\sigma_1^n\right)E_0\right]=\|\sigma_0^n-\sigma_1^n\|_1\ , \tag{5.1}$$

and $\sigma_{0/1}^n$ are average states analogous to the states (4.2), defined in this case as

$$\begin{aligned}\sigma_0^n &= \frac{\mathbb{1}_{n+1}\otimes\mathbb{1}_n}{d_{n+1}d_n}=\frac{\mathbb{1}_{AB}\otimes\mathbb{1}_C}{d_{AB}d_C}\,,\\ \sigma_1^n &= \frac{\mathbb{1}_n\otimes\mathbb{1}_{n+1}}{d_nd_{n+1}}=\frac{\mathbb{1}_A\otimes\mathbb{1}_{BC}}{d_Ad_{BC}}\,,\end{aligned} \tag{5.2}$$

where $\mathbb{1}_m$ stands for the projector onto the fully symmetric invariant subspace of m qubits, which has dimension $d_m = m + 1$. Sometimes, it turns out to be more convenient to use the subsystem labels, as on the right of Eq. (5.2).

Recall that the trace norm in Eq. (5.1) can be computed by switching to the total angular momentum basis, $\{|J, M\rangle\}$, and splitting it in the different contributions of the orthogonal Jordan subspaces (see Section 4.1 for details). The final answer is given by Eq. (4.11), with $n' = 1$. It takes the simple form

$$P_{\mathrm{e}}^{\mathrm{opt}} = \frac{1}{2} - \frac{1}{d_n^2 d_{n+1}} \sum_{k=0}^{n} k\sqrt{d_n^2 - k^2}\,, \tag{5.3}$$

where we have written the various values of the total angular momentum as $J = k + 1/2$. The formula (4.22) gives the asymptotic expression of $P_{\mathrm{e}}^{\mathrm{opt}}$ for large n, which in this case simply reads

$$P_e^{\mathrm{opt}} \simeq \frac{1}{6} + \frac{1}{3n}. \tag{5.4}$$

The leading order (1/6) coincides with the average error probability for *known* states $\int d\psi_0\, d\psi_1\, p_e^{\mathrm{opt}}(\psi_0, \psi_1)$, where $p_e^{\mathrm{opt}}(\psi_0, \psi_1)$ is the minimum error in discrimination between two given states $|\psi_0\rangle$ and $|\psi_1\rangle$.

The formulas above give an absolute lower bound to the error probability that can be physically attainable. We wish to show that this bound can actually be attained by a learning machine that uses a classical register to store all the relevant information obtained in the learning process regardless the size, $2n$, of the TS. A first hint that this may be possible is that the optimal measurement $\mathscr{E}$ can be shown to have positive partial transposition with respect to the partition TS/data qubit. Indeed this is a necessary condition for any measurement that consists of a local POVM on the TS which outcome is fed-forward to a second POVM on the data qubit. This class of one-way adaptive measurement can be characterized as

$$E_0 = \sum_{\mu} L_\mu \otimes D_\mu, \quad E_1 = \sum_{\mu} L_\mu \otimes (\mathbb{1}_1 - D_\mu), \tag{5.5}$$

where the positive operators L_μ (D_μ) act on the Hilbert space of the TS (data qubit we wish to classify), and $\sum_\mu L_\mu = \mathbb{1}_n \otimes \mathbb{1}_n$. The POVM $\mathscr{L} = \{L_\mu\}$ represents the learning process, and the parameter μ, which a priori may be discrete or continuous, encodes the information gathered in the measurement and required at the classification stage. For each possible value of μ, $\mathscr{D}_\mu = \{D_\mu, \mathbb{1}_1 - D_\mu\}$ defines the measurement on the data qubit, which two outcomes represent the classification decision (see Fig. 5.1). Clearly, the size of the required classical memory will be determined by the information content of the random variable μ.

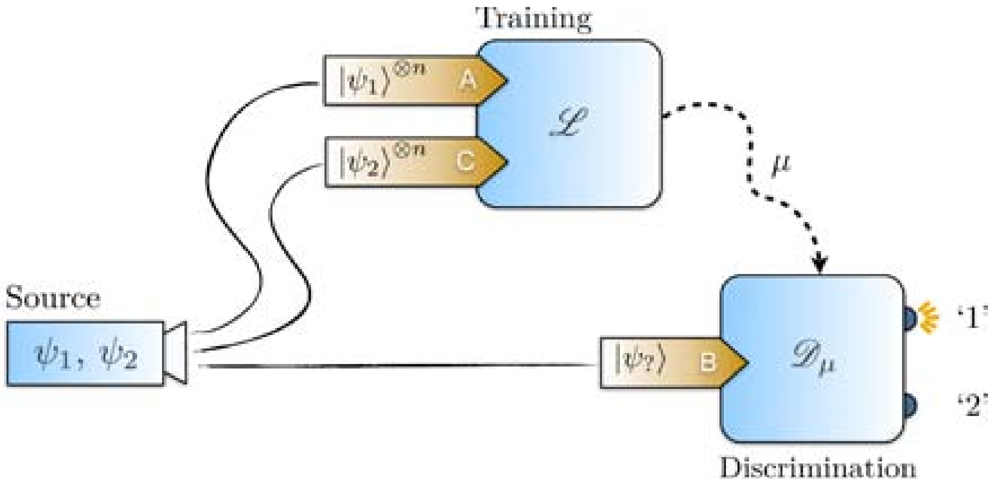


Figure 5.1. A learning protocol for qubit classification. First, a measurement $\mathscr{L}$ is performed over the TS (system AC), from which the information μ is extracted. Then, μ is used for defining a two-outcome measurement $\mathscr{D}_\mu$ that classifies the data qubit (system B) with some probability of success.

5.2 Covariance and structure of $\mathscr{L}$

We will next prove that the POVM $\mathscr{L}$, which extracts the relevant information from the TS, can be chosen to be covariant. This will also shed some light on the physical interpretation of the classical variable μ. The states (5.2) are by definition invariant under a rigid rotation acting on subsystems AC and B of the form $U = U_{AC} \otimes u$, where, throughout this Chapter, U stands for an element of the appropriate representation of SU(2), which should be obvious by context (in this case $U_{AC} = u^{\otimes 2n}$, where u is in the fundamental representation). Since $\operatorname{tr}(E_0 \sigma^n_{0/1}) = \operatorname{tr}(E_0 U^\dagger \sigma^n_{0/1} U) = \operatorname{tr}(U E_0 U^\dagger \sigma^n_{0/1})$, the positive operator $U E_0 U^\dagger$ gives the same error probability as E_0 for *any* choice of U [as can be seen from, e.g., Eq. (5.1)]. The same property thus holds for their average over the whole SU(2) group $\bar{E}_0 = \int du\, U E_0 U^\dagger$, which is invariant under rotations, and where du denotes the SU(2) Haar measure. By further exploiting rotation invariance (see Appendix B.1 for full details), $\bar{E}_0$ can be written as

$$\bar{E}_0 = \int du \left(U_{AC}\, \Omega\, U^\dagger_{AC} \right) \otimes \left(u[\uparrow] u^\dagger \right) \tag{5.6}$$

for some positive operator Ω, where we use the shorthand notation $[\uparrow] \equiv |\frac{1}{2},\frac{1}{2}\rangle\langle\frac{1}{2},\frac{1}{2}|$. Similarly, the second POVM element can be chosen to be an average, $\bar{E}_1$, of the form (5.6), with $[\downarrow] \equiv |\frac{1}{2},-\frac{1}{2}\rangle\langle\frac{1}{2},-\frac{1}{2}|$ instead of $[\uparrow]$. We immediately recognize $\bar{\mathscr{E}} = \{\bar{E}_0, \bar{E}_1\}$ to be of the form (5.5), where u,

$L_u \equiv U_{AC}\,\Omega\,U_{AC}^{\dagger}$ and $D_u \equiv u[\uparrow]u^{\dagger}$ play the role of μ, L_μ and D_μ, respectively. Hence, without loss of generality we can choose $\mathscr{L} = \{U_{AC}\,\Omega\,U_{AC}^{\dagger}\}_{\mathrm{SU}(2)}$, which is a covariant POVM with seed Ω. Note that u entirely defines the Stern-Gerlach measurement, $\mathscr{D}_u = \{u[\uparrow]u^{\dagger}, u[\downarrow]u^{\dagger}\}$, i.e., u specifies the direction along which the Stern-Gerlach has to be oriented. This is the relevant information that has to be retrieved from the TS and kept in the classical memory of the LM.

Covariance has also implications on the structure of Ω. In Appendix B.1, we show that this seed can always be written as

$$\Omega = \sum_{m=-n}^{n} \Omega_m \,; \quad \Omega_m \geqslant 0\,, \tag{5.7}$$

where

$$\sum_{m=-j}^{j} \langle j,m|\Omega_m|j,m\rangle = 2j+1, \quad 0 \leqslant j \leqslant n, \tag{5.8}$$

and j (m) stands for the total angular momentum j_{AC} (magnetic number m_{AC}) of the qubits in the TS. In other words, the seed is a direct sum of operators with a well defined magnetic number. As a result, we can interpret that Ω points along the z-axis. The constraint (5.8) ensures that $\mathscr{L}$ is a resolution of the identity.

To gain more insight into the structure of Ω, we trace subsystems B in the definition of Δ, given by the first equality in Eq. (5.1). For the covariant POVM (5.6), rotational invariance enables us to express this quantity as

$$\Delta^{\mathrm{LM}} = 2\max_{\Omega} \mathrm{tr}\,\{(\sigma_0^n - \sigma_1^n)\Omega \otimes [\uparrow]\} = 2\max_{\Omega} \mathrm{tr}\,(\Gamma_\uparrow \Omega), \tag{5.9}$$

where we have defined

$$\Gamma_\uparrow = \mathrm{tr}_B\{[\uparrow](\sigma_0^n - \sigma_1^n)\} \tag{5.10}$$

(the two resulting terms in the right-hand side are the post-measurement states of AC conditioned to the outcome $\uparrow$ after the Stern-Gerlach measurement $\mathscr{D}_{\boldsymbol{z}}$ is performed on B), and the maximization is over valid seeds (i.e., over positive operators Ω such that $\int du\, U_{AC}\,\Omega\,U_{AC}^{\dagger} = \mathbb{1}_{AC}$). We calculate $\Gamma_\uparrow$ in Appendix B.5. The resulting expression can be cast in the simple and transparent form

$$\Gamma_\uparrow = \frac{\hat{J}_z^A - \hat{J}_z^C}{d_n^2 d_{n+1}}, \tag{5.11}$$

where $\hat{J}_z^{A/C}$ is the z component of the total angular momentum operator acting on subsystem A/C, i.e., on the training qubits to which the human expert assigned the label 0/1. Eq. (5.11) suggests that the optimal Ω should project on the subspace of A (C) with maximum (minimum) magnetic number, which implies that $m_{AC} = 0$. An obvious candidate is

$$\Omega = [\phi^0]\,, \quad \left|\phi^0\right\rangle = \sum_{j=0}^{n} \sqrt{2j+1}\,|j,0\rangle\,. \tag{5.12}$$

Below we prove that indeed this seed generates the optimal LM POVM.

5.3 Optimality of the LM

We now prove our main result: the POVM $\bar{\mathscr{E}} = \{\bar{E}_0, \bar{E}_1\}$, generated from the seed state in Eq. (5.12), gives an error probability $P_{\rm e}^{\rm LM} = (1-\Delta^{\rm LM}/2)/2$ equal to the minimum error probability $P_{\rm e}^{\rm opt}$ of the optimal programmable discriminator, Eq. (5.3). It is, therefore, optimal and, moreover, it attains the absolute minimum allowed by quantum physics.

The proof goes as follows. From the very definition of error probability,

$$P_{\rm e}^{\rm LM} = \frac{1}{2}\left(\mathrm{tr}\,\sigma_1^n \bar{E}_0 + \mathrm{tr}\,\sigma_0^n \bar{E}_1\right), \tag{5.13}$$

we have

$$P_{\rm e}^{\rm LM} = \frac{\mathrm{tr}\,(\mathbb{1}_A \otimes \mathbb{1}_{BC}[\phi^0] \otimes [\uparrow]) + \mathrm{tr}\,(\mathbb{1}_{AB} \otimes \mathbb{1}_C[\phi^0] \otimes [\downarrow])}{2 d_n d_{n+1}}, \tag{5.14}$$

where we have used rotational invariance. We can further simplify this expression by writing it as

$$P_{\rm e}^{\rm LM} = \frac{\|\,\mathbb{1}_A \otimes \mathbb{1}_{BC}\,|\phi_0\rangle\,|\uparrow\rangle\,\|^2 + \|\,\mathbb{1}_{AB} \otimes \mathbb{1}_C\,|\phi_0\rangle\,|\downarrow\rangle\,\|^2}{2 d_n d_{n+1}}. \tag{5.15}$$

To compute the projections inside the norm signs we first write $|\phi^0\rangle\,|\uparrow\rangle$ ($|\phi^0\rangle\,|\downarrow\rangle$ will be considered below) in the total angular momentum basis $|J,M\rangle_{(AC)B}$, where the attached subscripts remind us how subsystems A, B and C are both *ordered* and *coupled* to give the total angular momentum J (note that a permutation of subsystems, prior to fixing the coupling, can only give rise to a global phase, thus not affecting the value of the norm we wish to compute). This is a trivial task since $|\phi^0\rangle\,|\uparrow\rangle \equiv |\phi^0\rangle_{AC}\,|\uparrow\rangle_B$, i.e., subsystems are ordered and coupled as the subscript $(AC)B$ specifies, so we just need the Clebsch-Gordan coefficients

$$\langle j \pm \tfrac{1}{2}, \tfrac{1}{2} | j, 0; \tfrac{1}{2}, \tfrac{1}{2}\rangle = \pm\sqrt{\frac{j + \frac{1}{2} \pm \frac{1}{2}}{2j+1}}. \tag{5.16}$$

The projector $\mathbb{1}_A \otimes \mathbb{1}_{BC}$, however, is naturally written as $\mathbb{1}_A \otimes \mathbb{1}_{BC} = \sum_{J,M} |J,M\rangle_{A(CB)}\langle J,M|$. This basis differs from that above in the coupling of the subsystems. To compute the projection $\mathbb{1}_A \otimes \mathbb{1}_{BC}\,|\phi^0\rangle\,|{\uparrow}\rangle$ we only need to know the overlaps between the two bases ${}_{A(CB)}\langle J,M|J,M\rangle_{(AC)B}$. Wigner's $6j$-symbols provide this information as a function of the angular momenta of the various subsystems (the overlaps are computed explicitly in Appendix B.2).

Using the Clebsch-Gordan coefficients and the overlaps between the two bases, it is not difficult to obtain

$$\mathbb{1}_A \otimes \mathbb{1}_{BC}|\phi^0\rangle\,|{\uparrow}\rangle = \sum_{j=1}^{n+1} \sqrt{j}\,\frac{\sqrt{d_n+j}-\sqrt{d_n-j}}{\sqrt{2}d_n}\,|j-\tfrac{1}{2},\tfrac{1}{2}\rangle_{A(CB)}\,, \tag{5.17}$$

An identical expression can be obtained for $\mathbb{1}_{AB} \otimes \mathbb{1}_C\,|\phi^0\rangle\,|{\downarrow}\rangle$ in the basis $|J,M\rangle_{(BA)C}$. To finish the proof, we compute the norm squared of Eq. (5.17) and substitute in Eq. (5.15). It is easy to check that this gives the expression of the error probability (5.3), i.e., $P_e^{\rm LM} = P_e^{\rm opt}$.

5.4 Memory of the LM

Let us go back to the POVM condition, specifically to the minimum number of unitary transformations needed to ensure that, given a suitable discretization $\int du \to \sum_\mu p_\mu$ of Eq. (5.6), $\{p_\mu U_\mu[\,\phi^0\,]U_\mu^\dagger\}$ is a resolution of the identity for arbitrary n. This issue is addressed in [Bagan *et al.*, 2001], where an explicit algorithm for constructing finite POVMs, including the ones we need here, is given. From the results there, we can bound the minimum number of outcomes of $\mathscr{L}$ by $2(n+1)(2n+1)$. This figure is important because its binary logarithm gives an upper bound to the minimum memory required. We see that it grows at most logarithmically with the size of the TS.

5.5 E&D machines

E&D machines can be discussed within this very framework, as they are particular instances of LMs. In this case the POVM $\mathscr{L}$ has the form $L_{\alpha i} = M_\alpha \otimes M'_i$, where $\mathscr{M} = \{M_\alpha\}$ and $\mathscr{M}' = \{M'_i\}$ are themselves POVMs on the TS subsystems A and C, respectively. The role of $\mathscr{M}$ and $\mathscr{M}'$ is to estimate (optimally) the qubit states in these subsystems [Holevo, 1982]. The measurement on B (the data qubit) now depends on the pair of outcomes of $\mathscr{M}$ and $\mathscr{M}'$: $\mathscr{D}_{\alpha i} = \{D_{\alpha i}, \mathbb{1}_1 - D_{\alpha i}\}$. It performs standard one-qubit discrimination according to the two pure-state specifications, say, the unit Bloch vectors $\boldsymbol{s}_0^\alpha$ and $\boldsymbol{s}_1^i$, estimated with $\mathscr{M}$ and $\mathscr{M}'$. In this Section, we wish to show that E&D machines perform worse than the optimal LM.

We start by tracing subsystems AC in Eq. (5.1), which for E&D reads

$$\Delta^{\text{E\&D}} = 2 \max_{\mathscr{M},\mathscr{M}'} \operatorname{tr}_B \max_{\{\mathscr{D}_{\alpha i}\}} \operatorname{tr}_{AC}[(\sigma_0^n - \sigma_1^n)E_0]. \tag{5.18}$$

If we write $\Delta^{\text{E\&D}} = \max_{\mathscr{M},\mathscr{M}'} \Delta_{\mathscr{M},\mathscr{M}'}$, we have

$$\Delta_{\mathscr{M},\mathscr{M}'} = \sum_{\alpha i} p_\alpha p'_i |\boldsymbol{r}_0^\alpha - \boldsymbol{r}_1^i|, \tag{5.19}$$

where $\boldsymbol{r}_0^\alpha$ and $\boldsymbol{r}_1^i$ are the Bloch vectors of the data qubit states

$$\rho_0^\alpha = \frac{1}{p_\alpha} \operatorname{tr}_A \left(\frac{\mathbb{1}_{n+1}^{AB}}{d_{n+1}} M_\alpha \right), \quad \rho_1^i = \frac{1}{p'_i} \operatorname{tr}_C \left(\frac{\mathbb{1}_{n+1}^{BC}}{d_{n+1}} M'_i \right), \tag{5.20}$$

conditioned to the outcomes α and i respectively, and $p_\alpha = d_n^{-1} \operatorname{tr} M_\alpha$, $p'_i = d_n^{-1} \operatorname{tr} M'_i$ are their probabilities. We now recall that optimal estimation necessarily requires that all elements of $\mathscr{M}$ must be of the form $M_\alpha = c_\alpha U_\alpha [\psi^0] U_\alpha^\dagger$, where $|\psi^0\rangle = |\frac{n}{2}, \frac{n}{2}\rangle$, $c_\alpha > 0$, and $\{U_\alpha\}$ are appropriate SU(2) rotations (analogous necessary conditions are required for $\mathscr{M}'$) [Derka *et al.*, 1998]. Substituting in Eq. (5.20) we obtain $p_\alpha = c_\alpha / d_n$, and

$$u_\alpha^\dagger \rho_0^\alpha u_\alpha = \frac{1}{d_{n+1}} \left(d_n [\uparrow] + [\downarrow] \right) \tag{5.21}$$

(a similar expression holds for ρ_1^i). This means that the Bloch vector of the data qubit conditioned to outcome α is proportional to $\boldsymbol{s}_0^\alpha$ (the Bloch vector of the corresponding estimate) and is shrunk by a factor $n/d_{n+1} = n/(n+2) \equiv \eta$. Note in passing that the shrinking factor η is independent of the measurements, provided it is optimal.

Surprisingly at first sight, POVMs that are optimal, and thus equivalent, for estimation may lead to different minimum error probabilities. In particular, the continuous covariant POVM is outperformed in the problem at hand by those with a finite number of outcomes. Optimal POVMs with few outcomes enforce large angles between the estimates $\boldsymbol{s}_0^\alpha$ and $\boldsymbol{s}_1^i$, and thus between $\boldsymbol{r}_0^\alpha$ and $\boldsymbol{r}_1^i$ ($\pi/2$ in the $n = 1$ example below). This translates into increased discrimination efficiency, as shown by Eq. (5.19), without compromising the quality of the estimation itself. Hence the orientation of $\mathscr{M}$ relative to $\mathscr{M}'$ (which for two continuous POVMs does not even make sense) plays an important role, as it does the actual number of outcomes. With an increasing size of the TS, the optimal estimation POVMs require also a larger number of outcomes and the angle between the estimates decreases in average, since they tend to fill the 2-sphere isotropically. Hence the minimum error probability is expected to approach that of two continuous POVMs. This is supported by numerical calculations. The problem of finding the optimal E&D machine for arbitrary n appears to be a hard one.Here we will give the absolute optimal E&D machine for $n=1$ and, also, we will compute the minimum-error probability for both $\mathscr{M}$ and $\mathscr{M}'$ being the continuous POVM that is optimal for estimation. The later, as mentioned, is expected to attain the optimal E&D error probability asymptotically.

We can obtain an upper bound on Eq. (5.19) by applying the Schwarz inequality. We readily find that

$$\begin{aligned}\Delta_{\mathscr{M},\mathscr{M}'} &\leqslant \sqrt{\sum_{\alpha i} p_\alpha p'_i |\boldsymbol{r}_0^\alpha - \boldsymbol{r}_1^i|^2} \\ &= \sqrt{\sum_{\alpha} p_\alpha |\boldsymbol{r}_0^\alpha|^2 + \sum_i p'_i |\boldsymbol{r}_1^i|^2}\,, \end{aligned} \tag{5.22}$$

where we have used that $\sum_\alpha p_\alpha \boldsymbol{r}_0^\alpha = \sum_i p'_i \boldsymbol{r}_1^i = 0$, as follows from the POVM condition on $\mathscr{M}$ and $\mathscr{M}'$. The maximum norm of $\boldsymbol{r}_0^\alpha$ and $\boldsymbol{r}_1^i$ is bounded by 1/3 (the shrinking factor η for $n=1$). Thus

$$\Delta_{\mathscr{M},\mathscr{M}'} \leqslant \sqrt{2}/3 < 1/\sqrt{3} = \Delta^{\mathrm{LM}}\,, \tag{5.23}$$

where the value of Δ^{LM} can be read off from Eq. (5.3). The E&D bound $\sqrt{2}/3$ is attained by the choices $M_{\uparrow/\downarrow} = [\uparrow/\downarrow]$ and $M'_{+/-} = [+/-]$, where we have used the definition $|\pm\rangle = (|\uparrow\rangle \pm |\downarrow\rangle)/\sqrt{2}$.

For arbitrary n, a simple expression for the error probability can be derived in the continuous POVM case, $\mathscr{M} = \mathscr{M}' = \{d_n U_{\boldsymbol{s}} [\,\psi^0\,] U_{\boldsymbol{s}}^\dagger\}_{\boldsymbol{s}\in\mathbb{S}^2}$, where $\boldsymbol{s}$ is a unit vector (a point on the 2-sphere $\mathbb{S}^2$) and $U_{\boldsymbol{s}}$ is the representation of the rotation that takes the unit vector along the z-axis, $\boldsymbol{z}$, into $\boldsymbol{s}$. Here $\boldsymbol{s}$ labels the outcomes of the measurement and thus plays the role of α and i. The continuous version of Eq. (5.19) can be easily computed to be

$$\Delta^{\mathrm{E\&D}} = \eta \int d\boldsymbol{s}\, |\boldsymbol{z} - \boldsymbol{s}| = \frac{4n}{3(n+2)}\,. \tag{5.24}$$

Asymptotically, we have $P_{\mathrm{e}}^{\mathrm{E\&D}} = 1/6 + 2/(3n) + \ldots$. Therefore, the excess risk, which we recall is the difference between the average error probability of the machine under consideration and that of the optimal discrimination protocol for *known* qubit states (1/6), is $R^{\mathrm{E\&D}} = 2/(3n) + \ldots$. This is twice the excess risk of the optimal programmable machine and the optimal LM, which can be read off from Eq. (5.4):

$$R^{\mathrm{LM}} = R^{\mathrm{opt}} = \frac{1}{3n} + \dots . \tag{5.25}$$

For $n = 1$, Eq. (5.23) leads to $R^{\mathrm{E\&D}} = (4-\sqrt{2})/12$. This value is already 15% larger than excess risk of the optimal LM: $R^{\mathrm{LM}} = (4 - \sqrt{3})/12$.

5.6 Robustness of LMs

So far we have adhered to the simplifying assumptions that the two types of states produced by the source are pure and, moreover, exactly equal in number. Neither of these two assumptions is likely to hold in practice, as both, interaction with the environment, i.e., decoherence and noise, and statistical fluctuations in the numbers of states of each type, will certainly take place. Here we prove that the performance of the optimal LM is not altered by these effects in the asymptotic limit of large TS. More precisely, the excess risk of the optimal LM remains equal to that of the optimal programmable discriminator to leading order in $1/n$ when noise and statistical fluctuations are taken into account.

Let us first consider the impact of noise, which we will assume isotropic and uncorrelated. Hence, instead of producing $[\psi_{0/1}]$, the source produces copies of

$$\rho_{0/1} = r[\psi_{0/1}] + (1-r)\frac{\mathbb{1}}{2} , \quad 0 < r \leqslant 1 . \tag{5.26}$$

In contrast to the pure qubits case, where $[\psi_{0/1}^{\otimes n}]$ belongs to the fully symmetric invariant subspace of maximum angular momentum $j = n/2$, the state of A/C is now a full-rank matrix of the form $\rho_{0/1}^{\otimes n}$. Hence, as showed in Section 3.4.1, it has projections on all the orthogonal subspaces $\mathscr{S}_j \otimes \mathbb{C}^{\nu_j^n}$, where $\mathscr{S}_j = \mathrm{span}(\{|j,m\rangle\}_{m=-j}^{j})$, $\mathbb{C}^{\nu_j^n}$ is the ν_j^n-dimensional multiplicity space of the representation with total angular momentum j, and j is in the range from 0 (1/2) to $n/2$ if n is even (odd). Therefore $\rho_{0/1}^{\otimes n}$ is block-diagonal in the total angular momentum eigenbasis. The multiplicity space $\mathbb{C}^{\nu_j^n}$ carries the label of the ν_j^n different equivalent representations of given j, which arise from the various ways the individual qubits can couple to produce total angular momentum j. For permutation invariant states (such as $\rho_{0/1}^{\otimes n}$), this has no physical relevance and the only effect of $\mathbb{C}^{\nu_j^n}$ in calculations is through its dimension ν_j^n, given by Eq. (3.70). The multiplicity space will hence be dropped throughout the rest of the Chapter.

The average states now become a direct sum of the form

$$\int d\psi_0\, d\psi_1\, \rho_0^{\otimes(n+1)} \otimes \rho_1^{\otimes n} \;=\; \sum_\xi p_\xi^n \sigma_{0,\xi}^n, \tag{5.27}$$

$$\int d\psi_0\, d\psi_1\, \rho_0^{\otimes n} \otimes \rho_1^{\otimes(n+1)} \;=\; \sum_\xi p_\xi^n \sigma_{1,\xi}^n, \tag{5.28}$$

where we use the shorthand notation $\xi = \{j_A, j_C\}$ [each angular momentum ranges from 0 (1/2) to $n/2$ for n even (odd)], and $p_\xi^n = p_{j_A}^n p_{j_C}^n$ is the probability of any of the two average states projecting on the block labeled ξ. Hence

$$\Delta^{\mathrm{LM}} = \sum_\xi p_\xi^n \left\| \sigma_{0,\xi}^n - \sigma_{1,\xi}^n \right\|_1 . \tag{5.29}$$

The number of terms in Eq. (5.29) is $[(2n+3\pm 1)/4]^2$ for even/odd n. It grows quadratically with n, in contrast to the pure state case for which there is a single contribution corresponding to $j_A = j_C = n/2$. In the asymptotic limit of large n, however, a big simplification arises because of the following two results[2]. The first result is that, for each ξ of the form $\xi = \{j, j\}$ $(j_A = j_C = j)$, the relation

$$\sigma_{0,\xi}^n - \sigma_{1,\xi}^n = \frac{r\langle \hat{J}_z\rangle_j}{j} \left(\sigma_0^{2j} - \sigma_1^{2j} \right) \tag{5.30}$$

holds, where $\sigma_{0/1}^{2j}$ are the average states (5.2) for a number of $2j$ *pure* qubits. Here $\langle \hat{J}_z\rangle_j$ is the expectation value restricted to $\mathscr{S}_j$ of the z-component of the angular momentum in the state $\rho^{\otimes n}$, where ρ has Bloch vector $r\boldsymbol{z}$. Eq. (5.30) is an exact algebraic identity that holds for any value of j, n and r (it bears no relation whatsoever to measurements of any kind). The second result is that, for large n, both $p_{j_A}^n$ and $p_{j_C}^n$ become continuous probability distributions, $p_n(x_A)$ and $p_n(x_C)$, where $x_{A/C} = 2j_{A/C}/n \in [0,1]$. Asymptotically, they approach Dirac delta functions peaked at $x_A = x_C = r$. Hence the only relevant contribution to Δ^{LM} comes from $\xi = \{rn/2, rn/2\}$. It then follows that in the asymptotic limit

$$\sum_\xi p_\xi^n \left(\sigma_{0,\xi}^n - \sigma_{1,\xi}^n \right) \simeq \frac{2\langle \hat{J}_z\rangle_{rn/2}}{n} \left(\sigma_0^{rn} - \sigma_1^{rn} \right) . \tag{5.31}$$

This last equation tells us that mixed-state quantum classification using a TS of size $2n$ is equivalent to its *pure*-state version for a TS of size $2nr$, provided n is asymptotically large. In particular, our proof of optimality above also

[2] Here we just state the results. We derive them in detail in Appendices B.3 and B.4.

holds for *arbitrary* $r \in (0, 1]$ if the TS is sizable enough, and $R^{\mathrm{LM}} \simeq R^{\mathrm{opt}}$. This result is much stronger than robustness against decoherence, which only would require optimality for values of r close to unity.

From Eqs. (5.29) and (5.31) one can easily compute Δ^{LM} for arbitrary r using that [Gendra *et al.*, 2012] $\langle \hat{J}_z \rangle_j \simeq j - (1-r)/(2r)$ up to exponentially vanishing terms. The trace norm of $\sigma_0^{rn} - \sigma_1^{rn}$ can be retrieved from, e.g., Eq. (5.25). For rn *pure* qubits one has $\|\sigma_0^{rn} - \sigma_1^{rn}\|_1 \simeq (4/3)[1 - 1/(rn)]$. After some trivial algebra we obtain

$$P_e^{\mathrm{LM}} = \frac{1}{2} - \frac{r}{3} + \frac{1}{3rn} + o(n^{-1}) \tag{5.32}$$

for the error probability, in agreement with the optimal programmable machine value given by Eq. (4.71), as claimed above. This corresponds to an excess risk of

$$R^{\mathrm{LM}} = \frac{1}{3rn} + o(n^{-1}) = R^{\mathrm{opt}} \,. \tag{5.33}$$

In the nonasymptotic case, the sum in Eq. (5.29) is not restricted to $\xi = \{j, j\}$ and the calculation of the excess risk becomes very involved. Rather than attempting to obtain an analytical result, for small training samples we have resorted to a numerical optimization. We first note that Eqs. (5.7) through (5.11) define a *semidefinite programming* optimization problem (SDP), for which very efficient numerical algorithms have been developed [Vandenberghe and Boyd, 1996]. In this framework, one maximizes the objective function Δ^{LM} [second equality in Eq. (5.9)] of the SDP variables $\Omega_m \geqslant 0$, subject to the linear condition (5.8). We use this approach to compute the error probability, or equivalently, the excess risk of a LM for mixed-state quantum classification of small samples ($n \leqslant 5$), where no analytical expression of the optimal seed is known. For mixed states the expression of $\Gamma_\uparrow$ and Ω_m can be found in the Appendix, Eqs. (B.5) through (B.7).

Our results are shown in Fig. 5.2, where we plot R^{LM} (shaped dots) and the lower bounds given by R^{opt} (solid lines) as a function of the purity r for up to $n = 5$. We note that the excess risk of the optimal LM is always remarkably close to the absolute minimum provided by the optimal programmable machine, and in the worst case ($n = 2$) it is only 0.4% larger. For $n = 1$ we see that $R^{\mathrm{LM}} = R^{\mathrm{opt}}$ for any value of r. This must be the case since for a single qubit in A and C one has $j_A = j_C = 1/2$, and Eq. (5.30) holds.

We now turn to robustness against statistical fluctuations in the number of states of each type produced by the source. In a real scenario one has to expect that $j_A = n_A/2 \neq n_C/2 = j_C$, $n_A + n_B = 2n$. Hence $\Gamma_\uparrow$ has the general form (B.5), which gives us a hint that our choice $\Omega = \Omega_{m=0}$ may not

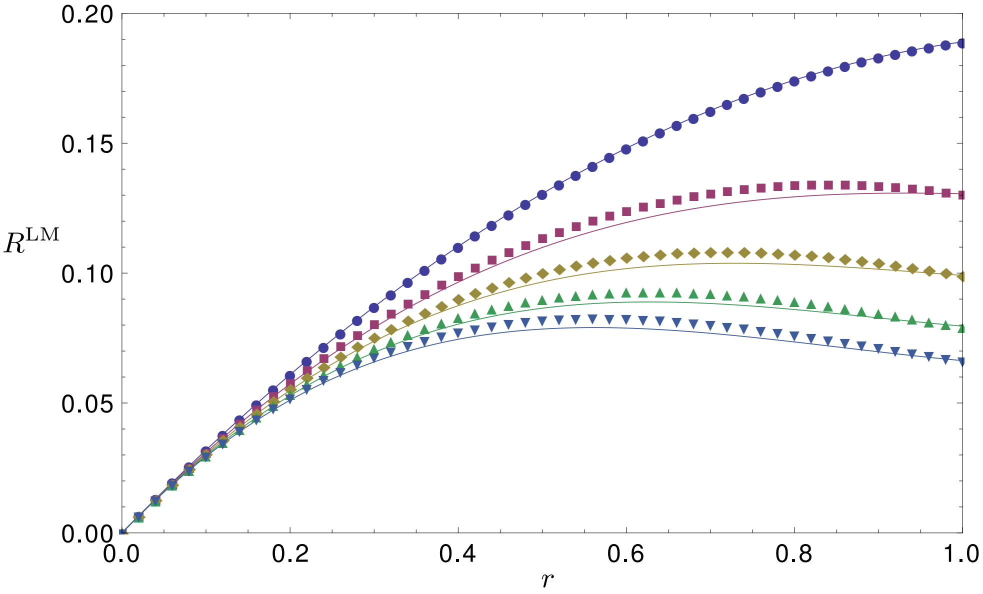


Figure 5.2. Excess risk R^{LM} (points) and its corresponding lower bound R^{opt} (lines), both as a function of the purity r, and for values of n ranging from 1 to 5 (from top to bottom).

be optimal for finite n. This has been confirmed by numerical analysis using the same SDP approach discussed above. Here, we show that the asymptotic performance (for large training samples) of the optimal LM, however, is still the same as that of the optimal programmable discriminator running under the same conditions (mixed states and statistical fluctuations in $n_{A/C}$).

Asymptotically, a real source for the problem at hand will typically produce $n_{A/C} = n \pm \delta\sqrt{n}$ mixed copies of each type. In Appendix B.4, it is shown that the relation (5.31) still holds in this case if n is large. It reads

$$\sigma^n_{0,\xi} - \sigma^n_{1,\xi} \simeq r\left(1 - \frac{1-r}{nr^2}\right)\left(\sigma_0^{rn} - \sigma_1^{rn}\right) \tag{5.34}$$

(δ first appears at order $n^{-3/2}$). Hence the effect of both statistical fluctuations in $n_{A/C}$ and noise (already considered above) is independent of the machine used for quantum classification (i.e., it is the same for LM, programmable machines, E&D, ...). In particular, the relation (5.33), $R^{\mathrm{LM}} = R^{\mathrm{opt}}$, between the excess rate of the optimal LM and its absolute limit given by the optimal programmable discriminator still holds asymptotically, which proves robustness.

To illustrate this, let us consider the effect of statistical fluctuations in $n_{A/C}$ for pure states. The optimal programmable machine for arbitrary

n_A, n_B and n_C is discussed in Appendix A.2. The error probability for the case at hand ($n_B = 1$) can be read off directly from Eq. (A.8), and its asymptotic form when n_A and n_C are both very large can be easily derived using Euler-Maclaurin's summation formula. The result up to subleading order is

$$P_e^{\rm opt} \simeq \frac{1}{6}\left(1+\frac{1}{n_A}+\frac{1}{n_C}\right),$$

which leads to

$$R^{\rm opt} = \frac{1}{6}\left(\frac{1}{n_A}+\frac{1}{n_C}\right)+\dots . \tag{5.35}$$

We see that when $n_{A/C} = n \pm \delta\sqrt{n}$ (i.e., when statistical fluctuations in $n_{A/C}$ are taken into account) one still has $R^{\rm opt} \simeq 1/(3n) \simeq R^{\rm LM}$.

5.7 Discussion

We have presented a *supervised* quantum learning machine that classifies a single qubit prepared in a pure but otherwise unknown state after it has been trained with a number of already classified qubits. Its performance attains the absolute bound given by the optimal programmable discrimination machine. This learning machine does not require quantum memory and can also be reused without retraining, which may save a lot of resources. The machine has been shown to be robust against noise and statistical fluctuations in the number of states of each type produced by the source. For small sized training sets the machine is very close to optimal, attaining an excess risk that is larger than the absolute lower limit by at most 0.4%. In the absence of noise and statistical fluctuations, the machine attains optimality for *any* size of the training set.

One may rise the question of whether or not the separated measurements on the training set and data qubit can be reversed in time; in a classical scenario where, e.g., one has to identify one of two faces based on a stack of training portraits, it is obvious that, without memory limitations, the order of training and data observation can be reversed (in both cases the final decision is taken based on the very same information). We will briefly show that this is not so in the quantum world. In the reversed setting, the machine first performs a measurement $\mathscr{D}$, with each element of rank one, $u_\mu[\uparrow]u_\mu^\dagger$, and stores the information (which of the possible outcomes is obtained) in the classical memory to control the measurement to be performed on the training set in a later time. The probability of error conditioned to one of the outcomes, say $\uparrow$, is given by the Helstrom formula $P_e^\uparrow = (1 - \|\Gamma_\uparrow\|_1/2)/2$, where $\Gamma_\uparrow$ is defined in Eq. (5.10). Using Eq. (5.11) one has $\|\Gamma_{\uparrow}\|_1 = d_n^{-2} d_{n+1}^{-1} \sum_{m,m'} |m-m'| = n/[3(n+1)]$. The averaged error probability is then

$$P_e^{\overset{\mathrm{LM}}{\leftarrow}} = \frac{1}{2}\left(1 - \frac{1}{6}\frac{n}{n+1}\right). \tag{5.36}$$

In the limit of infinite copies we obtain $P_e^{\overset{\mathrm{LM}}{\leftarrow}} \simeq 5/12$, which is way larger than $P_e^{\mathrm{LM}} \simeq 1/6$. The same minimum-error probability of Eq. (5.36) can be attained by performing a Stern-Gerlach measurement on the data qubit, which requires just one bit of classical memory. This is all the classical information that we can hope to retrieve from the data qubit, in agreement with Holevo's bound [Holevo, 1973]. This clearly limits the possibilities of a correct classification—very much in the same way as in face identification with limited memory size. In contrast, the amount of classical information "sent forward" in the optimal learning machine goes as the logarithm of the size of the training sample. This asymmetry also shows that, despite the separability of the measurements, nonclassical correlations between the training set and the data qubit play an important role in quantum learning.

CHAPTER 6

Quantum learning of coherent states

This Chapter analyzes the effect of uncertainty in discriminating between two coherent states in a learning context, following the scheme for qubits presented in the previous Chapter. Coherent states are the states produced by an ideal laser, and they comprise a very specific class among the states of continuous-variables (CV) systems, i.e., quantum systems with Hilbert spaces of infinite dimension like, for instance, the bosonic modes of an electromagnetic field. States of this type have been absent up to this point in the dissertation (only finite-dimensional systems have been considered so far), hence a few words about them are in order. Also, the mathematical toolbox required to deal with CV systems is quite different. For a technical overview on the basic tools needed for this Chapter, refer to Appendix C.

The quantum information research field divides itself in two branches, depending on the subject of study: finite dimensional systems, and CV systems. While traditionally the biggest efforts were put into the former type of systems, the study of CV systems as resources for quantum information processing has gradually become a matter of paramount importance. Supporting this assertion stands the great versatility that CV states have shown within the field, from the ease in their preparation and control in the experimental ground to their utility as subjects of genuinely quantum information processing tasks, such as quantum teleportation, quantum cloning, quantum key distribution, and quantum dense coding [Braunstein and van Loock, 2005; Eisert and Plenio, 2003; Cerf, 2007]. The class of Gaussian states, i.e., CV states that follow Gaussian statistics, receives most of the attention in the field of quantum information with CV systems [Weedbrook *et al.*, 2012].

This is mainly for two reasons: first, Gaussian states have a very simple mathematical characterization and, second, they describe appropriately the most common states of light that are realized with current technology.

The discrimination of Gaussian states plays a central role in the CV framework and, among all Gaussian states, coherent states stand out for its relevance in quantum optical communication theory. Lasers are widely used in current telecommunication systems, and the transmission of information can be theoretically modelled by bits encoded in the amplitude or phase modulation of a laser beam. The basic task of distinguishing two coherent states in an optimal way is thus of great interest, since lower chances of misidentification translate into higher transfer rates between the sender and the receiver.

The discrimination of coherent states has been considered within the two main approaches, that is minimum-error (Section 3.3.1) and unambiguous discrimination (Section 3.3.2), although the former is way more developed. Generically, a logical bit can be encoded in two possible coherent states $|0\rangle$ and $|2\alpha\rangle$, via amplitude modulation, or in the states $|\alpha\rangle$ and $|-\alpha\rangle$, via a phase shift. Both encoding schemes are equivalent, since one can move from one to the other by applying a displacement operator $\hat{D}(\alpha)$ in both states. In the minimum-error approach, the theoretical minimum for the probability of error is simply given by the Helstrom formula for pure states (3.34), as

$$P_e = \frac{1}{2}\left(1 - \sqrt{1 - e^{-4|\alpha|^2}}\right) , \tag{6.1}$$

where the overlap $|\langle\alpha|\beta\rangle|^2 = e^{-|\alpha-\beta|^2}$ has been used, and the probabilities of occurrence of each possible state have been taken to be equal for simplicity. A variety of implementations have been devised to achieve the task, e.g., the Kennedy receiver [Kennedy *et al.*, 1973], based on photon counting; the Dolinar receiver [Dolinar, 1973], a modification of the Kennedy receiver with real-time quantum feedback; and the homodyne receiver (see Section C.3.1)[1]. Concerning the unambiguous approach to the discrimination problem, results include the unambiguous discrimination between two known coherent states [Chefles and Barnett, 1998a; Banaszek, 1999], and its *programmable* version (see Chapter 4), i.e., when the value of the amplitude α is completely unknown [Sedlák *et al.*, 2007, 2009; Bartůšková *et al.*, 2008].

The purpose of this Chapter is to explore the fundamental task of discriminating between two coherent states with minimum error, when the available

[1] While the latter is the simplest procedure, it does not achieve optimality. However, for weak coherent states ($|\alpha|^2 < 0.4$), it yields an error probability very close to the optimal value P_e, and it is optimal among all Gaussian measurements [Takeoka and Sasaki, 2008]. In fact, the only optimal one of the three mentioned is the Dolinar receiver.

information about their amplitudes is incomplete. The simplest instance of such problem is a partial knowledge situation: the discrimination between the vacuum state, $|0\rangle$, and some coherent state, $|\alpha\rangle$, where the value of α is not provided beforehand in the classical sense, but instead embedded in a number n of ancillary modes in the state $|\alpha\rangle^{\otimes n}$. The question of whether such a discrimination scheme can be regarded as a learning protocol with equal performance than the most general protocol arises, thus extending the concepts settled in Chapters 4 and 5 to the CV realm.

Before starting with the calculations and to motivate the problem investigated in this Chapter, let me define the specifics of the setting in the context of a quantum-enhanced readout of classically-stored information.

6.1 Quantum reading of classical information

Imagine a classical memory register modelled by an array of cells, where each cell contains a reflective medium with two possible reflectivities r_0 and r_1. To read the information stored in the register, one shines light into one of the cells and analyzes its reflection. The task essentially consists in discriminating the two possible states of the reflected signal, which depend on the reflectivity of the medium and thus encode the logical bit stored in the cell. In the seminal paper of *quantum reading* [Pirandola, 2011], the author takes advantage of ancillary modes to prepare an initial entangled state between those and the signal. The reflected signal is sent together with the ancillas to a detector, where a joint discrimination measurement is performed. A purely quantum resource (entanglement) is thus introduced, enhancing the probability of a successful identification of the encoded bit[2]. The idea of using nonclassical light to retrieve classical information can be traced back to the precursory work of *quantum illumination* [Lloyd, 2008; Tan *et al.*, 2008], where the presence of a low-reflectivity object in a bright thermal-noise bath is detected with higher accuracy when entangled light is sent to illuminate the target region.

In this Chapter we consider a reading scenario with an imperfect coherent light source and no initial entanglement involved. The proposed scheme is as follows (see Fig. 6.1). We model an ideal classical memory by a *register* made of cells that contain either a transparent medium ($r_0 = 0$) or a highly reflective one ($r_1 = 1$). A *reader*, comprised by a *transmitter* and a *receiver*, extracts the information of each cell. The transmitter is a source that produces coherent states of a certain amplitude α. The value of α is not known

[2] In particular, Pirandola shows that a two-mode squeezed vacuum state outperforms any classical light, in the regime of few photons and high reflectivity memories.

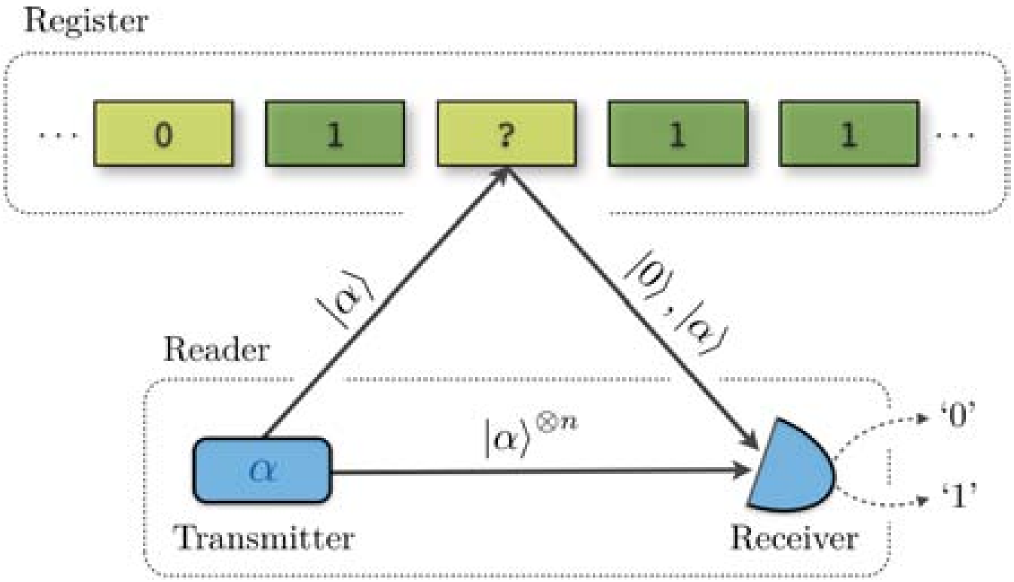


Figure 6.1. A quantum reading scheme that uses a coherent signal $|\alpha\rangle$, produced by a transmitter, to illuminate a cell of a register that stores a bit of information. A receiver extracts this bit by distinguishing between the two possible states of the reflected signal, $|0\rangle$ and $|\alpha\rangle$, assisted by n ancillary modes sent directly by the transmitter.

with certainty due, for instance, to imperfections in the source, but it can be statistically localized in a Gaussian distribution around some *known* α_0. A signal state $|\alpha\rangle$ is sent towards a cell of the register and, if it contains the transparent medium, it goes through; if it hits the highly reflective medium, it is reflected back to the receiver in an unperturbed form. This means that we have two possibilities at the entrance of the receiver upon arrival of the signal: either nothing arrives, and we represent this situation as the vacuum state $|0\rangle$, or it is the reflected signal, which is represented by the same signal state $|\alpha\rangle$. To aid in the discrimination of the signal, we alleviate the effects of the uncertainty in α by considering that n ancillary modes are produced by the transmitter in the global state $|\alpha\rangle^{\otimes n}$ and sent directly to the receiver. The receiver then performs measurements over the signal and the ancillas and outputs a binary result, corresponding with some probability to the bit stored in the irradiated cell.

We now set ourselves to answer the following questions: (i) which is the optimal (unrestricted) measurement, in terms of the error probability, that the receiver can perform? and (ii) is a *joint* measurement, performed over the signal together with the ancillas, necessary to achieve optimality? To do so, we first obtain the optimal minimum-error probability considering collective measurements (Section 6.2). Then, we devise an estimate-and-discriminate strategy, consisting in first estimating α by measuring the ancillary modes, and then using the acquired information to determine the signal state by a discrimination measurement tuned to distinguish the vacuum state $|0\rangle$ from a coherent state with the estimated amplitude (Section 6.3). We show that a collective measurement provides a lower *excess risk* (i.e., excess asymptotic average error over discrimination when α is known) than any Gaussian estimate-and-discriminate strategy, and we conjecture (and provide strong evidence) that this is the case for any LOCC strategy.

6.2 Collective strategy

The global state that arrives at the receiver can be expressed as either $[\alpha]^{\otimes n} \otimes [0]$ or $[\alpha]^{\otimes n} \otimes [\alpha]$, where the shorthand notation $[\,\cdot\,] \equiv |\,\cdot\,\rangle\langle\,\cdot\,|$ will be used throughout the Chapter. For simplicity, we take equal a priori probabilities of occurrence of each state. We will always consider the signal state to be that of the last mode, and all the previous modes will correspond to ancillas. First of all, note that the information carried by the ancillary modes can be conveniently "concentrated" into a single mode by means of a sequence of unbalanced beam splitters[3]. The action of a beam splitter over a pair of coherent states $|\alpha\rangle \otimes |\beta\rangle$ yields

$$|\alpha\rangle \otimes |\beta\rangle \longrightarrow \left|\sqrt{T}\alpha + \sqrt{R}\beta\right\rangle \otimes \left|-\sqrt{R}\alpha + \sqrt{T}\beta\right\rangle , \tag{6.2}$$

where T is the transmissivity of the beam splitter, R is its reflectivity, and $T + R = 1$. A balanced beam splitter ($T = R = 1/2$) acting on the first two ancillary modes thus returns $|\alpha\rangle \otimes |\alpha\rangle \longrightarrow \left|\sqrt{2}\alpha\right\rangle \otimes |0\rangle$. Since the beam splitter preserves the tensor product structure of the two modes, one can treat separately the first output mode and use it as input in a second beam splitter, together with the next ancillary mode. By choosing appropriately the values of T and R, the transformation $\left|\sqrt{2}\alpha\right\rangle \otimes |\alpha\rangle \longrightarrow \left|\sqrt{3}\alpha\right\rangle \otimes |0\rangle$ can be achieved. Applying this process sequentially over the n ancillary modes, we perform the transformation

$$|\alpha\rangle^{\otimes n} \longrightarrow \left|\sqrt{n}\alpha\right\rangle \otimes |0\rangle^{\otimes n-1} . \tag{6.3}$$

Note that this is a deterministic process, and that no information is lost, for it is contained completely in the complex parameter α. This operation allows

[3] See, e.g., Section III A in [Sedlák *et al.*, 2008] for details.

us to effectively deal with only two modes. The two possible global states entering the receiver become $[\sqrt{n}\alpha] \otimes [0]$ and $[\sqrt{n}\alpha] \otimes [\alpha]$.

The parameter α is not known with certainty. Building on the Bayesian ideas used in Chapter 4 to embed this lack of information into *average* global states, we immediately see that a flat prior distribution for α, as we considered for qubits, is not reasonable in this case. On the one hand, such prior would yield divergent average states of infinite energy, since the phase space is infinite. On the other hand, in a real situation it is not reasonable at all to assume that *all* amplitudes α are equally probable. The usual procedure in these cases is to consider that a small number of ancillary modes is used to make a rough estimation of α, such that our prior becomes a Gaussian probability distribution centred at α_0, which width goes as $\sim 1/\sqrt{n}$ [4]. Under these considerations, we express the true amplitude α as

$$\alpha \approx \alpha_0 + u/\sqrt{n}\,, \quad u \in \mathbb{C}\,, \tag{6.4}$$

where the parameter u follows the Gaussian distribution

$$G(u) = \frac{1}{\pi\mu^2} e^{-u^2/\mu^2}\,. \tag{6.5}$$

We have introduced the free parameter μ to define the width of $G(u)$. Once we come to the analysis of the asymptotic regime of large n, we will get rid of it by taking the limit $\mu \to \infty$, thus μ will not play any meaningful role.

Using the prior information acquired through the rough estimation, that is Eqs. (6.4) and (6.5), we compute the average global states arriving at the receiver

$$\sigma_1 = \int G(u) \left[\sqrt{n}\alpha_0 + u\right] \otimes [0]\, d^2u\,, \tag{6.6}$$

$$\sigma_2 = \int G(u) \left[\sqrt{n}\alpha_0 + u\right] \otimes \left[\alpha_0 + u/\sqrt{n}\right] d^2u\,. \tag{6.7}$$

The optimal measurement to determine the state of the signal is the Helstrom measurement for the discrimination of the states σ_1 and σ_2, that yields the

[4] Since we are interested in comparing the asymptotic performance of discrimination strategies in the limit of large n, the number of modes used for the rough estimation is negligible, i.e., $\tilde{n} = n^{1-\epsilon}$. Then, it can be shown that α belongs to a neighbourhood of size $n^{-1/2+\epsilon}$ centred at α_0, with probability converging to one (this is shown, though in a classical statistical context, in [Gill and Levit, 1995]). Moreover, this happens to be true for any model of i.i.d. quantum states ρ (regardless their dimensionality), hence the analysis of the asymptotic behaviour of any estimation model of this sort can be restricted to a local Gaussian model, centred at a fixed state ρ_0. This is known as *local asymptotic normality* [Gill and Guţă, 2013].

average minimum-error probability (see Eq. 3.33)

$$P_e^{\text{opt}}(n) = \frac{1}{2}\left(1 - \frac{1}{2}\left\|\sigma_1 - \sigma_2\right\|_1\right) . \tag{6.8}$$

[5]The technical difficulty in computing $P_e^{\text{opt}}(n)$ resides in that $\sigma_1 - \sigma_2$ is an infinite-dimensional full-rank matrix, hence its trace norm does not have a computable analytic expression for arbitrary finite n. Despite this, one can still resort to analytical methods in the asymptotic regime $n \to \infty$ by treating the states perturbatively. To ease this calculation, we first apply the displacement operator

$$\hat{D}(\alpha_0) = \hat{D}_1(-\sqrt{n}\alpha_0)\hat{D}_2(-\alpha_0) \tag{6.9}$$

to the states σ_1 and σ_2, where $\hat{D}_1$ ($\hat{D}_2$) acts on the first (second) mode, and we obtain the displaced global states

$$\begin{aligned} \bar{\sigma}_1 &= \hat{D}(\alpha_0)\sigma_1\hat{D}^\dagger(\alpha_0) = \int G(u)\,[u] \otimes [-\alpha_0]\,d^2u\,, & (6.10)\\ \bar{\sigma}_2 &= \hat{D}(\alpha_0)\sigma_2\hat{D}^\dagger(\alpha_0) = \int G(u)\,[u] \otimes \left[u/\sqrt{n}\right] d^2u\,. & (6.11) \end{aligned}$$

Since both states have been displaced the same amount, the trace norm does not change, i.e., $\left\|\sigma_0 - \sigma_1\right\|_1 = \left\|\bar{\sigma}_0 - \bar{\sigma}_1\right\|_1$. Eq. (6.10) directly yields

$$\bar{\sigma}_1 = \sum_{k=0}^{\infty} c_k \left|k\right\rangle\!\left\langle k\right| \otimes \left|-\alpha_0\right\rangle\!\left\langle -\alpha_0\right| , \tag{6.12}$$

where $c_k = \mu^{2k}/[(\mu^2+1)^{k+1}]$. Note that, as a result of the average, the first mode in Eq. (6.12) corresponds to a thermal state with average photon number μ^2. Note also that the n-dependence is entirely in $\bar{\sigma}_2$. In the limit $n \to \infty$, we can expand the second mode of $\bar{\sigma}_2$ by expressing it in the Fock basis as

$$\left|u/\sqrt{n}\right\rangle = e^{-\frac{|u|^2}{2n}} \sum_k \frac{(u/\sqrt{n})^k}{\sqrt{k!}} \left|k\right\rangle . \tag{6.13}$$

Then, up to order $1/n$ its asymptotic expansion gives

$$\begin{aligned} [u/\sqrt{n}] \sim{}& \left|0\right\rangle\!\left\langle 0\right| + \frac{1}{\sqrt{n}}\left(u\left|1\right\rangle\!\left\langle 0\right| + u^*\left|0\right\rangle\!\left\langle 1\right|\right) \\ &+ \frac{1}{n}\left\{|u|^2\left(\left|1\right\rangle\!\left\langle 1\right| - \left|0\right\rangle\!\left\langle 0\right|\right) + \frac{1}{\sqrt{2}}\left[u^2\left|2\right\rangle\!\left\langle 0\right| + (u^*)^2\left|0\right\rangle\!\left\langle 2\right|\right]\right\} . \end{aligned} \tag{6.14}$$

[5]Note that, *sensu stricto*, the dependence of $P_e^{\text{opt}}(n)$ on the localisation parameter α_0 should be made explicit. Keep in mind that, in general, all quantities computed in this Chapter will depend on α_0. Thus for the sake of notation clarity, we omit it hereafter when no confusion arises.

Putting Eq. (6.14) into Eq. (6.11) and computing the corresponding averages of each term in the expansion (see Appendix D.1), we obtain a state of the form

$$\bar{\sigma}_2 \sim \bar{\sigma}_2^{(0)} + \frac{1}{\sqrt{n}}\bar{\sigma}_2^{(1)} + \frac{1}{n}\bar{\sigma}_2^{(2)} \,, \tag{6.15}$$

where

$$\begin{aligned} \bar{\sigma}_2^{(0)} &= \sum_{k=0}^{\infty} c_k \left|k\middle\rangle\!\middle\langle k\right| \otimes \left|0\middle\rangle\!\middle\langle 0\right| \,, && (6.16) \\ \bar{\sigma}_2^{(1)} &= \sum_{k=0}^{\infty} d_{k+1} \left|k\middle\rangle\!\middle\langle k+1\right| \otimes \left|1\middle\rangle\!\middle\langle 0\right| + \tilde{d}_{k-1} \left|k\middle\rangle\!\middle\langle k-1\right| \otimes \left|0\middle\rangle\!\middle\langle 1\right| \,, && (6.17) \\ \bar{\sigma}_2^{(2)} &= \sum_{k=0}^{\infty} e_k \left|k\middle\rangle\!\middle\langle k\right| \otimes \left(\left|1\middle\rangle\!\middle\langle 1\right| - \left|0\middle\rangle\!\middle\langle 0\right|\right) \\ &\quad + f_{k+2} \left|k\middle\rangle\!\middle\langle k+2\right| \otimes \left|2\middle\rangle\!\middle\langle 0\right| + \tilde{f}_{k-2} \left|k\middle\rangle\!\middle\langle k-2\right| \otimes \left|0\middle\rangle\!\middle\langle 2\right| \,, && (6.18) \end{aligned}$$

and

$$\begin{aligned} d_{k+1} &= c_{k+1}\sqrt{k+1}\,, \quad \tilde{d}_{k-1} = c_k\sqrt{k}\,, \\ e_k &= c_{k+1}(k+1)\,, \\ f_{k+2} &= \frac{1}{\sqrt{2}}c_{k+2}\sqrt{(k+2)(k+1)}\,, \quad \tilde{f}_{k-2} = \frac{1}{\sqrt{2}}c_k\sqrt{k(k-1)}\,. \end{aligned}$$

6.2.1 Computation of the trace norm

We can now use Eqs. (6.12) and (6.15) to compute the trace norm $\left\|\bar{\sigma}_1 - \bar{\sigma}_2\right\|_1$ in the asymptotic regime of large n, up to order $1/n$, by applying perturbation theory. We express the trace norm as

$$\left\|\bar{\sigma}_1 - \bar{\sigma}_2\right\|_1 \sim \left\|A + B/\sqrt{n} + C/n \equiv \Gamma\right\|_1 = \sum_j \left|\gamma_j\right| \,, \tag{6.19}$$

where $A = \bar{\sigma}_1 - \bar{\sigma}_2^{(0)}$, $B = -\bar{\sigma}_2^{(1)}$, $C = -\bar{\sigma}_2^{(2)}$, and γ_j is the jth eigenvalue of Γ, which admits an expansion of the type $\gamma_j = \gamma_j^{(0)} + \gamma_j^{(1)}/\sqrt{n} + \gamma_j^{(2)}/n$. The matrix Γ belongs to the Hilbert space $\mathcal{H}_\infty \otimes \mathcal{H}_3$, i.e., the first mode is described by the infinite dimensional space generated by the Fock basis, and the second mode by the three-dimensional space spanned by the linearly independent vectors $\{\left|-\alpha_0\right\rangle, \left|0\right\rangle, \left|1\right\rangle\}$ (we will see that the contribution of $\left|2\right\rangle$ vanishes, hence it is not necessary to consider a fourth dimension). Writing the eigenvalue equation associated to γ_j and separating the expansion orders, we obtain the set of equations

$$
\begin{aligned}
&A\psi_j^{(0)} = \gamma_j^{(0)}\psi_j^{(0)}\,, && (6.20)\\
&A\psi_j^{(1)} + B\psi_j^{(0)} = \gamma_j^{(0)}\psi_j^{(1)} + \gamma_j^{(1)}\psi_j^{(0)}\,, && (6.21)\\
&A\psi_j^{(2)} + B\psi_j^{(1)} + C\psi_j^{(0)} = \gamma_j^{(0)}\psi_j^{(2)} + \gamma_j^{(1)}\psi_j^{(1)} + \gamma_j^{(2)}\psi_j^{(0)}\,, && (6.22)
\end{aligned}
$$

where $\psi_j = \psi_j^{(0)} + \psi_j^{(1)}/\sqrt{n} + \psi_j^{(2)}/n$ is the eigenvector associated to γ_j. Eq. (6.20) tells us that $\gamma_j^{(0)}$ is an eigenvalue of A with associated eigenvector $\psi_j^{(0)}$. We multiply (6.21) and (6.22) by $\left\langle\psi_j^{(0)}\right|$ to obtain

$$
\begin{aligned}
\gamma_j^{(1)} &= \left\langle\psi_j^{(0)}\right| B \left|\psi_j^{(0)}\right\rangle\,, && (6.23)\\
\gamma_j^{(2)} &= \left\langle\psi_j^{(0)}\right| C \left|\psi_j^{(0)}\right\rangle + \sum_{l\neq j} \frac{\left|\left\langle\psi_j^{(0)}\right| B \left|\psi_l^{(0)}\right\rangle\right|^2}{\gamma_j^{(0)} - \gamma_l^{(0)}}\,. && (6.24)
\end{aligned}
$$

Note that Eq. (6.24) assumes that there is no degeneracy in the spectrum of Γ at zero order (as we will see, this is indeed the case). From the structure of A we can deduce that the form of its eigenvector $\psi_j^{(0)}$ is

$$
\left|\psi_{i,\varepsilon}^{(0)}\right\rangle = |i\rangle \otimes |v_\varepsilon\rangle\,, \tag{6.25}
$$

where we have replaced the index j by the pair of indices i, ε. The index i represents the Fock state $|i\rangle$ in the first mode, and the vectors $|v_\varepsilon\rangle$ are eigenvectors of $|-\alpha_0\rangle\langle-\alpha_0| - |0\rangle\langle 0|$ and form a basis of $\mathcal{H}_3$ in the second mode. Every eigenvalue of Γ is now labeled by the pair of indices i, ε, where $i = 0, \ldots, \infty$ and $\varepsilon = +, -, 0$: the second mode in A has a positive, a negative, and a zero eigenvalue, to which we associate eigenvectors $|v_+\rangle$, $|v_-\rangle$ and $|v_0\rangle$, respectively. It is straightforward to see that the first two are

$$
|v_\pm\rangle = \frac{1}{2}\left(\frac{|-\alpha_0\rangle + |0\rangle}{N_+} \pm \frac{|-\alpha_0\rangle - |0\rangle}{N_-}\right)\,, \tag{6.26}
$$

where $N_\pm = \sqrt{1 \pm e^{-|\alpha_0|^2/2}}$. The zero-order eigenvalues of Γ with $\varepsilon = \pm$ are

$$
\gamma_{i,\pm}^{(0)} = \pm c_i \sqrt{1 - e^{-|\alpha_0|^2}}\,. \tag{6.27}
$$

The third eigenvector $|v_0\rangle$ is orthogonal to the subspace spanned by $|-\alpha_0\rangle$ and $|0\rangle$, and corresponds to the eigenvalue $\gamma_{i,0}^{(0)} = 0$ [6]. This eigenvector only

[6] Note that the zero-order eigenvalues $\gamma_{i,\varepsilon}^{(0)}$ are nondegenerate, hence Eq. (6.24) presents no divergence problems.

plays a role through the overlap $\langle 1|v_0\rangle$, which arises in Eqs. (6.23) and (6.24). We thus do not need its explicit form, but it will suffice to express $\langle 1|v_0\rangle$ in terms of known overlaps.

From (6.23) and (6.25) we readily see that $\gamma^{(1)}_{i,\varepsilon}=0$. Using (6.17), (6.18), (6.24) and (6.25) we can express $\gamma^{(2)}_{i,\varepsilon}$ as

$$\begin{aligned}\gamma^{(2)}_{i,\pm} &= e_i\left(|\langle 0|v_\pm\rangle|^2-|\langle 1|v_\pm\rangle|^2\right)\\ &+ \sum_\varepsilon \frac{d_i|\langle 0|v_\pm\rangle|^2|\langle 1|v_\varepsilon\rangle|^2}{\gamma^{(0)}_{i,\pm}-\gamma^{(0)}_{i-1,\varepsilon}}+\frac{\tilde{d}_i|\langle 1|v_\pm\rangle|^2|\langle 0|v_\varepsilon\rangle|^2}{\gamma^{(0)}_{i,\pm}-\gamma^{(0)}_{i+1,\varepsilon}}\,, \qquad (6.28)\\ \gamma^{(2)}_{i,0} &= 0\,, \qquad (6.29)\end{aligned}$$

where we have used that, by definition, $\langle 0|v_0\rangle=\langle\alpha_0|v_0\rangle=0$. The overlaps in (6.28) are

$$\begin{aligned}|\langle 0|v_\pm\rangle|^2 &= \frac{1}{2}\left(1\mp\sqrt{1-e^{-|\alpha_0|^2}}\right)\,, \qquad (6.30)\\ |\langle 1|v_\pm\rangle|^2 &= \frac{|\alpha_0|^2}{2}\frac{1\pm\sqrt{1-e^{-|\alpha_0|^2}}}{e^{|\alpha_0|^2}-1}\,, \qquad (6.31)\\ |\langle 1|v_0\rangle|^2 &= 1-\frac{|\langle 1|{-\alpha_0}\rangle|^2}{1-|\langle 0|{-\alpha_0}\rangle|^2}=1-\frac{|\alpha_0|^2e^{-|\alpha_0|^2}}{1-e^{-|\alpha_0|^2}}\,. \qquad (6.32)\end{aligned}$$

Now that we have computed the eigenvalues of Γ, comprised in Eqs. (6.27) and (6.28), we are finally in condition to evaluate the sum in the right-hand side of Eq. (6.19). It reads

$$\begin{aligned}\|\Gamma\|_1 &= \sum_{i,\varepsilon}\left|\gamma^{(0)}_{i,\varepsilon}+\frac{1}{n}\gamma^{(2)}_{i,\varepsilon}\right|\\ &= \sum_{i=0}^{\infty}\gamma^{(0)}_{i,+}+\frac{1}{n}\gamma^{(2)}_{i,+}-\gamma^{(0)}_{i,-}-\frac{1}{n}\gamma^{(2)}_{i,-}\\ &= \Lambda^{(0)}_+-\Lambda^{(0)}_-+\frac{1}{n}\left(\Lambda^{(2)}_+-\Lambda^{(2)}_-\right)\,, \qquad (6.33)\end{aligned}$$

where $\Lambda^{(0)}_\pm=\sum_{i=0}^\infty\gamma^{(0)}_{i,\pm}=\pm\sqrt{1-e^{-|\alpha_0|^2}}$ (recall that $\sum_{i=0}^\infty c_i=1$), and $\Lambda^{(2)}_\pm=\sum_{i=0}^\infty\gamma^{(2)}_{i,\pm}$. The analytic expression of $\Lambda^{(2)}_\pm$ is computable, but rather involved. The asymptotic form of the average minimum-error probability $P_e^{\mathrm{opt}}(n)$, defined in Eq. (6.8), hence becomes

$$P_e^{\mathrm{opt}}\equiv P_e^{\mathrm{opt}}(n\to\infty)\sim\frac{1}{2}\left[1-\sqrt{1-e^{-|\alpha_0|^2}}-\frac{1}{2n}\left(\Lambda^{(2)}_+-\Lambda^{(2)}_-\right)\right]\,. \qquad (6.34)$$

6.2.2 Excess risk

The figure of merit that we use to assess the performance of our protocol is the *excess risk*, defined as the difference between the asymptotic average error probability $P_e^{\rm opt}$ and the average error probability for the optimal strategy when α is perfectly known. As we said at the beginning of the section, the true value of α is $\alpha_0 + u/\sqrt{n}$ for a particular realization, thus knowing u equates to knowing α. The minimum-error probability for the discrimination between the *known* states $|0\rangle$ and $|\alpha_0 + u/\sqrt{n}\rangle$, $P_e^*(u,n)$, averaged over the Gaussian distribution $G(u)$, takes the form

$$\begin{aligned} P_e^*(n) &= \int G(u) P_e^*(u,n) d^2u \\ &= \int G(u) \frac{1}{2}\left(1-\sqrt{1-|\langle 0|\alpha_0+u/\sqrt{n}\rangle|^2}\right) d^2u \end{aligned} \qquad (6.35)$$

To compute this integral we do a series expansion of the overlap in the limit $n \to \infty$ and we use Eqs. (D.10), (D.11), and (D.12). After some algebra we obtain

$$P_e^* \equiv P_e^*(n\to\infty) \sim \frac{1}{2}\left(1-\sqrt{1-e^{-|\alpha_0|^2}}+\frac{1}{n}\Lambda^*\right), \qquad (6.36)$$

where

$$\Lambda^* = \frac{\mu^2\left[2\left(e^{-|\alpha_0|^2}-1\right)+|\alpha_0|^2\left(2-e^{-|\alpha_0|^2}\right)\right]}{4\left(e^{|\alpha_0|^2}-1\right)\sqrt{1-e^{-|\alpha_0|^2}}}. \qquad (6.37)$$

The excess risk is then defined through Eqs. (6.34) and (6.36) as

$$R^{\rm opt} = n\left(P_e^{\rm opt} - P_e^*\right) \qquad (6.38)$$

Finally, taking the limit $\mu \to \infty$ [recall that μ is a free parameter introduced to define the width of the Gaussian distribution (6.5)] we obtain

$$R^{\rm opt} = \frac{|\alpha_0|^2 e^{-|\alpha_0|^2/2}\left(2e^{|\alpha_0|^2}-1\right)}{16\left(e^{|\alpha_0|^2}-1\right)^{3/2}}. \qquad (6.39)$$

Note that the excess risk only depends on the module of α_0, i.e., on the average distance between $|\alpha\rangle$ and $|0\rangle$. The excess risk is thus phase-invariant, as it should.

6.3 Local strategy

An alternative—and more restrictive—strategy to determine the state of the signal consists in the natural combination of two fundamental tasks: state estimation, and state discrimination of known states. In such an "estimate-and-discriminate" (E&D) strategy, *all* ancillary modes are used to better estimate the unknown amplitude α. Then, the obtained information is used to tune a discrimination measurement over the signal that distinguishes the vacuum state from a coherent state with the estimated amplitude. In this Section we find the optimal E&D strategy based on Gaussian measurements and compute its excess risk $R^{\text{E\&D}}$. Then, we compare the result with that of the optimal collective strategy R^{opt}.

The most general Gaussian measurement that one can use to estimate the state of the ancillary mode $|\sqrt{n}\alpha\rangle$ is a *generalized heterodyne measurement* (see Appendix C.3.2), represented by a POVM with elements

$$E_{\bar{\beta}} = \frac{1}{\pi} \left|\bar{\beta}, r, \phi\rangle\langle\bar{\beta}, r, \phi\right| , \tag{6.40}$$

i.e., projectors onto pure Gaussian states with amplitude $\bar{\beta}$ and squeezing r along the direction ϕ. The outcome of such heterodyne measurement $\bar{\beta} = \sqrt{n}\beta$ is an estimate of $\sqrt{n}\alpha$, hence β stands for an estimate of α. Upon obtaining $\bar{\beta}$, the prior information that we have about α gets updated according to Bayes' rule, so that now the signal state can be either $|0\rangle\langle 0|$ or some state $\rho(\beta)$. The form of this second hypothesis is given by

$$\rho(\beta) = \int p(\alpha|\beta)\, |\alpha\rangle\langle\alpha|\, d^2\alpha\,, \tag{6.41}$$

where $p(\alpha|\beta)$ encodes the *posterior* information that we have acquired via the heterodyne measurement. It represents the conditional probability of the state of the ancillary mode being $|\sqrt{n}\alpha\rangle$, given that we obtained the estimate $\bar{\beta}$. Bayes' rule dictates

$$p(\alpha|\beta) = \frac{p(\beta|\alpha)p(\alpha)}{p(\beta)}\,, \tag{6.42}$$

where $p(\beta|\alpha)$ is given by (see Appendix D.2)

$$p(\beta|\alpha) = \frac{1}{\pi\cosh r} e^{-|\sqrt{n}\alpha-\bar{\beta}|^2 - \text{Re}[(\sqrt{n}\alpha-\bar{\beta})^2 e^{-i2\phi}]\tanh r}\,, \tag{6.43}$$

$p(\alpha)$ is the prior information that we have about α before the heterodyne measurement and

$$p(\beta) = \int p(\alpha)p(\beta|\alpha)d^2\alpha \tag{6.44}$$

is the total probability of obtaining the estimate β.

The error probability of the E&D strategy, averaged over all possible estimates β, is then

$$P_e^{\text{E\&D}}(n) = \frac{1}{2}\left(1 - \frac{1}{2}\int p(\beta)\, \| |0\rangle\langle 0| - \rho(\beta) \|_1 \, d^2\beta\right) . \tag{6.45}$$

Note that the estimate β depends ultimately on the number n of ancillary modes, hence the explicit dependence in the left-hand side of Eq. (6.45).

We are interested in the asymptotic expression of Eq. (6.45) when n is large, so let us now move to the asymptotic scenario $n \to \infty$. Recall that an initial rough estimation of α permits the localisation of the prior $p(\alpha)$ around a central point α_0, such that $\alpha \approx \alpha_0 + u/\sqrt{n}$, where u is distributed according to $G(u)$, defined in Eq. (6.5). Consequently, the estimate β will also be localised around the same point, i.e., $\beta \approx \alpha_0 + v/\sqrt{n}$, $v \in \mathbb{C}$. As a result, we can effectively shift from amplitudes α and β to a local Gaussian model around α_0, parametrized by u and v. According to this new model, we make the following transformations:

$$\begin{aligned}
p(\alpha) \quad &\to \quad G(u)\,, && (6.46)\\
p(\beta|\alpha) \quad &\to \quad p(v|u) = \frac{1}{\pi \cosh r} e^{-|u-v|^2 - \mathrm{Re}[(u-v)^2]\tanh r}\,, && (6.47)\\
p(\beta) \quad &\to \quad p(v) = \int p(v|u) G(u) du = \frac{1}{\pi \cosh r} \frac{1}{\sqrt{1 + \mu^2\left(2 + \frac{\mu^2}{\cosh^2 r}\right)}}\\
&\qquad \times \exp\left(\frac{|v|^2\left(1 + \frac{\mu^2}{\cosh^2 r}\right) + \mathrm{Re}[v^2]\tanh r}{\mu^4 \tanh^2 r - \left(\mu^2 + 1\right)^2}\right)\,, && (6.48)\\
p(\alpha|\beta) \quad &\to \quad p(u|v) = \frac{p(v|u)G(u)}{p(v)}\,, && (6.49)
\end{aligned}$$

where, for simplicity, we have assumed α_0 to be real. Note that this can be done without loss of generality. Note also that, by the symmetry of the problem, this assumption implies $\phi = 0$.

6.3.1 Computation of the trace norm

The shifting to the local model renders the trace norm in Eq. (6.45)

$$\| |0\rangle\langle 0| - \rho(\beta) \|_1 \quad \to \quad \| |-\alpha_0\rangle\langle -\alpha_0| - \rho(v) \|_1 \,, \tag{6.50}$$

where

$$\rho(v) = \int p(u|v) \left| u/\sqrt{n}\rangle\langle u/\sqrt{n} \right| d^2 u\,. \tag{6.51}$$

To compute the explicit expression of $\rho(v)$ we proceed as in the collective strategy. That is, we expand $|u/\sqrt{n}\rangle\langle u/\sqrt{n}|$ in the limit $n \to \infty$ up to order $1/n$, as in Eq. (6.14). We name the appearing integrals of $u, u^*, |u|^2, u^2$, and $(u^*)^2$ over the probability distribution $p(u|v)$ as I_1, I_1^*, I_2, I_3, and I_3^*, respectively. This allows us to write the trace norm (6.50) as

$$\left\| |-\alpha_0\rangle\langle-\alpha_0| - \rho(v) \right\|_1 \sim \left\| A' + B'/\sqrt{n} + C'/n \equiv \Phi \right\|_1 = \sum_\kappa |\lambda_\kappa| \,, \qquad (6.52)$$

where

$$\begin{aligned} A' &= |-\alpha_0\rangle\langle-\alpha_0| - |0\rangle\langle 0| \,, && (6.53)\\ B' &= -I_1 \, |1\rangle\langle 0| - I_1^* \, |0\rangle\langle 1| \,, && (6.54)\\ C' &= -I_2 \left(|1\rangle\langle 1| - |0\rangle\langle 0| \right) - \frac{1}{\sqrt{2}} \left(I_3 \, |2\rangle\langle 0| + I_3^* \, |0\rangle\langle 2| \right) \,, && (6.55) \end{aligned}$$

and λ_κ is the κth eigenvalue of Φ, which admits the perturbative expansion $\lambda_\kappa = \lambda_\kappa^{(0)} + \lambda_\kappa^{(1)}/\sqrt{n} + \lambda_\kappa^{(2)}/n$, just as its associated eigenvector $\varphi_\kappa = \varphi_\kappa^{(0)} + \varphi_\kappa^{(1)}/\sqrt{n} + \varphi_\kappa^{(2)}/n$. Up to order $1/n$, the matrix Φ has effective dimension 4 since it belongs to the space spanned by the set of linearly independent vectors $\{|-\alpha_0\rangle, |0\rangle, |1\rangle, |2\rangle\}$. Hence the index κ has in this case four possible values, i.e., $\kappa = +, -, 3, 4$. The zero-order eigenvalues $\lambda_\kappa^{(0)}$, which correspond to the eigenvalues of the rank-2 matrix A', are

$$\lambda_\pm^{(0)} = \pm\sqrt{1 - e^{-\alpha_0^2}} \,, \quad \lambda_3^{(0)} = \lambda_4^{(0)} = 0 \qquad (6.56)$$

(recall that $\alpha_0 \in \mathbb{R}$). Their associated eigenvectors are $\left|\varphi_\kappa^{(0)}\right\rangle = |v_\kappa\rangle$, where $|v_\pm\rangle$ is given by Eq. (6.26), and, by definition, $\langle v_\kappa|-\alpha_0\rangle = \langle v_\kappa|0\rangle = 0$ for $\kappa = 3, 4$. From analogous expressions to Eqs. (6.23) and (6.24) we can write the first and second-order eigenvalues as

$$\begin{aligned} \lambda_\kappa^{(1)} &= -I_1 \langle v_\kappa|1\rangle\langle 0|v_\kappa\rangle - I_1^* \langle v_\kappa|0\rangle\langle 1|v_\kappa\rangle \,, && (6.57)\\ \lambda_\kappa^{(2)} &= I_2 \left(|\langle v_\kappa|0\rangle|^2 - |\langle v_\kappa|1\rangle|^2 \right) - \frac{1}{\sqrt{2}} \left(I_3 \langle v_\kappa|2\rangle\langle 0|v_\kappa\rangle + I_3^* \langle v_\kappa|0\rangle\langle 2|v_\kappa\rangle \right) \\ &+ \sum_{\xi\neq\kappa} \left(|I_1|^2 \frac{|\langle v_\xi|1\rangle|^2 |\langle v_\kappa|0\rangle|^2 + |\langle v_\xi|0\rangle|^2 |\langle v_\kappa|1\rangle|^2}{\lambda_\kappa^{(0)} - \lambda_\xi^{(0)}} \right. \\ &+ \left. \frac{I_1^2 \langle v_\xi|1\rangle\langle v_\kappa|1\rangle\langle 0|v_\kappa\rangle\langle 0|v_\xi\rangle + (I_1^*)^2 \langle 1|v_\xi\rangle\langle 1|v_\kappa\rangle\langle v_\kappa|0\rangle\langle v_\xi|0\rangle}{\lambda_\kappa^{(0)} - \lambda_\xi^{(0)}} \right) . && (6.58) \end{aligned}$$

The needed overlaps for computing $\lambda_\kappa^{(1)}$ and $\lambda_\kappa^{(2)}$ are given by Eqs. (6.30), (6.31), and

$$\langle v_\pm|0\rangle = \frac{1}{2}\left(N_+ \mp N_-\right) , \tag{6.59}$$

$$\langle v_\pm|1\rangle = \frac{1}{2}(-\alpha_0)e^{-\alpha_0^2/2}\left(\frac{1}{N_+} \pm \frac{1}{N_-}\right) , \tag{6.60}$$

$$|\langle v_3|1\rangle|^2 = 1 - \frac{|\langle 1|{-\alpha_0}\rangle|^2}{1-|\langle 0|{-\alpha_0}\rangle|^2 - |\langle 2|{-\alpha_0}\rangle|^2} , \tag{6.61}$$

$$|\langle v_4|1\rangle|^2 = \frac{|\langle 1|{-\alpha_0}\rangle|^2|\langle 2|{-\alpha_0}\rangle|^2}{\left(1-|\langle 0|{-\alpha_0}\rangle|^2\right)\left(1-|\langle 0|{-\alpha_0}\rangle|^2 - |\langle 2|{-\alpha_0}\rangle|^2\right)} . \tag{6.62}$$

The expressions for the overlaps (6.61) and (6.62) actually depend on the dimension of the space that we are considering (four in this case), and they are not unique: there are infinitely many possible orientations of the orthogonal pair of vectors $\{|v_3\rangle, |v_4\rangle\}$ such that both of them are orthogonal to the plane formed by $\{|{-\alpha_0}\rangle, |0\rangle\}$, which is the only requirement we have. Note, however, that this degeneracy has no effect on the excess risk, thus we are free to choose the particular orientation that, in addition, verifies $\langle v_3|2\rangle = 0$, yielding the simple expressions (6.61) and (6.62).

Finally, we write down the trace norm as

$$\begin{aligned} \|\Phi\|_1 &= \sum_\kappa |\lambda_\kappa^{(0)} + \frac{1}{\sqrt{n}}\lambda_\kappa^{(1)} + \frac{1}{n}\lambda_\kappa^{(2)}| \\ &= \lambda_+^{(0)} - \lambda_-^{(0)} + \frac{1}{\sqrt{n}}\left(\lambda_+^{(1)} - \lambda_-^{(1)}\right) \\ &\quad + \frac{1}{n}\left(\lambda_+^{(2)} - \lambda_-^{(2)} + |\lambda_3^{(2)}| + |\lambda_4^{(2)}|\right) , \end{aligned} \tag{6.63}$$

which we use now to obtain the asymptotic expression for the average error probability, defined in Eq. (6.45). Recall Eq. (6.48) and note that we have to average Eq. (6.63) over the probability distribution $p(v)$. Regarding this average, it is worth taking into account the following considerations. First, the v-dependence of the eigenvalues comes from I_1, I_2, I_3, and its complex conjugates. The integrals needed are calculated in Appendix D.3. Second, because of Eq. (D.18), $\lambda_\kappa^{(1)} = 0$ and hence the order $1/\sqrt{n}$ term vanishes, as it should. And third, the second-order eigenvalues $\lambda_3^{(2)}$ and $\lambda_4^{(2)}$ are v-independent and positive, so we can ignore the absolute values in Eq. (6.63). Putting all together, we can express the asymptotic average error probability of the E&D strategy as

$$P_e^{\text{E\&D}} \equiv P_e^{\text{E\&D}}(n\to\infty) \sim \frac{1}{2}\left(1 - \sqrt{1-e^{-\alpha_0^2}} + \frac{1}{n}\Delta^{\text{E\&D}}\right) , \tag{6.64}$$

where

$$\Delta^{\mathrm{E\&D}} = -\frac{1}{2}\left(\lambda_3^{(2)} + \lambda_4^{(2)} + \int p(v)(\lambda_+^{(2)} - \lambda_-^{(2)})dv\right) . \tag{6.65}$$

6.3.2 Excess risk

The excess risk associated to the E&D strategy is generically expressed as

$$R^{\mathrm{E\&D}}(r) = n \lim_{\mu\to\infty} \left(P_e^{\mathrm{E\&D}} - P_e^*\right) , \tag{6.66}$$

where P_e^* is the error probability for known α, given in Eq. (6.36), and $P_e^{\mathrm{E\&D}}$ is the result from the previous section, i.e., Eq. (6.64). Although $R^{\mathrm{E\&D}}(r)$ has a closed analytical expression, we do not show it here since its form is not very illuminating. Note that we have to take the limit $\mu \to \infty$ in the excess risk, as we did for the collective case. Note also that all the expressions calculated so far explicitly depend on the squeezing parameter r (apart from α_0). This parameter stands for the squeezing of the generalized heterodyne measurement in Eq. (6.40), which we have left unfixed on purpose. As a result, we now define, through the squeezing r, the optimal heterodyne measurement over the ancillary mode to be that which yields the lowest excess risk (6.66), i.e.,

$$R^{\mathrm{E\&D}} = \min_r R^{\mathrm{E\&D}}(r) . \tag{6.67}$$

To find the optimal r, we look at the parameter estimation theory of Gaussian models (see, e.g., [Gill and Guţă, 2013]). In a generic two-dimensional Gaussian shift model, the optimal measurement for the estimation of a parameter $\theta = (q,p)$ is a generalized heterodyne measurement[7] of the type (6.40). Such measurement yields a quadratic risk of the form

$$R_{\hat\theta} = \int p(\theta)((\hat\theta - \theta)^T G(\hat\theta - \theta))d^2\theta , \tag{6.68}$$

where $p(\theta)$ is some probability distribution, $\hat\theta$ is an estimator of θ, and G is a two-dimensional matrix. One can always switch to the coordinates system in which $G = \mathrm{diag}(g_q, g_p)$ is diagonal to write

$$R_{\hat\theta} = g_q \int p(\theta)(\hat q - q)^2 d^2\theta + g_p \int p(\theta)(\hat p - p)^2 d^2\theta . \tag{6.69}$$

It can be shown [Gill and Guţă, 2013] that the optimal squeezing of the estimation measurement, i.e., that for which the quadratic risk $R_{\hat\theta}$ is minimal,

[7] This is the case whenever the covariance of the Gaussian model is known, and the mean is a linear transformation of the unknown parameter.

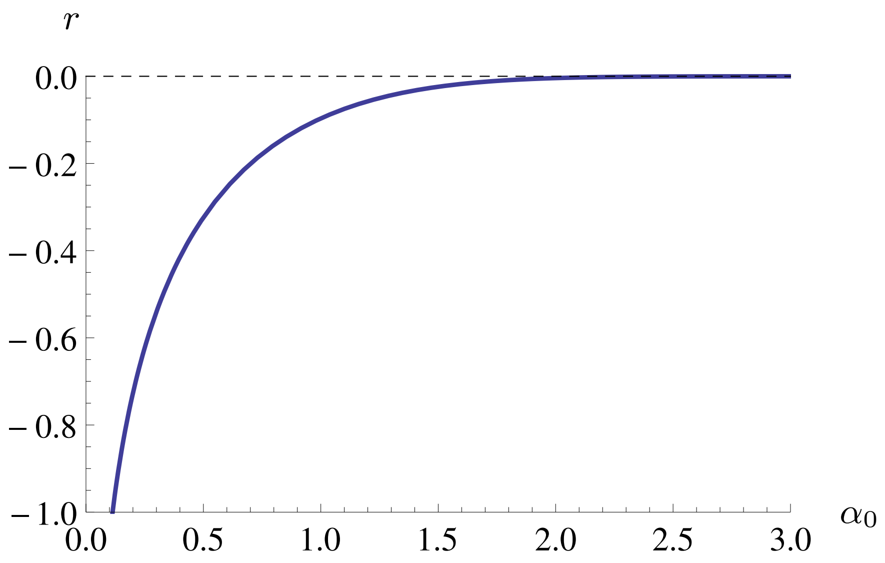


Figure 6.2. Optimal squeezing r for the generalized heterodyne measurement in a E&D strategy, as a function of α_0.

is given by

$$r = \frac{1}{4} \ln \left(\frac{g_q}{g_p} \right) . \tag{6.70}$$

We can then simply compare Eq. (6.69) with Eq. (6.66) to deduce the values of g_q and g_p for our case. By doing so, we obtain that the optimal squeezing reads

$$r = \frac{1}{4} \ln \left(\frac{f(\alpha_0) + \alpha_0^2}{f(\alpha_0) - \alpha_0^2} \right) , \tag{6.71}$$

where

$$f(\alpha_0) = 2e^{\alpha_0^2} \left(e^{\alpha_0^2} - 1 \right) \left(\sqrt{1 - e^{-\alpha_0^2}} - 1 \right) + \alpha_0^2 \left(1 - 2e^{\alpha_0^2} \sqrt{1 - e^{-\alpha_0^2}} \right) . \tag{6.72}$$

Eq. (6.71) tells us that the optimal squeezing r is a function of α_0 that takes negative values, and asymptotically approaches zero when α_0 is large (see Fig. 6.2). This means that the optimal estimation measurement over the ancillary mode is comprised by projectors onto coherent states antisqueezed along the line between α_0 and the origin (which represents the vacuum) in phase space. In other words, the estimation is tailored to have better resolution along that axis because of the subsequent discrimination of the signal state. This makes sense: since the error probability in the discrimination depends primarily on the distance between the hypotheses, it is more important

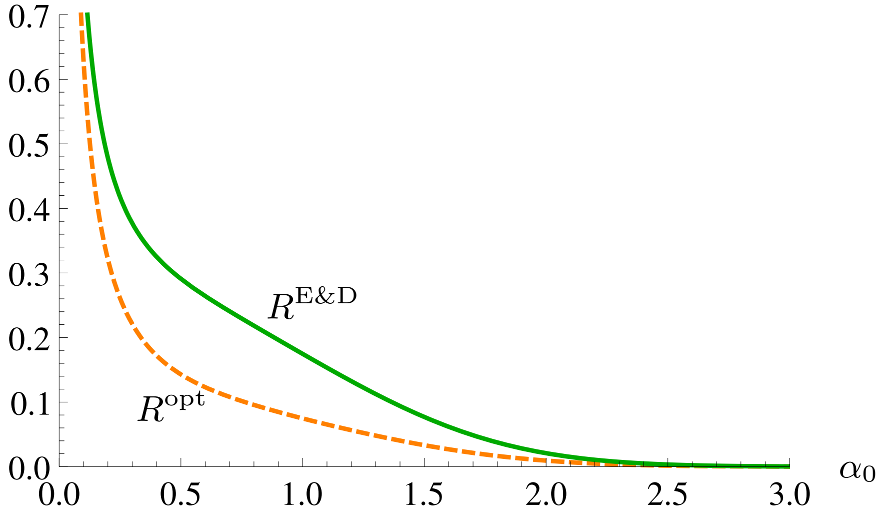


Figure 6.3. Excess risk for the collective strategy, R^{opt}, and for the local strategy, $R^{\text{E\&D}}$, as a function of α_0.

to estimate this distance more accurately rather than along the orthogonal direction. For large amplitudes, the estimation converges to a (standard) heterodyne measurement with no squeezing. As α_0 approaches 0 the states of the signal become more and more indistinguishable, and the projectors of the heterodyne measurement approach infinitely squeezed coherent states, thus converging to a homodyne measurement.

Inserting Eq. (6.71) into Eq. (6.67) we finally obtain the expression of $R^{\text{E\&D}}$ as a function of α_0, which we can now compare with the excess risk for the collective strategy R^{opt}, given in Eq. (6.39). We plot both functions in Fig. 6.3. For small amplitudes in the range $\alpha_0 \sim (0.3 - 1.5)$ there is a noticeable difference in the performance of the two strategies, reaching more than a factor two at some points. We also observe that the gap closes for large amplitudes. This behaviour is expected, since the problem becomes classical when the energy of the signal is sufficiently large. Very weak energies also render the strategies almost equivalent.

6.4 Completely general estimation measurements

We have showed that a local strategy based on the estimation of the ancillary state via a generalized heterodyne measurement, followed by the corresponding discrimination measurement on the signal mode, performs worse than the most general (collective) strategy. However, this procedure does not encompass *all* local strategies. The heterodyne measurement, although with some nonzero squeezing, still observes the phase space around α_0 in a Gaussian way. A more general measurement that produces a non-Gaussian probability distribution for the estimate β might perform better in terms of the excess risk.

The intuition behind this is the following. Imagine that we restrict α further to be on the real axis. Then, the true α is either to the left of α_0 or to the right, depending on the sign of the local parameter u. In the former case, α is closer to the vacuum, so the error in discriminating between them is larger than for the states on the other side. One would then expect that it is desirable to estimate better the negative parameters u, compared to the positive ones. Gaussian measurements like the heterodyne do not contemplate this situation, as they are translationally invariant, and that might be the reason behind the gap in Fig. 6.3.

To test this intuition we design the following simple example. Since the required methods are a straightforward extension of the ones used in the previous sections, we only sketch the procedure without showing any explicit calculation. Imagine now that the true value of α is not Gaussian distributed around α_0, but it can only take the values $\alpha = \alpha_0 \pm 1/\sqrt{n}$, representing the states that are closer to the vacuum and further away. Having only two possibilities for α allows us to solve analytically the most general local strategy, since estimating the ancillary state becomes a discrimination problem between the states $|\sqrt{n}\alpha_0 + 1\rangle$ and $|\sqrt{n}\alpha_0 - 1\rangle$. The measurement that distinguishes the two possibilities is a two-outcome POVM $\mathcal{E} = \{|e_+\rangle\langle e_+|\,, |e_-\rangle\langle e_-|\}$[8]. We use the displacement operator (6.9) to shift to the local model around α_0, such that the state of the ancillary mode is now either $|1\rangle$ or $|-1\rangle$. Then, the probabilities of correctly identifying each state are

$$p_+ = |\langle e_+|1\rangle|^2 \equiv c^2 \quad \text{and} \quad p_- = |\langle e_-|-1\rangle|^2 = 1 - c^2\,. \tag{6.73}$$

Since the vectors $|e_+\rangle$ and $|e_-\rangle$ are orthogonal by definition, the only freedom in choosing the POVM $\mathcal{E}$ is its relative orientation with respect to the pair of vectors $|1\rangle$ and $|-1\rangle$, which is parametrized by the overlap c. Then, if our intuition is right, it should happen that $c < 1/2$, i.e., that the probability of a correct identification is greater for the state $|-1\rangle$ than for $|1\rangle$.

From now on we proceed as for the E&D strategy: we first compute the posterior state of the signal mode according to Bayes' rule, and then we compute the optimal error probability in the discrimination of $|-\alpha_0\rangle\langle-\alpha_0|$ and the

[8] Note that, by choosing the POVM elements to be rank-1 projectors, we rule out higher-rank POVMs. However, it is easy to show by the convexity properties of the trace norm that the optimal POVM, in terms of the error probability, has rank-1 elements.

posterior state, which is a combination of $|1/\sqrt{n}\rangle\langle 1/\sqrt{n}|$ and $|-1/\sqrt{n}\rangle\langle -1/\sqrt{n}|$, weighted by the corresponding posterior probabilities. The c-dependence is carried by these probabilities. Going to the asymptotic limit $n \to \infty$, applying perturbation theory for computing the trace norm, and averaging the result over the two possible outcomes in the discrimination of the signal state, we finally obtain the asymptotic average error probability for the local strategy as a function of c. The asymptotic average error probability for the optimal collective strategy in this simple case is obtained exactly along the same lines as shown in Section 6.2, and the one for *known* states is given by the asymptotic expansion of Eq. (6.35), substituting the average over $G(u)$ appropriately.

Now we can compute the excess risk for the local and collective strategy, and optimize the local one over c. Surprisingly, the optimal solution yields $c = 1/2$, i.e., the POVM $\mathcal{E}$ is symmetric with respect to the vectors $|1\rangle$ and $|-1\rangle$, hence both hypotheses receive the same treatment by the measurement in charge of determining the state of the ancillary mode. Moreover, the gap between the excess risk of both strategies remains. This result impels us to conjecture that the optimal collective strategy performs better than *any* local strategy.

6.5 Discussion

In this Chapter we have proposed a learning scheme for coherent states of light, similar to the one proposed for qubits in Chapter 5. We have presented it in the context of a quantum-enhanced readout of classically-stored binary information, following a recent research line initiated in [Pirandola, 2011]. The reading of information, encoded in the state of a signal that comes reflected by a memory cell, is achieved by measuring the signal and deciding its state to be either the vacuum state or some coherent state of *unknown* amplitude. The effect of this uncertainty is palliated by supplying a large number of ancillary modes in the same coherent state. We present two strategies that make different uses of this (quantum) side information to determine the state of the signal: a collective strategy, consisting in measuring all modes at once and making the binary decision, and a local strategy, based on first estimating—*learning*—the unknown amplitude, then using the acquired knowledge to tune a discrimination measurement over the signal. We show that the former outperforms any local strategy that uses a Gaussian estimation measurement over the ancillary modes. Furthermore, we conjecture that this is indeed the case for *any* local strategy, on the light of a simplification of the original setting that allows us to consider completely general measurements.

Previous works on quantum reading rely on the use of specific preparations of nonclassical—entangled—states of light to improve the reading performance of a classical memory [Pirandola, 2011; Nair, 2011; Spedalieri *et al.*, 2012; Tej *et al.*, 2013]. Our results indicate that, when there exists some uncertainty in the states produced by the source (and, consequently, the possibility of preparing a specific entangled signal state is highly diminished), quantum resources (collective measurements) still enhance the reading of classical information using classicaly correlated light. It is worth mentioning that there are precedents of classically correlated coherent states exhibiting quantum phenomena of this sort. As an example, in the context of estimation of product coherent states, the optimal measure-and-prepare strategy on identical copies of $|\alpha\rangle$ can be achieved by LOCC (according to the fidelity criterion), but bipartite product states $|\alpha\rangle|\alpha^*\rangle$ require entangled measures [Niset *et al.*, 2007].

On a final note, the quantum enhancement found here is relevant on the regime of low energy signals[9] (small amplitudes). This is in accordance to the advantage regime provided by nonclassical light sources, as discussed in other works. A low energy readout of memories is, in fact, of very practical interest. While—mathematically—the success probability of any readout protocol could be arbitrarily increased by sending signals with infinite energy, there are many situations where this is highly discouraged. For instance, the readout of photosensitive organic memories requires a high level of control over the amount of energy irradiated per cell. In those situations, the use of signals with very low energy benefits from quantum-enhanced performance, whereas highly energetic classical light could easily damage the memory.

[9] Note that here we have only considered sending a single-mode signal. However, in what coherent states are concerned, increasing the number of modes of the signal and increasing the energy of a single mode are equivalent situations.

CHAPTER 7

Decomposition of quantum measurements

The growth of quantum information theory and, in particular, the development of a vast variety of quantum processing techniques in the past few decades has drawn major attention towards the measurement process in quantum mechanics. Because no complete knowledge of the state of a quantum system can be retrieved from a single measurement, in general there are different incompatible measurement strategies that may yield very different results when applied to the same scenario. Hence, most often the design of a quantum processing technique involves finding which measurement best accomplishes a specific task, or which sequence of measurements is statistically optimal. These problems are the keystone of quantum estimation theory [Helstrom, 1976], and its solutions stand as a characteristic feature of many quantum processing tasks.

Recent advances in experimental techniques have rendered many of these tasks realizable in a laboratory, where a minimum resource perspective prevails. The sought for the minimum resources needed to implement a certain task has a paradigmatic example in quantum state preparation: to prepare all pure states of a bipartite system, it is enough to prepare only one maximally entangled pure state; then, by means of local operations and classical communication, one can obtain any bipartite pure state [Nielsen and Chuang, 2000]. The mathematical object that represents a general quantum measurement is a POVM (see Section 2.3), and therefore these kind of questions concern to the mathematical structure of POVMs. The aim of this Chapter is to address the following minimum resource problem: given a certain POVM, what are the simplest resources needed, and how one can implement it in terms of them?

POVMs form a convex set. This means that, given two known POVMs, any randomized implementation of them is also a POVM: just as mixed states are probabilistic mixtures of pure states, one can talk about measurements that can be regarded as probabilistic mixtures of POVMs. Those that cannot be expressed as combinations of other measurements are called extremal POVMs. Since many measurement optimization problems consist in maximizing a convex figure of merit, which leads to an extremal solution, this type of POVM appears quite frequently. It is no wonder then that the characterization of extremal POVMs has been extensively addressed in the literature[1].

It is clear that the set of all extremal POVMs comprise the toolbox needed to effectively implement any measurement, as an appropriate convex combination of extremal POVMs will reproduce its statistics. A number of works have been devoted to prove the existence of such decompositions of measurements into extremals for finite [D'Ariano *et al.*, 2005; Haapasalo *et al.*, 2011] as well as infinite dimensional systems [Chiribella *et al.*, 2007]. However, the question of which are the minimal resources needed to implement a given POVM remains unclear from an operational point of view. In this Chapter we provide a clear answer to this question by designing a constructive and efficient algorithm that takes as input any POVM with an arbitrary (but finite) number of outcomes and gives as output a convex combination of extremal POVMs that reproduces its statistics. We show that only rank-1 extremal POVMs are needed if one allows for a classical post-processing of the outcomes (in agreement to a similar result shown in [Haapasalo *et al.*, 2011]). The number of extremals that this algorithm produces is upper bounded by $(N-1)d+1$, where N is the number of outcomes of the input POVM and d is the dimension of its associated Hilbert space. This bound is significantly lower than the best previously known upper bound [D'Ariano *et al.*, 2005], which scaled as d^2. As a byproduct of our analysis, we obtain a simple geometrical characterization of extremal POVMs in terms of the generalized Bloch vectors associated to their elements.

In Section 7.1 we fix the notation and illustrate how the algorithm works in a few simple cases. In Section 7.2 we set the mathematical tools we rely on and we derive from them a geometrical characterization of extremal POVMs. Section 7.3 is devoted to the full description of the algorithm, and Section 7.4 to the discussion of further improvements. We finally summarize our results.

[1] See, e.g., [D'Ariano *et al.*, 2005; Chiribella *et al.*, 2010; Pellonpää, 2011; Heinosaari and Pellonpää, 2012].

7.1 Simple cases

Let us start by fixing the notation and conventions used throughout this Chapter. A POVM is a set $\mathbb{P} = \{E_i\}$ of positive semidefinite operators acting on a Hilbert space $\mathcal{H}$ of dimension d, which satisfy the normalization condition $\sum_i E_i = \mathbb{I}$. The operator E_i is called a *POVM element*, and it is associated to the outcome i of the POVM. In this Chapter we focus on POVMs with a finite number of outcomes. The elements E_i might be zero for some i, meaning that the corresponding outcomes have zero probability of occurrence. Two POVMs that differ only in the number or position of their zero elements are considered to be physically equivalent. When characterizing a POVM by its number of outcomes we will refer only to those with physical meaning, that is to the outcomes with a non-zero operator associated. In this spirit, we denote by $\mathbb{P}_N$ a POVM $\mathbb{P}$ with N non-zero elements, and we will refer to it as a N-outcome POVM.

A convex combination of two POVMs is also a POVM: suppose that $\mathbb{P}_3^{(1)} = \{E_1, E_2, E_3, 0, 0\}$ and $\mathbb{P}_3^{(2)} = \{0, 0, E_3, E_4, E_5\}$ are two 3-outcome POVMs, then $\mathbb{P}_5 \equiv p_1\mathbb{P}_3^{(1)} + p_2\mathbb{P}_3^{(2)} = \{p_1E_1, p_1E_2, (p_1 + p_2)E_3, p_2E_4, p_2E_5\}$ is also a POVM, where $p_1 + p_2 = 1$. The convex combination $\mathbb{P}_5$ is the weighted sum element-by-element of $\mathbb{P}_3^{(1)}$ and $\mathbb{P}_3^{(2)}$.

In this Chapter we are faced with the reverse situation: given a POVM, we want to find a decomposition into a convex combination of smaller (i.e. with less outcomes) POVMs. As a simple example of this type of decomposition, consider the POVM needed in the eavesdropping of the "BB84" protocol [Nielsen and Chuang, 2000]

$$\mathbb{P}_4 = \left\{ \frac{1}{2} |0\rangle\langle 0| , \frac{1}{2} |1\rangle\langle 1| , \frac{1}{2} |+\rangle\langle +| , \frac{1}{2} |-\rangle\langle -| \right\} . \tag{7.1}$$

Note that $\mathbb{P}_4$ can be expressed as

$$\mathbb{P}_4 = \frac{1}{2}\mathbb{P}_2^{(z)} + \frac{1}{2}\mathbb{P}_2^{(x)} , \tag{7.2}$$

where

$$\mathbb{P}_2^{(z)} = \{|0\rangle\langle 0| , |1\rangle\langle 1| , 0, 0\} \tag{7.3}$$

$$\mathbb{P}_2^{(x)} = \{0, 0, |+\rangle\langle +| , |-\rangle\langle -|\} . \tag{7.4}$$

Thus, the POVM $\mathbb{P}_4$ can be effectively implemented by tossing an unbiased coin, and then performing either $\mathbb{P}_2^{(x)}$ or $\mathbb{P}_2^{(z)}$ based on the outcome of this toss. In this case it is trivial to identify at sight the two pairs of orthogonal operators and their weights in the decomposition. This will not be so for an

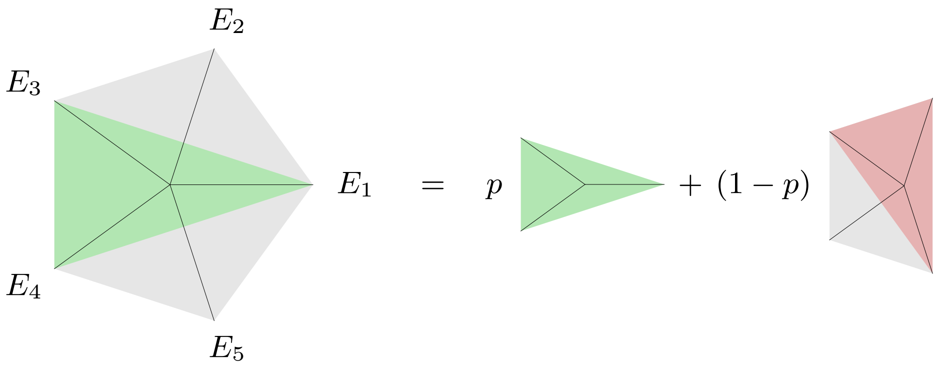


Figure 7.1. First step of the decomposition of $\mathbb{P}_5$. The selection of elements (green) form the trine $\mathbb{P}_3^{(1)}$ which appears in the decomposition with associated probability p. After extracting it, we are left with $\mathbb{P}_4^{(\mathrm{aux})}$ with associated probability $(1-p)$. In the second step we select another trine (red) from $\mathbb{P}_4^{(\mathrm{aux})}$.

arbitrary measurement. The next example is presented to gain insight on how this operation can be performed algorithmically. Consider the POVM with five outcomes

$$\mathbb{P}_5 = \left\{ \frac{2}{5}E_1, \frac{2}{5}E_2, \frac{2}{5}E_3, \frac{2}{5}E_4, \frac{2}{5}E_5 \right\}, \tag{7.5}$$

where E_i are rank-1 projectors lying on the equator of the Bloch sphere and aligned on the directions shown in Fig 7.1. To carry out its decomposition, one first notices that some subsets of $\{E_i\}$ may form a smaller POVM by themselves with appropriate weights. Then, by selecting one of these subsets (for instance the trine formed by elements 1, 3 and 4), one can rewrite the original POVM as

$$\mathbb{P}_5 = p\mathbb{P}_3^{(1)} + (1-p)\mathbb{P}_4^{(\mathrm{aux})}, \tag{7.6}$$

where $p = 1/\sqrt{5}$ and

$$\mathbb{P}_3^{(1)} = \left\{ \tfrac{2}{\sqrt{5}}E_1, 0, \left(1-\tfrac{1}{\sqrt{5}}\right)E_3, \left(1-\tfrac{1}{\sqrt{5}}\right)E_4, 0 \right\}, \tag{7.7}$$

$$\mathbb{P}_4^{(\mathrm{aux})} = \left\{ 0, \tfrac{2}{5-\sqrt{5}}E_2, \tfrac{3-\sqrt{5}}{5-\sqrt{5}}E_3, \tfrac{3-\sqrt{5}}{5-\sqrt{5}}E_4, \tfrac{2}{5-\sqrt{5}}E_5 \right\}. \tag{7.8}$$

Note that both $\mathbb{P}_3^{(1)}$ and $\mathbb{P}_4^{(\mathrm{aux})}$ are strictly smaller POVMs than $\mathbb{P}_5$. The operation just performed consists in algebraically extracting $\mathbb{P}_3^{(1)}$, in such a way that the remaining elements form a POVM with at least one less outcome (in the following section we prove that such an operation is always possible). Note also that $\mathbb{P}_4^{(\mathrm{aux})}$ is further decomposable. Proceeding iteratively, one can select this time the elements 2, 3 and 5 and write the convex combination of trines

$$\mathbb{P}_4^{(\mathrm{aux})} = p'\mathbb{P}_3^{(2)} + (1-p')\mathbb{P}_3^{(3)}\,, \tag{7.9}$$

where $p' = 1/2$ and

$$\mathbb{P}_3^{(2)} = \left\{0, \left(1-\tfrac{1}{\sqrt{5}}\right)E_2, \left(1-\tfrac{1}{\sqrt{5}}\right)E_3, 0, \tfrac{2}{\sqrt{5}}E_5\right\}\,, \tag{7.10}$$

$$\mathbb{P}_3^{(3)} = \left\{0, \tfrac{2}{\sqrt{5}}E_2, 0, \left(1-\tfrac{1}{\sqrt{5}}\right)E_4, \left(1-\tfrac{1}{\sqrt{5}}\right)E_5\right\}\,. \tag{7.11}$$

Finally, the original 5-outcome POVM can be expressed as a convex combination of 3-outcome POVMs as

$$\mathbb{P}_5 = p_1\mathbb{P}_3^{(1)} + p_2\mathbb{P}_3^{(2)} + p_3\mathbb{P}_3^{(3)} \tag{7.12}$$

where $p_1 = p$, $p_2 = (1-p)p'$ and $p_3 = (1-p)(1-p')$.

Note that both $\mathbb{P}_5$ and $\mathbb{P}_4$ in the previous examples are rank-1 POVMs[2], and hence we need no more than convex combinations of rank-1 POVMs to implement them. However, consider the full-rank 2-outcome POVM

$$\mathbb{P}_2 = \left\{\frac{1}{2}\left|0\right\rangle\!\left\langle 0\right|, \frac{1}{2}\left|0\right\rangle\!\left\langle 0\right| + \left|1\right\rangle\!\left\langle 1\right|\right\}\,. \tag{7.13}$$

In this case it will be enough to measure $\mathbb{P}_2^{(z)} = \{\left|0\right\rangle\!\left\langle 0\right|, \left|1\right\rangle\!\left\langle 1\right|\}$ and, if its first outcome is obtained, then toss an unbiased coin to decide between the two outcomes of $\mathbb{P}_2$. The projector $\left|0\right\rangle\!\left\langle 0\right|$, an element of $\mathbb{P}_2^{(z)}$, is associated with more than one element of $\mathbb{P}_2$, thus the association of the obtained outcome with an original outcome is not immediate. This situation requires an additional step: classical post-processing of the outcomes. This kind of operation has been previously introduced in the literature under the name of *relabeling* [Haapasalo *et al.*, 2011]. In general, the post-processing step will be necessary whenever $\operatorname{rank}(\mathbb{P}_N) > 1$. For any original element E_i such that $\operatorname{rank}(E_i) > 1$, we will split it into a combination of rank-1 operators (by writing it in its eigenbasis) and consider such operators as additional outcomes, thus obtaining a rank-1 POVM that is statistically equivalent to the original one. Of course, to reproduce the statistics accordingly, a map from such new outcomes to the original ones is needed. We address in full detail the case of POVMs of higher rank and the inclusion of a post-processing step in Section 7.3.

[2] A POVM is called rank-1 iff all its non-zero elements E_i are rank-1 operators, i.e. they can be written as $E_i = e_i P_i$, where $0 < e_i \leqslant 1$ and P_i is a normalized one-dimensional projector.

We have seen in this Section examples of measurements that are mixtures of other measurements. The mathematical structure of POVMs is convex: any inner point of the set of POVMs corresponds to a mixed measurement, i.e. it can be expressed as a convex combination of two different POVMs. We denote by $\mathcal{P}_N$ the convex set of POVMs with N elements on $\mathcal{H}$. Note that for any $\mathbb{P} \in \mathcal{P}_N$ we can construct a physically equivalent POVM $\tilde{\mathbb{P}} \in \mathcal{P}_M$, with $M \geqslant N$, just by adding zero-elements to $\mathbb{P}$. The limit of infinite elements yields the convex set of all POVMs $\mathcal{P}$.

An *extremal* POVM is a measurement that cannot be expressed as a mixture of two other POVMs. The 2- and 3-outcome POVMs obtained in the examples above are extremal. If a POVM with N elements $\mathbb{P}$ is extremal in the convex set $\mathcal{P}_N$, then any physically equivalent POVM with M elements $\tilde{\mathbb{P}}$, $M \geqslant N$, is also extremal in $\mathcal{P}_M$. Ultimately, $\mathbb{P}$ will be associated with a set of extremal points of $\mathcal{P}$. So far we have used an apparently more restricted definition of extremality. From the logic of the decompositions presented, it follows that we are considering a rank-1 POVM $\mathbb{P}_N = \{E_i\}$ to be extremal iff there does not exist any subset $\{E_k\} \subset \mathbb{P}_N$, $k = 1, \ldots, M < N$ such that $\mathbb{P}_M = \{a_k E_k\}$ is itself a POVM for a suitable set of positive coefficients $\{a_k\}$. We have seen that if such a subset exists, then $\mathbb{P}_N$ can be split in $\mathbb{P}_M$ plus another POVM. We are therefore considering only decompositions into extremals formed by a subset of elements of the original $\mathbb{P}_N$. However, we prove in Section 7.2 that looking for such subsets is sufficient to check for extremality of a given POVM.

7.2 Selection of extremal POVMs and geometric characterization

The decomposition of the POVMs presented as examples above is achieved through the selection of subsets of their elements capable of forming a POVM by themselves. In order to give some insight on how to perform this selection for a general POVM $\mathbb{P}$ with N outcomes, we now examine the conditions under which a set of n arbitrary rank-1 operators $\{E_i\}$ can comprise a POVM, that is, there is a set of positive coefficients $\{a_i\}$ such that $\sum_{i=1}^{n} a_i E_i = \mathbb{1}$. For simplicity and w.l.o.g. we will assume the operators E_i to be normalized (i.e., $\operatorname{tr} E_i = 1$). Recall that, for a d-dimensional Hilbert space, we can express E_i in a generalized Bloch-like representation as

$$E_i = \left(\frac{1}{d}\mathbb{1} + \frac{1}{2} \sum_j \langle \hat{\lambda}_j \rangle_i \hat{\lambda}_j \right) , \tag{7.14}$$

where the operators $\hat{\lambda}_j$, $j = 1, \ldots, d^2-1$ are an orthogonal basis of generators of SU(d) and the generalized Bloch vector $\boldsymbol{v}_i$ is defined with their expectation values: $\boldsymbol{v}_i \equiv (\langle\hat{\lambda}_1\rangle_i, \ldots, \langle\hat{\lambda}_{d^2-1}\rangle_i)$. In this representation, pure states have associated a generalized Bloch vector of fixed length $|\boldsymbol{v}| = \sqrt{2(d-1)/d}$. Then, the POVM condition may be equivalently written as

$$\sum_i a_i = d\,, \tag{7.15}$$

$$\sum_i a_i \boldsymbol{v}_i = \boldsymbol{0}\,, \tag{7.16}$$

that is a system of d^2 linear equations. At this point we are only interested in checking the consistency of (7.15) and (7.16). Therefore, the existence of the set $\{a_i\}$ can be cast as a *linear programming feasibility problem.*

Before proceeding further, let us briefly overview the standard linear programming formalism (for an extensive review on the topic see e.g. [Boyd and Vandenberghe, 2004; Todd, 2002]). A general *linear program* (LP) has the standard form

$$\begin{array}{rl} \text{min} & c^T x \\ \text{subject to} & Ax = b \\ & x \geqslant 0\,, \end{array} \tag{7.17}$$

where $A \in \mathbb{R}^{p\times q}$, $b \in \mathbb{R}^p$ and $c \in \mathbb{R}^q$ are the given data, and the vector $x \in \mathbb{R}^q$ is the variable to optimize. We call (7.17) *feasible* if there exists $x \in \mathbb{R}^q$ such that $Ax = b$, $x \geqslant 0$. Any LP of the standard form above has a *dual problem* of the form

$$\begin{array}{rl} \text{max} & -b^T \nu \\ \text{subject to} & A^T\nu + c \geqslant 0\,, \end{array} \tag{7.18}$$

where $\nu \in \mathbb{R}^p$. Let us assume that both LPs (7.17) and (7.18) are feasible. Then, we may write

$$c^T x + b^T \nu = x^T c + x^T A^T \nu = x^T (c + A^T \nu) \geqslant 0\,. \tag{7.19}$$

In order to obtain feasibility conditions of the LP (7.17), we now set $c = 0$ and solve it. The existence of a solution implies that (7.17) is feasible and, from (7.18) and (7.19), that for all vectors ν, $A^T\nu \geqslant 0$ implies $b^T\nu \geqslant 0$. If the dual problem does not have a solution, then its corresponding LP neither has one. Conversely, the existence of a vector ν that verifies the conditions

$$A^T\nu \;\leqslant\; 0\,, \tag{7.20}$$

$$b^T\nu \;>\; 0\,, \tag{7.21}$$

implies the infeasibility of (7.17). Notice that finding a ν subject to $A^T\nu \geqslant 0$, $b^T\nu < 0$ is an equivalent problem.

We are now in the position to reinterpret the problem of finding the set of coefficients $\{a_i\}$ within the general linear program scheme presented above. The components of the vector x are the coefficients we want to determine, that is $x = \{a_1, a_2, \ldots, a_n\}$. Conditions (7.15) and (7.16) can be cast together in the $Ax = b$ equation: A is a matrix whose columns are given by vectors $v_i = (\boldsymbol{v}_i, 1)$, and $b = (\mathbf{0}, d)$. Therefore, the dimensions of this linear program are given by $p \equiv d^2, q \equiv n$. In the dual problem the vector ν has dimension d^2 and is unrestricted. However, for later convenience and w.l.o.g. let us choose the specific form $\nu = (\beta\boldsymbol{\nu}, \alpha)$, where $\alpha \in \mathbb{R}, \beta \in \mathbb{R}^+$ are arbitrary constants and $|\boldsymbol{\nu}| = \sqrt{2(d-1)/d}$. From Eqs. (7.20) and (7.21) we have

$$\beta\boldsymbol{v}_i \cdot \boldsymbol{\nu} + \alpha \leqslant 0\,, \tag{7.22}$$

$$\alpha > 0\,. \tag{7.23}$$

A vector ν will simultaneously satisfy these conditions if and only if $\boldsymbol{v}_i \cdot \boldsymbol{\nu} < -\alpha/\beta$. We can always choose β sufficiently large such that $-\alpha/\beta \to 0$, so the least restrictive condition has the form

$$\boldsymbol{v}_i \cdot \boldsymbol{\nu} < 0 \tag{7.24}$$

[taking the complementary equations to (7.20) and (7.21) would have led to the equivalent condition $\boldsymbol{v}_i \cdot \boldsymbol{\nu} > 0$]. To summarize, as long as there exists a vector $\boldsymbol{\nu}$ whose scalar product with every other generalized Bloch vector $\boldsymbol{v}_i$ is negative, we can always choose two positive constants α, β such that $\nu = (\beta\boldsymbol{\nu}, \alpha)$ satisfies Eqs. (7.20) and (7.21). Hence, the LP (7.17) is infeasible and the set of operators $\{E_i\}$ cannot form a POVM.

Condition (7.24) has a clear geometrical interpretation: $\boldsymbol{\nu}$ defines a hyperplane in $\mathbb{R}^{d^2-1}$ which includes the $\mathbf{0}$ point and splits a (d^2-2)-sphere such that all $\boldsymbol{v}_i$ points are situated at one side of the hyperplane. Obviously, if the vectors $\boldsymbol{v}_i$ do not span $\mathbb{R}^{d^2-1}$ but a subspace of smaller dimension d', it will suffice to consider hyperplanes of dimension $d'-1$. This hyperplane condition is equivalent to stating that the convex hull of the $\boldsymbol{v}_i$ points does not contain the $\mathbf{0}$ point.

We now state and prove next that, given a POVM with $n > d^2$ non-zero elements, it is always possible to select a subset of at most d^2 which is also a POVM, up to a suitable redistribution of weights. This is easily derived from the LP feasibility formulation: Eqs. (7.15) and (7.16) represent a system of d^2 equality conditions and n variables; if such a system is feasible, it would have a single solution for some value of $n \leqslant d^2$. For $n > d^2$ its solution will have $n - d^2$ extra degrees of freedom, and hence we will always be able to fix $n - d^2$ variables to zero. Since this statement is not valid when $n \leqslant d^2$ (except for the case in which vectors $\boldsymbol{v}_i$ span a smaller subspace of $\mathbb{R}^{d^2-1}$), it follows that an extremal POVM will have at most d^2 non-zero elements, as it has been noted in previous works [D'Ariano *et al.*, 2005; Haapasalo *et al.*, 2011].

The geometrical interpretation of the POVM condition provides a clear and useful picture of the results in the previous paragraph in terms of the distribution of vectors $\boldsymbol{v}_i$. Note that the number of vectors needed to subtend a solid angle in $\mathbb{R}^{d^2-1}$ is $d^2 - 1$. The conical hull defined by such vectors contains a portion of a hypersphere S^{d^2-2}. It is then easy to convince oneself that the minimum number of vectors required to cover the whole S^{d^2-2} as a union of conical hulls is d^2 [note that such a distribution necessarily implies the violation of condition (7.24) and, therefore, the fulfilment of (7.16)]. This means that, given such a set of d^2 vectors, if we add an extra vector, it will necessarily fall in a conical hull defined by a certain subset of $d^2 - 1$ vectors of the original set and thus it could be expressed as a conical combination of those (i.e. as a linear combination with nonnegative coefficients). Hence, given d^2+1 POVM elements whose Bloch vectors satisfy condition (7.16), one can always choose one of the vectors and replace it by a conical combination of d^2-1 other vectors: the remaining set of d^2 vectors still satisfies condition (7.16).

In general, Bloch vectors $\boldsymbol{v}_i$ will be contained in $\mathbb{R}^{d^2-1}$. When $n < d^2$, additional restrictions over vectors $\boldsymbol{v}_i$ derive from (7.24). If $n = 2$ then the generalized Bloch vectors $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$ should span a 1-dimensional space in order to be able to violate condition (7.24). In fact, the condition is violated only if $\boldsymbol{v}_1 = -\boldsymbol{v}_2$. If $n = 3$, vectors $\boldsymbol{v}_1, \boldsymbol{v}_2$ and $\boldsymbol{v}_3$ should lie on a plane and not belong to the same semicircle (defined by a line). For any n we should have

$$\{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n\} \in S^{n-2} \subset \mathbb{R}^{n-1}\,, \tag{7.25}$$

where vectors $\boldsymbol{v}_i$ do not belong to any hemisphere of S^{n-2}. Note that the extremality statement in the previous paragraph extends to $\mathbb{R}^{n-1}$: if we have $n' \geqslant n + 1$ vectors (whose associated operators form a POVM) that span $\mathbb{R}^{n-1}$, then we can always find subsets of at most n vectors which violate condition (7.24), and thus are able to form an extremal POVM.

To finish this section and for clarity purposes, note that it has been assumed that the solutions of the LP feasibility problem correspond to extremal POVMs, i.e. extremal points not only of the set of feasible points but also of the set of all POVMs. This is indeed the case: on one hand, such a solution corresponds to a set of linearly independent POVM elements $\{E_i\}$; on the other hand, any POVM with at most d^2 rank-1 linearly independent elements is extremal (see, e.g., Proposition 3 in [Haapasalo *et al.*, 2011]).

7.3 The algorithm

In this section, we present our constructive algorithm for decomposing a POVM into extremals. We first address the case of rank-1 POVMs, and then we extend the algorithm to higher-rank cases. We are given a rank-1 POVM $\mathbb{P}_N = \{a_i E_i\}$, $i = 1, \ldots, N$, where E_i are normalized operators given by (7.14) and $a_i > 0$. Our aim is to express it as

$$\mathbb{P}_N = \sum_k p_k \mathbb{P}_n^{(k)}, \tag{7.26}$$

where $\mathbb{P}_n^{(k)}$ is an extremal rank-1 POVM with $n \leqslant d^2$ outcomes. This means that in order to implement $\mathbb{P}_N$ it will suffice to randomly select a value of k from the probability distribution p_k, and then perform $\mathbb{P}_n^{(k)}$. The algorithm we propose to carry out such a decomposition works as follows:

We first define the LP feasibility problem

$$\begin{aligned} \text{find} \quad & x \\ \text{subject to} \quad & Ax = b \\ & x \geqslant 0\,, \end{aligned} \tag{7.27}$$

where x is a vector of N variables, A is a matrix whose columns are given by vectors $v_i = (\boldsymbol{v}_i, 1)$, and $b = (\mathbf{0}, d)$. The set of feasible points of this LP, i.e. the values of x compatible with the conditions of the LP, define a convex polytope K in the space of coefficients:

$$K = \{x \,/\, Ax = b, x \geqslant 0\} \subset \mathbb{R}^N. \tag{7.28}$$

The vertices of K are its extremal points, and the region of $\mathbb{R}^N$ defined by the convex hull of all the vertices contains all the points that can be expressed as convex combinations of these extremal points. Dantzig's *simplex method* for solving LPs [Todd, 2002] starts at a vertex of K, and it moves from vertex to vertex minimizing a cost function, until there is no preferred direction of minimization; then, the optimal solution has been found. Since there is no cost function in a feasibility problem, the simplex method applied to (7.27) terminates at its first step: when it finds the first vertex. The convex polytope K is isomorphic to a subset of $\mathcal{P}_N$, i.e. there is a one-to-one correspondence between all their elements, and they behave equivalently. Therefore, such a vertex $x^{(1)} = \{x_i^{(1)}\}$ found as the solution of the LP corresponds to the set of coefficients of an extremal POVM, and as such $x^{(1)}$ will have at most d^2 and at least d non-zero elements. The vertices of the polytope K correspond to all the extremal POVMs that one can comprise using only the original elements $\{E_i\}$, and its interior region contains all the possible POVMs generated by these extremals.

Once we have found $x^{(1)}$, we algebraically subtract it from the original set of coefficients $\{a_i\}$. To illustrate this operation, let us assume $d = 2$ and $x^{(1)} = \{x_1^{(1)}, x_2^{(1)}, 0, \ldots, 0\}$. Then, $\{a_i\}$ is rewritten as

$$\{a_1, a_2, a_3, \ldots, a_N\} = p\, x^{(1)} + (1-p) x^{(\mathrm{aux})}\,, \tag{7.29}$$

$$x^{(\mathrm{aux})} = \left\{ \tfrac{a_1 - p\, x_1^{(1)}}{1-p}, \tfrac{a_2 - p\, x_2^{(1)}}{1-p}, \tfrac{a_3}{1-p}, \ldots, \tfrac{a_N}{1-p} \right\}. \tag{7.30}$$

For $x^{(\mathrm{aux})}$ to be an element of K, the inequality

$$p \leqslant a_i / x_i^{(1)} \leqslant 1 \tag{7.31}$$

has to hold for all i such that $x_i^{(1)} > 0$. To guarantee the left-hand side of (7.31), we take

$$p = \min_i \frac{a_i}{x_i^{(1)}}\,. \tag{7.32}$$

Let us reorder the coefficients $\{a_i\}$ and $x^{(1)}$ such that $p = a_1 / x_1^{(1)}$. This choice of p makes the first coefficient of $x^{(\mathrm{aux})}$ to be zero (it could happen that more than one element turns to be zero, thus accelerating the algorithm, but we consider from now on the worst case scenario in which one element is eliminated at a time). Also, the right-hand side of (7.31) is immediately satisfied since $a_1 < x_1^{(1)}$. Note that $p \in [0,1]$, thus it is a probability. Now, (7.29) can be understood as a probabilistic (convex) combination of $x^{(1)}$ and $x^{(\mathrm{aux})}$, both set of coefficients corresponding to an extremal POVM $\mathbb{P}_2^{(1)}$ and a POVM with $N-1$ outcomes $\mathbb{P}_{N-1}^{(\mathrm{aux})}$. Hence, as a result of the first step of the algorithm, we can write

$$\mathbb{P}_N = p\, \mathbb{P}_2^{(1)} + (1-p)\, \mathbb{P}_{N-1}^{(\mathrm{aux})}\,. \tag{7.33}$$

We then repeat this process redefining the LP with $\mathbb{P}_{N-1}^{(\mathrm{aux})}$ as the initial POVM, which gives us another vertex $x^{(2)}$ associated to an extremal POVM with n outcomes $\mathbb{P}_n^{(2)}$, a remainder $\mathbb{P}_{N-2}^{(\mathrm{aux})}$ and its corresponding probabilities. Of course, in general $d \leqslant n \leqslant d^2$. We iterate this process $N - n_L$ times, where n_L is the number of outcomes of the last extremal POVM obtained.

At the last step the simplex algorithm will identify a unique solution with probability 1, corresponding to the input set $x^{(\text{aux})} = x^{(N-n_L)}$.

It is important to stress that the polytopes of the LPs at each step of the algorithm, K^k, are subsequent subsets of each other, that is

$$K \supset K^1 \supset \ldots \supset K^{N-n_L+1}. \tag{7.34}$$

The result of each step is the elimination of one of the original elements $\{E_i\}$, and with it all the vertices that required that element. Thus, each step projects the polytope onto a subspace of the space of coefficients by reducing its dimension by one. As a consequence, in the end all the vertices selected by the simplex algorithm were vertices of the original K.

When the rank of $\mathbb{P}_N$ is higher than 1 we can still apply the same algorithm, just adding two extra steps: one preparation step and one post-processing step. The preparation step works as follows: for every i such that $\text{rank}\,(E_i) > 1$, express E_i in its eigenbasis $\{|v_{ij}\rangle\}$ as

$$E_i = \sum_j \lambda_j \, |v_{ij}\rangle\!\langle v_{ij}| = \sum_j E_{ij}. \tag{7.35}$$

Consider each rank-1 operator E_{ij} as a new outcome and denote the new (rank-1) POVM by $\mathbb{P}_{\bar{N}} = \{\bar{E}_l\}_{l=1}^{\bar{N}}$, where $\bar{N} = \sum_i \text{rank}\,(E_i) > N$. The label $l(i,j)$ carries the information contained in labels i and j. Now, the algorithm described above can be applied directly over $\mathbb{P}_{\bar{N}}$. The post-processing step is needed for associating the outcomes of the measure finally performed (l) to the outcomes of the original $\mathbb{P}_N$ (i).

A generic algorithm for decomposing a point in a convex set into a combination of extremal points of that set can be found in [D'Ariano *et al.*, 2005]. Although in this paper D'Ariano *et al.* specialize it for a general $\mathbb{P} \in \mathcal{P}_N$, we would like to remark that significant differences stand between our algorithm and the one presented there. The algorithm of [D'Ariano *et al.*, 2005] consists in a recursive splitting of an inner point of the convex set into a convex combination of two points that lie on a facet of the convex set (and thus a subset of a strictly smaller dimension). After enough steps it yields a number of extremal points along with some weights in a tree-like form, thus statistically reproducing the original point as a mixture of extremal points. The direction in which the splitting is done at each step is determined through an eigenvalue evaluation. The particular decomposition we have presented in this Chapter may be considered within this general scheme (we also do binary partitions at each step), however two main differences arise. On one hand, the process of obtaining extremal points (i.e. the direction of splitting) is radically different. We associate a polytope K to a subset of the convex set $\mathcal{P}_N$ via an isomorphism, and then we move efficiently along the directions marked by the vertices of K. Thus, there is no need to analyze the whole convex set $\mathcal{P}_N$ (which is strongly convex, i.e. its extremal points are not isolated but lie on a continuum) for a given $\mathbb{P}$: our algorithm does not optimize a direction among a continuum of possibilities at each step but selects any direction of a given finite set. On the other hand, the authors in [D'Ariano *et al.*, 2005] state that their algorithm provides a minimal decomposition, with a number of extremals upperbounded by $(N-1)d^2+1$. We have found that our algorithm yields the tighter bound $(N-1)d+1$.

7.4 Ordered decompositions

The algorithm described in Section 7.3 will produce one of many possible decompositions of the initial POVM into at most $N - n_L + 1$ extremals (recall that n_L ranges from d to d^2), even if we only consider extremals made of original elements. Because at each step any of the vertices of the polytope could be identified and extracted, the final decomposition obtained is not unique and depends on the particular implementation of the simplex method for solving the LP. That being said, one could be interested in a particular decomposition that exhibits certain properties. We observe that there is room in our algorithm for these extra requirements while maintaining its structure, that is to efficiently produce decompositions into at most $N-n_L+1$ extremals obtained through a LP solved by the simplex method. To obtain a particular decomposition with this structure that verifies a certain desired property we will simply have to establish some ranking among the vertices of the polytope in agreement to that property or associated criterion, and tweak the algorithm to choose first the ones at the top of the ranking. This is what we call an *ordered* decomposition.

A desirable ordering from the point of view of an experimental realization may be, for instance, to prioritize the vertices with more zero elements, if there is any. Those vertices would correspond to extremals with less outcomes. In the case of $d = 2$, for instance, extremal POVMs can have 2, 3 or 4 outcomes. Such a decomposition would seek first for 2-outcome (Stern-Gerlach measurements), then 3-outcome and finally 4-outcome POVMs.

The simplex method is an efficient way of finding the optimal vertex of a polytope according to some criterion, which is implemented as a cost function. This is done by minimizing or maximizing such a cost function. In the description of the algorithm we chose this function to be independent of the variables, because we were only interested in finding a feasible point. The choice of the cost function will vary the direction taken by the simplex algorithm when it moves from one vertex to another, and it is therefore a way to establish a ranking among the vertices. Consider for instance the cost function

$$Q_n = \sum_{i=1}^{n} x_i^2 \,. \tag{7.36}$$

The maximization of Q_n on its own could in principle work for finding the vertices with more zeros: if we would have no other constraint but a fixed quantity d to distribute among the n parties x_i, the strategy that maximizes Q_n is to give all to one party and zero to the others. But we have more constraints in (7.27). Let us take a look on the minimum and maximum values of Q_4, that is for extremals with 4 outcomes. The value of Q_4 will only depend on the geometric distribution of the outcomes of the extremal. On one hand, Q_4 takes its minimum value when $d = \sum_i x_i$ is equally distributed among the variables x_i, that is when the 4 associated Bloch vectors $\boldsymbol{v}_i$ are orthogonal in pairs (i.e. the POVM is a combination of two Stern-Gerlachs). This value is $Q_4^{\min} = (d/4)^2 \times 4 = d^2/4$. On the other hand, Q_4 reaches its maximum value if three of the vectors are parallel and the fourth is orthogonal to all the others (this is the way to put a maximum weight on one of the x_i), that is $Q_4^{\max} = (d/2)^2 + (d/6)^2 \times 3 = d^2/3$. Applying the same reasoning for 3-outcome extremals we have $Q_3^{\min} = d^2/3$ and $Q_3^{\max} = 3d^2/8$, and 2-outcomes can only give $Q_2 = d^2/2$. Since

$$Q_2 > Q_3^{\max} > Q_3^{\min} = Q_4^{\max} > Q_4^{\min} \,, \tag{7.37}$$

the maximization of function Q_n prioritizes the extremals with fewer outcomes at least for $d = 2$, when the maximum number of nonzero elements in a vertex is $n = 4$. This, unfortunately, stops being valid for $n > 4$, which in general happens if $d > 2$.

The general problem of maximizing a convex function over a convex set of feasible points is called *convex maximization*. The problem at hand belongs to this category. While the more standard class of *convex minimization* problems (i.e. minimizing a convex function over a convex polytope) count on efficient solving algorithms, this is not the case for convex maximization, except for very special cases. The efficiency of the convex minimization relies on the uniqueness of the convex function's minimum, which is an inner point of the polytope. Conversely, its maxima are located on the vertices of the polytope and all but one are *local* maxima. This fact makes the convex maximization problems intractable in general, and so it is the maximization of (7.36). The difficulty lies on the fact that an algorithm might find a local maximum (a vertex), but there is no way to certificate its global optimality (although there are algorithms that, despite no proof certificate, provide good guesses [Fortin and Tseveendorj, 2010]).

Any global search algorithm (able to guarantee global optimality) for convex maximization somehow *enumerates* all the vertices, and thus its efficiency highly depends on the number of those. Of course, the ordered decomposition we are looking for is immediately obtained if one enumerates all the vertices of K. With such a list, we would just have to pick up first those vertices with more zero elements, corresponding to the extremals with fewer outcomes (or according to any other criterion we may wish). Furthermore, no additional optimization is required since we can extract from the same list the vertex required at each step, thus keeping us from solving a LP for doing so. The problem of enumerating the vertices of a bounded polyhedron is NP hard in the general case [Khachiyan *et al.*, 2008], but has efficient algorithms able to generate all vertices in polynomial time (typically linear in the number of vertices) for several special cases. For instance, in [Avis and Fukuda, 1992] there is an algorithm that enumerates the v vertices of a convex polyhedron in $\mathbb{R}^m$ defined by a system of D linear inequalities in time $O(mDv)$. Our polytope K is of this type, and hence we could use the algorithm for our purpose. Note however that v has a direct dependence on m and D. The problem of computing v for a given polytope is NP-hard, but a bound can be provided [Barvinok, 2012]: the number of vertices of our polytope $K \subset \mathbb{R}^m$ is at least exponential in m.

In summary, an ordered decomposition of a POVM can be carried out in two ways. On one hand, nonlinear programming techniques can be used to maximize a cost function subject to the constraints of (7.27), but none of them will perform with perfect accuracy. We have found a cost function that prioritizes the extremals with less outcomes for $d = 2$, but not for greater dimensions. Finding a cost function is problem-specific, and it seems to be highly non-trivial: its maximization should lead first to a vertex of the polytope, and secondly it should move from one to another maximizing the desired property. On the other hand, an alternative method is to enumerate all the vertices of the polytope K defined by the constraints of (7.27), but the number of vertices and thus the time required to carry out the enumeration grows exponentially with the number of elements of the original POVM.

7.5 Discussion

We have presented an efficient algorithm to decompose any POVM $\mathbb{P} \in \mathcal{P}_N$ into extremal ones. The decomposition achieved consists of a convex combination of at least $N - n_L + 1$ (if $\mathbb{P}$ is rank-1) and at most $Nd - n_L + 1$ (if $\mathbb{P}$ is full-rank) extremal measurements, where n_L ranges from d to d^2 and its value is determined by each particular $\mathbb{P}$. In the case in which $\mathbb{P}$ presents some symmetry (as the BB84 POVM shown as an example in Section 7.1), more than one element may be eliminated in one step of the algorithm and thus the number of extremals would be even less. We have shown that only extremal rank-1 POVMs are required to effectively implement $\mathbb{P}$ by introducing a classical post-processing of the outcomes. The decomposition is efficiently carried out by an algorithm based on resolutions of LPs using the simplex method, within polynomial time in N and d. The efficiency is achieved by restricting the analysis to a polytope-shaped subset of $\mathcal{P}_N$ for a given $\mathbb{P}$, and thus by taking into consideration only a finite number of extremals (the vertices of the polytope), in contrast to what other authors have considered so far (see e.g. [D'Ariano *et al.*, 2005]). Furthermore, in [D'Ariano *et al.*, 2005], a generic decomposition algorithm that yields a certain maximum number of extremals is provided. We have found that our algorithm beats this performance in a worst case scenario.

Since a given POVM admits many decompositions, we also explore the possibility of obtaining a particular decomposition that exhibits a certain desired property, introduced in the algorithm as an input. We call these decompositions *ordered*, and they are based on prioritizations of extremals that can be made out of subsets of the elements of $\mathbb{P}$. As an example we give a method to prioritize extremal POVMs with less outcomes in the case of $d = 2$, and show that either efficiency or accuracy necessarily get compromised.

Outlook

The specific conclusions of the research projects addressed in this thesis have already been discussed at the end of each corresponding chapter. Here, I would like to finish by giving a brief outlook on future research lines and open problems that naturally arise from within the covered topics.

The group-theoretic concepts used in Chapter 4 to compute the optimal programmable discrimination machine for qubits can also be applied to higher-dimensional systems. In fact, some results are already available in the literature for pure states of arbitrary dimension [Hayashi *et al.*, 2005, 2006; Akimoto and Hayashi, 2011], but the mixed states case remains an open problem, and so does the fully universal discrimination machine, for states of more than two dimensions. In this line of generalizations, the extreme case of infinite dimensions, i.e., programmable discrimination of continuous-variables systems, has only been discussed before for coherent states and unambiguous discrimination [Sedlák *et al.*, 2007, 2009]. Although Chapter 6 provides an instance of programmable minimum-error discrimination with coherent states, there is much work to be done. Extending the applicability of programmable discrimination protocols to general Gaussian states, or even more complex cases such as multimode entangled states, would be of great fundamental and practical interest.

In Chapter 5, I analysed the classification of qubit states in a *supervised* learning scenario. The most obvious generalization, and the most promising one, is to consider *unsupervised* scenarios, where no human expert classifies the training sample. This is a challenging problem with direct practical applications in quantum control and information processing, and, although the literature on the topic is not abundant, it is beginning to raise much attention (see, e.g., [Lloyd *et al.*, 2013]).

Another generalization of both programmable and learning machines is to consider more than two possible states, although the scarcity of results in general multihypothesis quantum state discrimination is somewhat discouraging—it is expected that only very special cases will be analytically tractable. A more promising extension is to analyse the behaviour of the proposed programmable and learning machines under the more general scheme of discrimination with an error margin. On the one hand, programmable discrimination of mixed states has yet to be considered when a limiting margin is imposed on the rate of errors. On the other hand, a very interesting question that remains unanswered to date is whether the optimality of the learning protocol proposed in Chapter 5 is compromised when one allows for some proportion of inconclusive answers.

APPENDIX A

Technical details of Chapter 4

A.1 Wigner's $6j$-symbols

Let us consider three angular momenta j_1, j_2, j_3 that couple to give a total J. Note that there is no unique way to carry out this coupling; we might first couple j_1 and j_2 to give a resultant j_{12}, and couple this to j_3 to give J, or alternatively, we may couple j_1 to the resultant j_{23} of coupling j_2 and j_3. Moreover, the intermediate couplings can give in principle different values of j_{12} or j_{23} which, when coupled to j_3 or j_1, end up giving the same value of J. All these possibilities lead to linearly independent states with the same J and M, thus they must be distinguished by specifying the intermediate angular momentum and the order of coupling. There exists a unitary transformation that maps the states obtained from the two possible orderings of the coupling; Wigner's 6j-symbols [Edmonds, 1960], denoted in the next equation by $\{ \begin{smallmatrix} \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \end{smallmatrix} \}$, provide the coefficients of this transformation:

$$\begin{aligned} &\langle (j_1\, j_2) j_{12}, j_3; J, M | j_1, (j_2\, j_3) j_{23}; J, M \rangle \\ &\qquad = (-1)^{j_1+j_2+j_3+J} \sqrt{(2j_{12}+1)(2j_{23}+1)} \begin{Bmatrix} j_1 & j_2 & j_{12} \\ j_3 & J & j_{23} \end{Bmatrix} . \end{aligned} \tag{A.1}$$

Note that this overlap is independent of M.

A.2 Arbitrary number of copies

In this Section we present the probabilities for unambiguous and minimum-error discrimination when the number of copies n_A, n_B, n_C loaded at the machine ports is completely arbitrary. Note that, in this case, the global states σ_1 and σ_2 [cf. Eq. (4.2)] may have different dimensions, for $d_1 = (n_A + n_B + 1)(n_C + 1)$ is in general not equal to $d_2 = (n_A + 1)(n_B + n_C + 1)$. One can easily convince oneself that the support of the state with smallest dimension is always contained in the support of the other, and hence the problem can be solved in very much the same way as in the main text as far as the intersection of the supports is concerned. The remaining of the state with higher dimension yields a trivial contribution to the error probabilities. Without loss of generality we can assume from now on that $n_A \geq n_C$. As discussed in the main text, the error probabilities are computed by adding the pairwise contributions of the state bases in the common support, the main difference being that σ_1 and σ_2 do not have equal coefficients in front of the projectors and hence the prior probabilities of each pair of states are different. Also, the overlaps in Eq. (4.6) will have a slightly more complicated expression. Here we have $j_A = n_A/2$, $j_B = n_B/2$, $j_C = n_C/2$, $j_{AB} = (n_A + n_B)/2$ and $j_{BC} = (n_B + n_C)/2$. The minimum J available for σ_1 is $j_B + j_A - j_C \equiv J^1_{\text{min}}$, and $|j_B + j_C - j_A| \equiv J^2_{\text{min}}$ for σ_2. The maximum angular momentum $j_A + j_B + j_C \equiv J_{\text{max}}$ is reachable for both states. For equal prior probabilities for σ_1 and σ_2, we can write

$$\frac{1}{2}\sigma_1 = \sum_{J=J^1_{\text{min}}}^{J_{\text{max}}} \sum_{M=-J}^{J} p_J\, \pi^1_J [j_{AB}; JM]\,, \tag{A.2}$$

$$\frac{1}{2}\sigma_2 = \sum_{J=J^2_{\text{min}}}^{J_{\text{max}}} \sum_{M=-J}^{J} p_J\, \pi^2_J [j_{BC}; JM]\,, \tag{A.3}$$

where $p_J = \frac{1}{2}\left(\frac{1}{d_1} + \frac{1}{d_2}\right)$, $\pi^1_J = \frac{1}{2p_J\, d_1}$, $\pi^2_J = \frac{1}{2p_J\, d_2}$ for $J^1_{\text{min}} \leq J \leq J_{\text{max}}$, whereas $p_J = \frac{1}{2d_2}$, $\pi^1_J = 0$, $\pi^2_J = 1$ for $J^2_{\text{min}} \leq J < J^1_{\text{min}}$. We view p_J as the probability of obtaining the outcome (M) J in a measurement of the (z component of the) total angular momentum on the unknown state. Likewise, we view π^1_J, $\pi^2_J = 1 - \pi^1_J$ as the probabilities that the unknown state be $[j_{AB}; JM]$ or $[j_{BC}; JM]$ for that specific pair of outcomes J and M (note that these probabilities are actually independent of M). If the condition

$$\frac{c_J^2}{1 + c_J^2} \leq \pi^{AB}_J \leq \frac{1}{1 + c_J^2}\,, \tag{A.4}$$

where $c_J = |\langle j_{AB}; JM | j_{BC}; JM\rangle|$ is given by Eq. (4.6), holds, then the probability of obtaining an inconclusive answer when we finally discriminate between $[j_{AB}; JM]$ and $[j_{BC}; JM]$ is $Q_J = 2\sqrt{\pi^1_J \pi^2_J} c_J$ [cf. Eq. (3.47)]. If Eq. (A.4) is satisfied for $\hat{J} = J_{\max} - 1$, then it will be satisfied all over this range of J, since c_J is a monotonically increasing function of J. The overlap $c_{\hat{J}}$ has the very simple form

$$c^2_{\hat{J}} = \frac{n_A n_C}{(n_A + n_B)(n_B + n_C)} . \tag{A.5}$$

Thus Eq. (A.4) is equivalent to

$$\begin{aligned} \frac{n_A n_C}{(n_A + n_B)(n_B + n_C)} &\leq \frac{(n_A + n_B + 1)(n_C + 1)}{(n_B + n_C + 1)(n_A + 1)} \\ &\leq \frac{(n_A + n_B)(n_B + n_C)}{n_A n_C} , \end{aligned} \tag{A.6}$$

which is clearly true. Eq. (A.4) does not hold if $J = J_{\max}$, for which we have $Q_{J_{\max}} = 1$. Note that since no error is made for $J^2_{\min} \leq J < J^1_{\min}$, for $\pi^1_J = 0$, the total inconclusive probability reads $Q = \sum_{J=J^1_{\min}}^{J_{\max}} p_J\,(2J+1) Q_J$, which has the explicit expression

$$\begin{aligned} Q = \frac{1}{2}\left(\frac{1}{\sqrt{d_1}} - \frac{1}{\sqrt{d_2}}\right)^2 d_{ABC} + \frac{1}{\sqrt{d_1 d_2}} \sum_{k=0}^{n_C} (n_A + n_B - n_C + 2k + 1) \\ \times \sqrt{\frac{\binom{n_A + n_B - n_C + k}{n_B}\binom{n_B + k}{n_B}}{\binom{n_A + n_B}{n_B}\binom{n_C + n_B}{n_B}}} , \end{aligned} \tag{A.7}$$

where $d_{ABC} = n_A + n_B + n_C + 1$. Note also that, when $n_A = n_C$, the term proportional to d_{ABC} vanishes and the square root term simplifies, so we recover the closed form given in the main text [cf. Eq. (4.8)].

The minimum-error probability can be computed entirely along the same lines. For a pair of states we have $P_{e,J} = \frac{1}{2}\left(1 - \sqrt{1 - 4\pi^1_J \pi^2_J c^2_J}\right)$ [cf. Eq. (3.34)], and the total error probability reads

$$\begin{aligned} P_e &= \frac{1}{4}\left\{1 + \frac{d_1}{d_2} - \frac{d_1 + d_2}{d_1 d_2} \sum_{k=0}^{n_C} (n_A + n_B - n_C + 2k + 1) \right. \\ &\quad \left. \times \sqrt{1 - 4\frac{d_1 d_2}{(d_1 + d_2)^2} \frac{\binom{n_A + n_B - n_C + k}{n_B}\binom{n_B + k}{n_B}}{\binom{n_A + n_B}{n_B}\binom{n_C + n_B}{n_B}}}\right\} . \end{aligned} \tag{A.8}$$

This expression coincides with Eq. (31) of [Akimoto and Hayashi, 2011].

A.3 Averaged C_j^n coefficients

Here we compute the average of the coefficients [see Eq. (4.42)]

$$C_j^n = \frac{1}{2j+1}\left(\frac{1-r^2}{4}\right)^{n/2-j}\sum_{k=-j}^{j}\left(\frac{1-r}{2}\right)^{j-k}\left(\frac{1+r}{2}\right)^{j+k} \tag{A.9}$$

for the hard-sphere, Bures and Chernoff priors, given by Eqs. (4.74) through (4.76), considered in the fully universal discrimination machine.

For the hard-sphere prior we have

$$\langle C_j^n\rangle_{\rm HS} = 3\int C_j^n r^2 dr = 6\frac{\Gamma(n/2+j+2)\Gamma(n/2-j+1)}{\Gamma(n+4)}\,. \tag{A.10}$$

The Bures distribution yields

$$\langle C_j^n\rangle_{\rm Bu} = \frac{4}{\pi}\int C_j^n \frac{r^2}{\sqrt{1-r^2}}dr = \frac{4}{\pi}\frac{\Gamma(n/2+j+3/2)\Gamma(n/2-j+1/2)}{\Gamma(n+3)}\,. \tag{A.11}$$

The averages for the Chernoff prior are a bit more involved, but still can be given in a closed form as

$$\begin{aligned}\langle C_j^n\rangle_{\rm Ch} &= \frac{1}{\pi-2}\int C_j^n \frac{\left(\sqrt{1+r}-\sqrt{1-r}\right)^2}{\sqrt{1-r^2}}dr \\ &= \frac{2}{(\pi-2)(2j+1)}\sum_{m=-j}^{j}\left[B_{1/2}\left(\tfrac{n+1-2m}{2},\tfrac{n+1+2m}{2}\right)\right. \\ &\qquad \left. -2B_{1/2}\left(\tfrac{n-2m+2}{2},\tfrac{n+2m+2}{2}\right)\right]\,, \end{aligned} \tag{A.12}$$

where $B_x(a,b)=\int_0^x t^{a-1}(1-t)^{b-1}dt$ is the incomplete beta function [Abramowitz and Stegun, 1972].

APPENDIX B

Technical details of Chapter 5

B.1 Covariance and structure of $\mathscr{L}$

We start with a POVM element of the form $\bar{E}_0 = \int du\, U\, E_0\, U^\dagger$. Since D_μ must be a rank-one projector, it can always be written as $D_\mu = u_\mu\, [\uparrow]\, u_\mu^\dagger$ for a suitable SU(2) rotation u_μ. Thus,

$$\bar{E}_0 = \sum_\mu \int du \left(U_{AC} L_\mu U_{AC}^\dagger \right) \otimes \left(u u_\mu [\uparrow] u_\mu^\dagger u^\dagger \right).$$

We next use the invariance of the Haar measure du to make the change of variable $u\, u_\mu \to u'$ and, accordingly, $U_{AC} \to U'_{AC} U^\dagger_{\mu\, AC}$. After regrouping terms we have

$$\begin{aligned}
\bar{E}_0 &= \sum_\mu \int du' \left(U'_{AC} U^\dagger_{\mu\, AC} L_\mu U_{\mu\, AC} U'^\dagger_{AC} \right) \otimes \left(u' [\uparrow] u'^\dagger \right) \\
&= \int du' \left[U'_{AC} \left(\sum_\mu U^\dagger_{\mu\, AC} L_\mu U_{\mu\, AC} \right) U'^\dagger_{AC} \right] \otimes \left(u' [\uparrow] u'^\dagger \right) \\
&= \int du \left(U_{AC}\, \Omega\, U_{AC}^\dagger \right) \otimes \left(u [\uparrow] u^\dagger \right), \qquad \text{(B.1)}
\end{aligned}$$

where we have defined

$$\Omega = \sum_\mu U^\dagger_{\mu\, AC} L_\mu U_{\mu\, AC} \geq 0\,.$$

The POVM element $\bar{E}_1$ is obtained by replacing $[\uparrow]$ by $[\downarrow]$ in the expressions above. From the POVM condition $\sum_\mu L_\mu = \mathbb{1}_{AC}$ it immediately follows that

$$\int du U_{AC} \Omega U_{AC}^\dagger = \mathbb{1}_{AC} \,,$$

where $\mathbb{1}_{AC}$ is the identity on the Hilbert space of the TS, i.e., $\mathbb{1}_{AC} = \mathbb{1}_A \otimes \mathbb{1}_C$. Therefore $\mathscr{L} = \{U_{AC}\, \Omega\, U_{AC}^\dagger\}_{\mathrm{SU}(2)}$ is a covariant POVM. The positive operator Ω is called the seed of the covariant POVM $\mathscr{L}$.

Now, let $u_z(\varphi)$ be a rotation about the z-axis, which leaves $[\uparrow]$ invariant. By performing the change of variables $u \to u' u_z(\varphi)$ [and $U_{AC} \to U'_{AC} U_{zAC}(\varphi)$] in Eq. (B.1), we readily see that Ω and $U_{zAC}(\varphi)\, \Omega\, U_{zAC}^\dagger(\varphi)$ both give the same average operator $\bar{E}_0$ for any $\varphi \in [0, 4\pi)$. So, its average over φ,

$$\int_0^{4\pi} \frac{d\varphi}{4\pi} U_z(\varphi)\, \Omega\, U_z^\dagger(\varphi),$$

can be used as a seed without loss of generality, where we have dropped the subscript AC to simplify the notation. Such a seed is by construction invariant under the group of rotations about the z-axis (just like $[\uparrow]$) and, by Schur's lemma, a direct sum of operators with well defined magnetic number. Therefore, in the total angular momentum basis for AC, we can always choose the seed of $\mathscr{L}$ as

$$\Omega = \sum_{m=-n}^{n} \Omega_m \,; \quad \Omega_m \geqslant 0.$$

The constraint (5.8) follows from the POVM condition $\mathbb{1}_{AC} = \int du\, U\, \Omega\, U^\dagger$ and Schur's lemma. The result also holds if A and C have different number of copies (provided they add up to $2n$). It also holds for mixed states.

B.2 Overlaps

For the proof of optimality of the LM, we couple subsystems A, B and C in two ways: $A(CB)$ and $(AC)B$ to produce the states $|j_A, (j_C\, j_B) j_{CB}; J, M\rangle$ and $|(j_A\, j_C) j_{AC}, j_B; J, M\rangle$, which we denote by $|J, M\rangle_{A(CB)}$ and $|J, M\rangle_{(AC)B}$ respectively for short. The various angular momenta involved are fixed to $j_A = j_C = \frac{n}{2}$, $j_B = \frac{1}{2}$, $j_{AC} = j$, $j_{CB} = \frac{n}{2} + \frac{1}{2}$, whereas $J = j \pm \frac{1}{2}$. With these values, the general expression (A.1) gives us the overlaps that we need:

$${}_{A(CB)}\langle j \pm \tfrac{1}{2}, \tfrac{1}{2} | j \pm \tfrac{1}{2}, \tfrac{1}{2} \rangle_{(AC)B} = \sqrt{\frac{n + \frac{3}{2} \pm (j + \frac{1}{2})}{2(n+1)}} \,.$$

B.3 Measurement of a block-diagonal $\boldsymbol{\rho}^{\otimes \boldsymbol{n}}$

The state $\rho^{\otimes n}$ of n identical copies of a general qubit state ρ with purity r and Bloch vector $r\boldsymbol{s}$, has a block diagonal form in the basis of the total angular momentum (see Section 3.4.1) given by

$$\rho^{\otimes n} = \sum_j p_j^n \rho_j \otimes \frac{\mathbb{1}_j}{\nu_j^n}.$$

Here $j = 0\,(1/2), \ldots, n/2$ if n is even (odd), $\mathbb{1}_j$ is the identity in the multiplicity space $\mathbb{C}^{\nu_j^n}$, of dimension ν_j^n (the multiplicity of the representation with total angular momentum j), where

$$\nu_j^n = \binom{n}{n/2-j} \frac{2j+1}{n/2+j+1}.$$

The normalized state ρ_j, which is supported on the representation subspace $\mathscr{S}_j = \operatorname{span}\{|j,m\rangle\}$ of dimension $2j+1 = d_{2j}$, is

$$\rho_j = U_{\boldsymbol{s}} \left(\sum_{m=-j}^{j} a_m^j \; [j,m] \right) U_{\boldsymbol{s}}^\dagger,$$

where

$$a_m^j = \frac{1}{c_j} \left(\frac{1-r}{2}\right)^{j-m} \left(\frac{1+r}{2}\right)^{j+m}, \tag{B.2}$$

and

$$c_j = \frac{1}{r} \left\{ \left(\frac{1+r}{2}\right)^{2j+1} - \left(\frac{1-r}{2}\right)^{2j+1} \right\},$$

so that $\sum_{m=-j}^{j} a_m^j = 1$, and we stick to our shorthand notation $[\,\cdot\,] \equiv |\,\cdot\,\rangle\langle\,\cdot\,|$, i.e., $[j,m] \equiv |j,m\rangle\langle j,m|$. The measurement on $\rho^{\otimes n}$ defined by the set of projectors on the various subspaces $\mathscr{S}_j$ will produce ρ_j as a posterior state with probability

$$p_j^n = \nu_j^n c_j \left(\frac{1-r^2}{4}\right)^{n/2-j}.$$

One can easily check that $\sum_j p_j^n = 1$.

In the large n limit, we can replace p_j^n for a continuous probability distribution $p_n(x)$ in $[0,1]$, where $x = 2j/n$. Applying Stirling's approximation to p_j one obtains:

$$p_n(x) \simeq \sqrt{\frac{n}{2\pi}} \frac{1}{\sqrt{1-x^2}} \frac{x(1+r)}{r(1+x)} \; \mathrm{e}^{-nH(\frac{1+x}{2} \| \frac{1+r}{2})},$$

where $H(s \parallel t)$ is the (binary) relative entropy

$$H(s \parallel t) = s \log \frac{s}{t} + (1-s) \log \frac{1-s}{1-t}.$$

The approximation is valid for x and r both in the open unit interval $(0,1)$. For non-vanishing r, $p_n(x)$ becomes a Dirac delta function peaked at $x = r$, $p_\infty(x) = \delta(x-r)$, which corresponds to $j = nr/2$.

B.4 Derivation of Eqs. (5.30) and (5.34)

Let us start with the general case where $\xi = \{j, j'\}$. To obtain $\sigma^n_{0,\xi}$ we first write Eqs. (5.27) and (5.28) as the SU(2) group integrals

$$\begin{aligned}\sigma^n_{0,\xi} &= \int du\, U_{AB} \left(\sum_{m=-j}^{j} a^j_m [j,m]_A \otimes \rho^B_0 \right) U^\dagger_{AB} \\ &\qquad \otimes \int du'\, U'_C \left(\sum_{m=-j'}^{j'} a^{j'}_m [j',m]_C \right) U'^\dagger_C \,,\end{aligned}$$

where a^j_m is given in Eq. (B.2), and ρ^B_0 is the mixed state ρ_0, Eq. (5.26), of the qubit B. We next couple A with B (more precisely, their subspaces of angular momentum j) using the Clebsch-Gordan coefficients

$$\begin{aligned}|\langle j+\tfrac{1}{2}, m+\tfrac{1}{2} | j,m; \tfrac{1}{2}, \tfrac{1}{2}\rangle|^2 &= \frac{j+m+1}{2j+1}\,, \\ |\langle j-\tfrac{1}{2}, m+\tfrac{1}{2} | j,m; \tfrac{1}{2}, \tfrac{1}{2}\rangle|^2 &= \frac{j-m}{2j+1}\,.\end{aligned}$$

The resulting expressions can be easily integrated using Schur's lemma. Note that the integrals of crossed terms of the form $|j,m\rangle\langle j',m|$ will vanish for all $j \neq j'$. We readily obtain

$$\sigma^n_{0,\xi} = \sum_{m=-j}^{j} a^j_m \left(\frac{j+1+mr}{d_{2j}} \frac{\mathbb{1}^{AB}_{2j+1}}{d_{2j+1}} + \frac{j-mr}{d_{2j}} \frac{\mathbb{1}^{AB}_{2j-1}}{d_{2j-1}} \right) \otimes \frac{\mathbb{1}^C_{2j'}}{d_{2j'}}\,,$$

where $\mathbb{1}_{2j}$ is the projector on $\mathscr{S}_j$ and $d_{2j} = 2j+1 = \dim \mathscr{S}_j$. The superscripts attached to the various projectors specify the subsystems to which they refer. These projectors are formally equal to those used in Eq. (5.2) (i.e., $\mathbb{1}_{2j}$ projects onto the fully symmetric subspace of $2j$ qubits), hence we stick to the same notation. Note that $\operatorname{tr} \sigma^n_{0,\xi} = 1$, as it should be.

We can further simplify this expression by introducing $\langle\hat{J}_z\rangle_j = \sum_m m\, a^j_m$, i.e., the expectation value of the z-component of the total angular momentum in the state ρ_j (i.e., of $\mathbb{1}_{2j}\hat{J}_z\mathbb{1}_{2j}$ in the state $\rho^{\otimes n}_{0/1}$) for a Bloch vector $r\boldsymbol{z}$:

$$\sigma^n_{0,\xi} = \left(\frac{j+1+r\langle\hat{J}_z\rangle_j}{d_{2j}}\,\frac{\mathbb{1}^{AB}_{2j+1}}{d_{2j+1}} + \frac{j-r\langle\hat{J}_z\rangle_j}{d_{2j}}\,\frac{\mathbb{1}^{AB}_{2j-1}}{d_{2j-1}}\right)\otimes\frac{\mathbb{1}^{C}_{2j'}}{d_{2j'}}.$$

Using the relation

$$\mathbb{1}^{AB}_{2j-1} = \mathbb{1}^{A}_{2j}\otimes\mathbb{1}^{B}_{1} - \mathbb{1}^{AB}_{2j+1},$$

and $(j+1)/d_{2j+1} = j/d_{2j-1} = 1/2$, we can write

$$\sigma^n_{0,\xi} = \left(\frac{r\langle\hat{J}_z\rangle_j}{j}\,\frac{\mathbb{1}^{AB}_{2j+1}}{d_{2j+1}} + \frac{j-r\langle\hat{J}_z\rangle_j}{j}\,\frac{\mathbb{1}^{A}_{2j}}{d_{2j}}\otimes\frac{\mathbb{1}^{B}_{1}}{2}\right)\otimes\frac{\mathbb{1}^{C}_{2j'}}{d_{2j'}}. \tag{B.3}$$

Similarly, we can show that

$$\sigma^n_{1,\xi} = \frac{\mathbb{1}^{A}_{2j}}{d_{2j}}\otimes\left(\frac{r\langle\hat{J}_z\rangle_{j'}}{j'}\,\frac{\mathbb{1}^{BC}_{2j'+1}}{d_{2j'+1}} + \frac{j'-r\langle\hat{J}_z\rangle_{j'}}{j'}\,\frac{\mathbb{1}^{B}_{1}}{2}\otimes\frac{\mathbb{1}^{C}_{2j'}}{d_{2j'}}\right). \tag{B.4}$$

Therefore, if $j' = j$,

$$\sigma^n_{0,\xi} - \sigma^n_{1,\xi} = \frac{r\langle\hat{J}_z\rangle_j}{j}\left(\frac{\mathbb{1}^{AB}_{2j+1}}{d_{2j+1}}\otimes\frac{\mathbb{1}^{C}_{2j}}{d_{2j}} - \frac{\mathbb{1}^{A}_{2j}}{d_{2j}}\otimes\frac{\mathbb{1}^{BC}_{2j+1}}{d_{2j+1}}\right).$$

Comparing with Eq. (5.2), the two terms in the second line can be understood as the average states for a number of $2j$ pure qubits, i.e., as σ^{2j}_0 and σ^{2j}_1 respectively. Hence, if $\xi = \{j,j\}$ we have the relation

$$\sigma^n_{0,\xi} - \sigma^n_{1,\xi} = \frac{r\langle\hat{J}_z\rangle_j}{j}\left(\sigma^{2j}_0 - \sigma^{2j}_1\right),$$

which is Eq. (5.30). It is important to emphasize that this equation is exact (i.e., it holds for any value of j, n and r) and bears no relation whatsoever to measurements, for it is just an algebraic identity between the various operators involved.

In the asymptotic limit, for n_A and n_C of the form $n_{A/C} \simeq n \pm bn^a$, $n \gg 1$, $a < 1$, the probabilities p^n_j and $p^n_{j'}$ are peaked at $j \simeq rn_A/2$ and $j' \simeq rn_C/2$, as was explained above. Hence only the average state components $\sigma^n_{0/1,\xi}$ with $\xi = \{j,j'\}$ such that $j \simeq (r/2)n(1+bn^{a-1})$ and $j' \simeq (r/2)n(1-bn^{a-1})$ are important. From Eqs. (B.3) and (B.4) it is straightforward to obtain

$$\sigma^n_{0,\xi} - \sigma^n_{1,\xi} \simeq r\left(1 - \frac{1-r}{nr^2}\right)\left(\sigma^{rn}_0 - \sigma^{rn}_1\right) + o(n^{-1}),$$

where we have used that [Gendra *et al.*, 2012] $\langle\hat{J}_z\rangle_j \simeq j-(1-r)/(2r)$ up to exponentially vanishing terms. This relation, for the particular value of $a=1/2$, is used in the proof of robustness, Eq. (5.34).

B.5 Calculation of $\boldsymbol{\Gamma}_\uparrow$

Here we calculate $\Gamma_{\uparrow,\xi}=\mathrm{tr}_B\{[\uparrow](\sigma^n_{0,\xi}-\sigma^n_{1,\xi})\}$, where the average states are defined in Eqs. (5.27) and (5.28), and explicitly given in Eqs. (B.3) and (B.4) for $\xi=\{j,j'\}$. Let us first calculate the conditional state $\mathrm{tr}_B([\uparrow]\sigma^n_{0,\xi})$. For that, we need to express $\mathbb{1}^{AB}_{2j+1}=\sum_m[j+\frac{1}{2},m]$ in the original product basis $\{|j_A,m_A\rangle\otimes|\uparrow/\downarrow\rangle\}$. Recalling the Clebsch-Gordan coefficients $|\langle\frac{1}{2},\frac{1}{2};j,m|j+\frac{1}{2},m+\frac{1}{2}\rangle|^2=(j+m+1)/(2j+1)$, one readily obtains

$$\mathrm{tr}_B\left([\uparrow]\frac{\mathbb{1}^{AB}_{2j+1}}{d_{2j+1}}\right)=\sum_{m=-j}^{j}\frac{j+1+m}{2(j+1)d_{2j}}[j,m]_A,$$

which can be written as

$$\mathrm{tr}_B\left([\uparrow]\frac{\mathbb{1}^{AB}_{2j+1}}{d_{2j+1}}\right)=\frac{1}{2}\left(\frac{\mathbb{1}^{A}_{2j}}{d_{2j}}+\frac{1}{d_{2j}}\frac{\hat{J}^A_z}{j+1}\right),$$

where $\hat{J}^A_z$ is the z component of the total angular momentum operator acting on subsystem A. An analogous expression is obtained for $\mathrm{tr}_B\left([\uparrow]\mathbb{1}^{BC}_{2j'+1}\right)$. Substituting in Eqs. (B.3) and (B.4) and subtracting the resulting expressions, one has $\Gamma_\uparrow=\sum_\xi p^n_\xi\Gamma_{\uparrow,\xi}$, with

$$\Gamma_{\uparrow,\xi}=\frac{1}{2d_{2j_A}d_{2j_C}}\left(\frac{r\langle\hat{J}_z\rangle_{j_A}}{j_A}\frac{\hat{J}^A_z}{j_A+1}-\frac{r\langle\hat{J}_z\rangle_{j_C}}{j_C}\frac{\hat{J}^C_z}{j_C+1}\right),\tag{B.5}$$

where we have written $\xi=\{j_A,j_C\}$, instead of $\xi=\{j,j'\}$ used in the derivation. For pure states, $r=1$, $j_A=j_C=n/2$, $\langle\hat{J}_z\rangle_{n/2}=n/2$, and we recover Eq. (5.11).

In order to minimize the excess risk using SDP, we find it convenient to write Eq. (5.9) in the form

$$\Delta^{\mathrm{LM}}=2\max_{\{\Omega_{m,\xi}\}}\sum_\xi p^n_\xi\mathrm{tr}\,(\Gamma_{\uparrow,\xi}\Omega_{m,\xi}),\tag{B.6}$$

where we recall that $m=m_{AC}=m_A+m_C$, and we assumed w.l.o.g. that the seed of the optimal POVM has the block form $\Omega_m=\sum_\xi\Omega_{m,\xi}$. The POVM condition, Eq. (5.8) must now hold on each block, thus for $\xi = \{j_A, j_C\}$, we must impose that

$$\sum_{m=-j}^{j} \langle j, m|\Omega_{m,\xi}|j, m\rangle = 2j + 1, \ |j_A - j_C| \leqslant j \leqslant j_A + j_C. \tag{B.7}$$

APPENDIX C

Continuous-variables systems

A continuous-variables (CV) system is a bosonic system described by a Hilbert space of infinite dimension. CV systems provide the appropriate description of the states of light, and they have earned an outstanding role in quantum information and communication, as quantum optical settings allow to successfully implement, with current technology, quantum processing tasks such as quantum teleportation [Furusawa *et al.*, 1998], quantum key distribution [Grosshans *et al.*, 2003], and quantum dense coding [Li *et al.*, 2002]. Special tools are required for describing this type of systems. The purpose of this Section is to give an overview on the formalism of CV systems that underlies in Chapter 6. For more complete reviews on the topic, see [Braunstein and van Loock, 2005; Eisert and Plenio, 2003; Cerf, 2007].

A CV system of N canonical bosonic modes is described by a Hilbert space $\mathcal{H} = \bigotimes_{i=1}^{N} \mathcal{H}_i$, resulting from the tensor product structure of infinite dimensional spaces $\mathcal{H}_i$, each of them associated to a single mode. Each mode is described by a pair of canonical conjugate operators $\hat{q}_i$ and $\hat{p}_i$, acting on $\mathcal{H}_i$. These operators may correspond, for instance, to position and momentum operators associated to a second quantized electromagnetic field, which Hamiltonian

$$\hat{H} = \sum_{i=1}^{N} \hbar\omega_i \left(\hat{a}_i^\dagger \hat{a}_i + \frac{1}{2} \right) \tag{C.1}$$

describes a system of N noninteracting harmonic oscillators with different frequencies ω_i, the *modes* of the field. Another example susceptible of a canonical description is the collective spin of a polarized ensemble of atoms

[Julsgaard *et al.*, 2001]. The ladder operators $\hat{a}_k$ and $\hat{a}_k^\dagger$ relate to the quadrature phase operators (position and momentum) according to

$$\hat{q}_k = \frac{\hat{a}_k + \hat{a}_k^\dagger}{\sqrt{2}}\,, \quad \hat{p}_k = \frac{\hat{a}_k - \hat{a}_k^\dagger}{i\sqrt{2}}\,, \tag{C.2}$$

and they obey the canonical commutation relation (CCR)

$$[\hat{a}_k, \hat{a}_l^\dagger] = \delta_{kl}\,, \quad [\hat{a}_k, \hat{a}_l] = [\hat{a}_k^\dagger, \hat{a}_l^\dagger] = 0\,, \tag{C.3}$$

which, in terms of $\hat{q}_k$ and $\hat{p}_k$, reads[1]

$$[\hat{q}_k, \hat{p}_k] = i\mathbb{1}_k\,, \tag{C.4}$$

where $\mathbb{1}_k$ is the identity operator on mode k. The canonical operators of all modes of the system can be grouped in the vector $\hat{R} = (\hat{q}_1, \hat{p}_1, \ldots, \hat{q}_N, \hat{p}_N)^T$. In this notation, the CCR (C.4) reads

$$[\hat{R}_k, \hat{R}_l] = i\Omega_{kl}\,, \tag{C.5}$$

where $k, l = 1, 2, \ldots, 2N$, and Ω is the symplectic matrix

$$\Omega = \bigoplus_{i=1}^{N} \omega\,, \quad \omega = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}\,. \tag{C.6}$$

C.1 The phase-space picture

The states of a CV system are the set of positive trace-class operators $\{\rho\}$ on the Hilbert space $\mathcal{H} = \bigotimes_{i=1}^N \mathcal{H}_i$. The complete description of any state ρ of such an infinite-dimensional system can be conveniently provided by the (0-ordered) *characteristic function*

$$\chi(\xi) = \mathrm{tr}\,(\rho \hat{D}_\xi)\,, \tag{C.7}$$

where $\xi \in \mathbb{R}^{2N}$, and $\hat{D}_\xi$ is a Weyl operator (see below). The vector ξ belongs to the $2N$-dimensional real vector space $\Gamma(\mathbb{R}^{2N}, \Omega)$ called *phase space*. From the form of Eq.(C.7) one can readily see that the tensor-product structure of the Hilbert space is replaced by a direct sum structure in the phase space, such that $\Gamma = \bigoplus_{i=1}^N \Gamma_i$, where $\Gamma_i(\mathbb{R}^2, \omega)$ is the local phase space of mode i.

[1] The canonical operators are chosen to be adimensional, hence $\hbar$ does not appear explicitly in any of the equations.

The Weyl operator $\hat{D}_\xi$ acts in the states as a translation in the phase space. It is defined as

$$\hat{D}_\xi = e^{-i\hat{R}^T \Omega \xi} \,, \tag{C.8}$$

and its action over an arbitrary vector of canonical operators $\hat{R}$ yields

$$\hat{D}_\xi^\dagger \hat{R}_i \hat{D}_\xi = \hat{R}_i - \xi_i \mathbb{1} \,. \tag{C.9}$$

The characteristic function $\chi(\xi)$ is related, via a Fourier transform, to the so-called *Wigner function*

$$W(\xi) = \frac{1}{(2\pi)^{2N}} \int_{\mathbb{R}^{2N}} d^{2N}\kappa \chi(\kappa) e^{i\kappa^T \Omega \xi} \,, \tag{C.10}$$

that constitutes an alternative complete description of quantum states for CV systems. The Wigner function is a real-valued *quasi-probability distribution*. This denomination is motivated from the fact that the function $W(\xi)$ might be negative or ill-behaved in certain regions of the phase space, and nevertheless it quantifies the probability with which one might expect to obtain the values ξ upon measuring simultaneously the canonical operators $\hat{R}$. There exist, however, alternative ways of defining quasi-probability distributions for CV states for which the Wigner function is not an appropriate description. These variations are derived from alternative definitions of the characteristic function [Leonhardt, 1997]. The following properties are worth remarking:

1. $W(\xi)$ is normalized, i.e.,

$$\int_{\mathbb{R}^{2N}} d^{2N}\kappa W(\kappa) = \operatorname{tr}\rho = \chi(0) = 1 \,. \tag{C.11}$$

2. In terms of $W(\xi)$, the purity of a state ρ is expressed as

$$\int_{\mathbb{R}^{2N}} d^{2N}\kappa W^2(\kappa) = \int_{\mathbb{R}^{2N}} d^{2N}\xi |\chi(\xi)|^2 = \operatorname{tr}\rho^2 = \mu \,. \tag{C.12}$$

3. The overlap between two states ρ_1 and ρ_2 corresponds to

$$\operatorname{tr}(\rho_1\rho_2) = 2\pi \int_{\mathbb{R}^{2N}} d^{2N}\kappa W_1(\kappa) W_2(\kappa) \,. \tag{C.13}$$

The phase-space formulation offers the theoretical tools to map states and operations of infinite-dimensional CV systems into relations in finite real spaces. Both the density matrix and the Wigner function provide a complete description of the state of a CV system, hence a one-to-one correspondence between them exists. For a single mode state, i.e., $\xi = (q, p)$, it is of the form

$$W(q,p) = \frac{1}{\pi} \int_{-\infty}^{\infty} dx \, \langle q + x|\rho|q - x\rangle \, e^{-2ipx} \,, \tag{C.14}$$

which is Wigner's legendary formula [Wigner, 1932].

C.2 The states of light

As stated above, the (N-mode) electromagnetic field, the paradigm of CV systems, can be modelled by the Hamiltonian of N noninteracting harmonic oscillators given in Eq. (C.1). The states of the harmonic oscillator associated to the ith mode belong to the Hilbert space $\mathcal{H}_i$, and this space is spanned by the eigenstates of the number operator $\hat{n}_i = \hat{a}_i^\dagger \hat{a}_i$ that represents the corresponding Hamiltonian. These states form the so-called Fock basis $\{|n\rangle_i\}$, verifying

$$\hat{n}_i\,|n\rangle_i = n_i\,|n\rangle_i\ , \tag{C.15}$$

where $n_i = 0, \ldots, \infty$ gives the quanta of excitations of mode i. The Hamiltonian of each mode is bounded from below, thus ensuring the stability of the system. For the ith mode, the ground state of the oscillator or *vacuum state* of the field is that which is annihilated by the operator $\hat{a}_i$, i.e., $\hat{a}_i\,|0\rangle_i = 0$. The vacuum state of the global Hilbert space is just $|0\rangle = \bigotimes_i |0\rangle_i$. The Fock state $|n\rangle_i$ can be regarded as the nth excitation (photon) of the vacuum of mode i, obtained by the action of the annihilation ($\hat{a}_i$) and creation ($\hat{a}_i^\dagger$) operators (recall that $\hat{a}^\dagger\,|n\rangle = \sqrt{n+1}\,|n+1\rangle$ and $\hat{a}\,|n\rangle = \sqrt{n}\,|n-1\rangle$), i.e.,

$$|n\rangle_i = \frac{(\hat{a}_i^\dagger)^n}{\sqrt{n_i!}}\,|0\rangle_i \tag{C.16}$$

The Fock states, with the exception of the vacuum, belong to the broader class of non-Gaussian states. In general, non-Gaussian states are difficult to handle, both mathematically and experimentally. By contrast, Gaussian states exhibit much nicer properties and comprise an extremely relevant class of CV states, since the vast majority of the states prepared in quantum optics laboratories are of this type[2].

The set of *Gaussian states* is, by definition, the set of states with Gaussian characteristic functions and quasi-probability distributions on the multimode quantum phase space. Gaussian states include, among others, coherent, squeezed, and thermal states. From its very definition, it follows that a Gaussian state ρ is completely characterized by the first and second statistical moments of the quadrature field operators, embodied in the vector of first moments $\bar{R}$ and the *covariance matrix* (CM) $\boldsymbol{\sigma}$, respectively, which elements are

$$\begin{aligned} \bar{R}_i &= \langle \hat{R}_i \rangle\,, & \text{(C.17)} \\ \sigma_{ij} &= \langle \hat{R}_i\hat{R}_j + \hat{R}_j\hat{R}_i \rangle - 2\langle \hat{R} \rangle_i \langle \hat{R}_j \rangle\,, & \text{(C.18)} \end{aligned}$$

[2] For a review on the uses of Gaussian states in quantum information applications, see [Weedbrook *et al.*, 2012].

and where $i, j = 1, \ldots, 2N$. The Wigner function of a Gaussian state ρ has the form

$$W(X) = \frac{1}{\pi^N \sqrt{\det \boldsymbol{\sigma}}} e^{-(X-\bar{R})\boldsymbol{\sigma}^{-1}(X-\bar{R})^T} , \tag{C.19}$$

where X stands for the real phase-space vector $(q_1, p_1, \ldots, q_n, p_n) \in \Gamma$.

The vector of first moments $\bar{R}$ can be arbitrarily adjusted by local unitary operations, namely displacements in phase space by means of Weyl operators (C.8). Since the reduced state resulting from a partial trace operation over a subset of modes of a Gaussian state is still Gaussian, one can apply single-mode Weyl operators to locally re-center each such reduced Gaussian. Such operations leave all the informationally relevant properties of the state invariant, hence in general the first moments can be adjusted to 0 without loss of generality. It follows that, despite the infinite dimension of the associated Hilbert space, the complete description of an arbitrary Gaussian state (up to local unitary operations) is given by its $2N \times 2N$ CM $\boldsymbol{\sigma}$. For a CM to describe a proper physical state, it must verify the condition

$$\boldsymbol{\sigma} + i\Omega \geqslant 0 , \tag{C.20}$$

analogous to the semidefinite-positive condition for the density matrix $\rho \geqslant 0$.

Generically, a N-mode Gaussian state has a CM $\boldsymbol{\sigma}$ that can be written in terms of 2×2 submatrices as

$$\boldsymbol{\sigma} = \begin{pmatrix} \boldsymbol{\sigma}_1 & \boldsymbol{\epsilon}_{1,2} & \cdots & \boldsymbol{\epsilon}_{1,N} \\ \boldsymbol{\epsilon}^T_{1,2} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \boldsymbol{\epsilon}_{N-1,N} \\ \boldsymbol{\epsilon}^T_{1,N} & \cdots & \boldsymbol{\epsilon}^T_{N-1,N} & \boldsymbol{\sigma}_N \end{pmatrix} . \tag{C.21}$$

The diagonal block $\boldsymbol{\sigma}_i$ is the local CM of the corresponding reduced state of mode i. On the other hand, the off-diagonal matrices $\boldsymbol{\epsilon}_{i,j}$ encode the intermodal correlations (both classical and quantum) between modes i and j. A product state has no off-diagonal terms, hence its CM is simply the direct sum of the local CMs. Properties like the entanglement of a state and its purity, and linear transformations of first moments in phase space (symplectic transformations), can all be described within the CM formalism.

The three most important types of single-mode Gaussian states are coherent, squeezed, and thermal states.

C.2.1 Coherent states

Coherent states are the states produced by an ideal laser. They are ubiquitous in CV quantum information, and, among all CV states, their dynamics

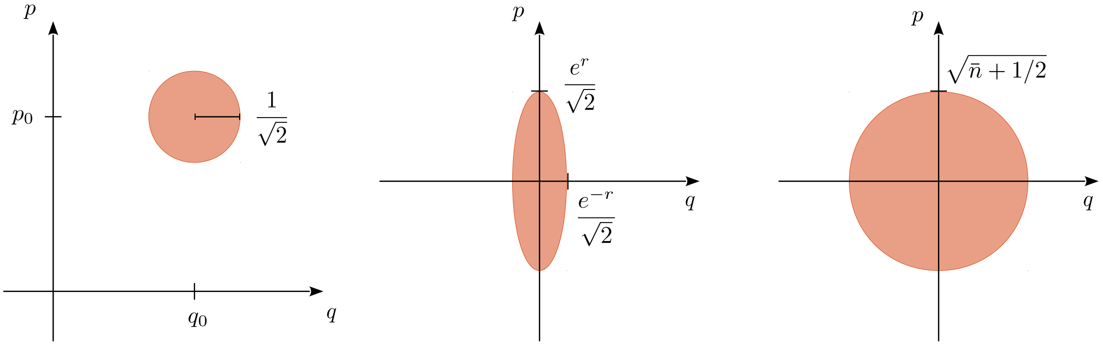


Figure C.1. From left to right: a coherent state of amplitude $\alpha = (q_0+ip_0)/\sqrt{2}$, a squeezed vacuum state with a squeezing parameter r, and a thermal state with average photon number $\bar{n}$.

is the one that most resembles the behavior of a classical electromagnetic field. Coherent states have minimal quantum uncertainty, which means that fluctuations are symmetrically distributed between its quadratures.

Coherent states can be defined as the eigenstates of the annihilation operator $\hat{a}$

$$\hat{a}\,|\alpha\rangle = \alpha\,|\alpha\rangle \ , \tag{C.22}$$

where the eigenvalue α, in general complex, is the *amplitude* of the state $|\alpha\rangle$, and it is related to the quadratures through

$$\alpha = \frac{q+ip}{\sqrt{2}} \,, \tag{C.23}$$

i.e., $q = \sqrt{2}\mathrm{Re}(\alpha)$ and $p = \sqrt{2}\mathrm{Im}(\alpha)$. The state $|\alpha\rangle$ results from applying the single-mode displacement operator $\hat{D}(\alpha)$ to the vacuum, that is

$$|\alpha\rangle = \hat{D}(\alpha)\,|0\rangle \ , \tag{C.24}$$

where

$$\hat{D}(\alpha) = e^{\alpha\hat{a}^\dagger - \alpha^*\hat{a}} = e^{-|\alpha|^2/2} e^{\alpha\hat{a}^\dagger} e^{-\alpha^*\hat{a}} \,. \tag{C.25}$$

The displacement operator can be identified with the single-mode Weyl operator (C.8) by using the relations in Eq. (C.2). In the Heisenberg picture, the action of $\hat{D}(\alpha)$ over the operator $\hat{a}$ yields the displacement

$$\hat{D}^\dagger(\alpha)\,\hat{a}\,\hat{D}(\alpha) = \hat{a} + \alpha \,. \tag{C.26}$$

Another useful property is

$$\hat{D}^\dagger(\alpha)\hat{D}(\beta) = e^{-\frac{1}{2}(\alpha\beta^* - \beta\alpha^*)}\hat{D}(\beta-\alpha) \,. \tag{C.27}$$

One can use the definition of the displacement operator, that is Eq. (C.25), to express a coherent state in terms of Fock states:

$$\begin{aligned}|\alpha\rangle &= \hat{D}(\alpha)\,|0\rangle = e^{-|\alpha|^2/2}e^{\alpha\hat{a}^\dagger}e^{-\alpha^*\hat{a}}\,|0\rangle \\ &= e^{-|\alpha|^2/2}\sum_{n=0}^{\infty}\frac{(\alpha\hat{a}^\dagger)^n}{n!}\,|0\rangle \\ &= e^{-|\alpha|^2/2}\sum_{n=0}^{\infty}\frac{\alpha^n}{n!}\,|n\rangle\ . \end{aligned} \tag{C.28}$$

The Fock representation (C.28) shows that a coherent state has the Poissonian photon statistics

$$P(n) = |\langle n|\alpha\rangle|^2 = \frac{|\alpha|^{2n}}{n!}e^{-|\alpha|^2}\,. \tag{C.29}$$

Note that the average photon number, or *intensity*, of a coherent state is $\langle\hat{n}\rangle = \langle\hat{a}^\dagger\hat{a}\rangle = |\alpha|^2$.

The overlap between two coherent states $|\alpha\rangle$ and $|\beta\rangle$ can be readily seen to give, by means of Eq. (C.27),

$$\begin{aligned}\langle\alpha|\beta\rangle &= \langle 0|\hat{D}^\dagger(\alpha)\hat{D}(\beta)|0\rangle \\ &= e^{-\frac{1}{2}(\alpha\beta^*-\beta\alpha^*)}\,\langle 0|\hat{D}(\beta-\alpha)|0\rangle \\ &= e^{-\frac{1}{2}(\alpha\beta^*-\beta\alpha^*)}\langle 0|\beta-\alpha\rangle \\ &= e^{-|\alpha|^2/2-|\beta|^2/2+\alpha^*\beta}\,, \end{aligned} \tag{C.30}$$

hence

$$|\langle\alpha|\beta\rangle|^2 = e^{-|\alpha-\beta|^2}\,. \tag{C.31}$$

Eq. (C.31) shows that two coherent states approach orthogonality only when their amplitude difference is large. Despite being nonorthogonal, coherent states form a basis in $\mathcal{H}$ (which is an *overcomplete* basis, for this very reason), and fulfil the completeness relation

$$\frac{1}{\pi}\int d^2\alpha\,|\alpha\rangle\langle\alpha| = \mathbb{1}\,. \tag{C.32}$$

The characteristic function of a coherent state can be straightforwardly obtained through Eqs. (C.7) and (C.25). One can then compute the corresponding Wigner function using Eq. (C.10), and compare the result with the Wigner function of a general Gaussian state, given by Eq. (C.19). This analysis shows that a coherent state $|\alpha\rangle$ has a displacement vector $\bar{R} = (q,p)$, and a CM $\boldsymbol{\sigma} = \mathbb{1}$. This means that it has the same minimal fluctuations as the vacuum state, but displaced in phase space. Thus a coherent state can be depicted as a displaced circle of radius $1/\sqrt{2}$ in phase space (see Fig. C.1).

C.2.2 Squeezed states

Squeezed states are states that have an asymmetrical distribution of fluctuations among their quadratures. That means to say, it is possible to reduce the uncertainty in one of the quadratures of a state, but this comes always at the expense of an increase in the noise of its conjugate variable, in accordance to Heisenberg's uncertainty principle. The preparation procedure of a squeezed state, that is the squeezing transformation, uses nonlinear optic elements and does not conserve the total photon number.

The single-mode squeezing operator is described by

$$\hat{S}(r,\phi) = e^{\frac{r}{2}\left(\hat{a}^2 e^{-2i\phi} - \hat{a}^{\dagger 2} e^{2i\phi}\right)}, \tag{C.33}$$

where $r \geqslant 0$ is the *squeezing parameter*. Its effect over the operators $\hat{a}$ and $\hat{a}^\dagger$ is

$$\begin{aligned} \hat{S}^\dagger(r,\phi)\hat{a}\hat{S}(r,\phi) &= \hat{a}\cosh r - \hat{a}^\dagger e^{i\phi}\sinh r & \text{(C.34)} \\ \hat{S}^\dagger(r,\phi)\hat{a}^\dagger\hat{S}(r,\phi) &= \hat{a}^\dagger\cosh r - \hat{a}e^{-i\phi}\sinh r\,. & \text{(C.35)} \end{aligned}$$

Applied instead to the rotated quadrature operators $\hat{q}_\phi = \hat{q}\cos\phi + \hat{p}\sin\phi$ and $\hat{p}_\phi = -\hat{q}\sin\phi + \hat{p}\cos\phi$, it yields

$$\begin{aligned} \hat{S}^\dagger(r,\phi)\hat{q}_\phi\hat{S}(r,\phi) &= \hat{q}_\phi e^{-r}\,, & \text{(C.36)} \\ \hat{S}^\dagger(r,\phi)\hat{p}_\phi\hat{S}(r,\phi) &= \hat{p}_\phi e^{r}\,. & \text{(C.37)} \end{aligned}$$

In the Fock representation, using Eq. (C.33) and taking $\phi = 0$ for simplicity, the action of $\hat{S}$ over a vacuum state $|0\rangle$ results in the squeezed vacuum state

$$|0,r\rangle = \hat{S}(r,0)\,|0\rangle = \frac{1}{\sqrt{\cosh r}}\sum_{n=0}^{\infty}\tanh^n r\,\frac{\sqrt{(2n)!}}{2^n n!}\,|2n\rangle\,. \tag{C.38}$$

The CM of the state $|0,r\rangle$ takes the simple form $\boldsymbol{\sigma} = \mathrm{diag}(e^{-2r}, e^{2r})$, which accounts for the difference in the quadrature variances $\Delta^2 q = \langle\hat{q}^2\rangle - \langle\hat{q}\rangle^2 = e^{-2r}/2$ and $\Delta^2 p = \langle\hat{p}^2\rangle - \langle\hat{p}\rangle^2 = e^{2r}/2$. A squeezed vacuum state is thus depicted in phase space as an ellipse with an area equal to that of a minimal uncertainty state, i.e., $\pi/2$ (see Fig. C.1).

C.2.3 Thermal states

The state of a single-mode field in thermal equilibrium with its environment is a thermal state, with density operator

$$\rho_{\mathrm{th}} = (1 - e^{-\beta})\sum_{n=0}^{\infty} e^{-\beta n}\,|n\rangle\langle n|\,, \tag{C.39}$$

where $\beta = \omega/k_B T$ denotes the ratio between the energy ω and the temperature T (k_B stands for the Boltzmann's constant). To justify Eq. (C.39), recall that in thermal equilibrium the density operator must be diagonal in the energy representation and that photons obey the Bose-Einstein statistics. The average photon number for the thermal state ρ_{th} is

$$\bar{n} = \operatorname{tr}(\rho_{\text{th}}\hat{n}) = (1 - e^{-\beta}) \sum_{n=0}^{\infty} n e^{-\beta n} = \frac{1}{e^{\beta} - 1}\,, \tag{C.40}$$

thus ρ_{th} can be expressed in terms of $\bar{n}$ as

$$\rho_{\text{th}} = \frac{1}{1+\bar{n}} \sum_{n=0}^{\infty} \left(\frac{\bar{n}}{1+\bar{n}}\right)^{n} |n\rangle\langle n|\;. \tag{C.41}$$

One can easily check that the Wigner function associated to the state ρ_{th} is Gaussian. The first moments vanish, hence its displacement is $\bar{R} = (0,0)$. Its CM is the diagonal matrix $\boldsymbol{\sigma} = (2\bar{n}+1)\mathbb{1}$. The form of $\boldsymbol{\sigma}$ tells us that a thermal state has symmetric variances of its quadratures, and that these are proportional to $\bar{n}$ and, in turn, dependent on the temperature T. Thus a thermal state is a symmetric state of greater than minimal uncertainty. It can be depicted in phase space as a circle of radius $\sqrt{\bar{n}+1/2}$ (see Fig. C.1).

The most general *mixed* Gaussian state is obtained by the sequential action of the squeezing (C.33) and displacement (C.25) operators on a thermal state (C.39):

$$\rho(\alpha, r, \phi) = \hat{D}(\alpha)\hat{S}(r,\phi)\rho_{\text{th}}\hat{S}^{\dagger}(r,\phi)\hat{D}^{\dagger}(\alpha)\,. \tag{C.42}$$

The most general *pure* Gaussian state is achieved by setting $\bar{n} = 0$. This corresponds to a rotated, squeezed and displaced state $|\alpha, r, \phi\rangle = \hat{D}(\alpha)\hat{S}(r,\phi)\,|0\rangle$.

C.3 The measurements of light

Quantum measurements of CV systems can be theoretically described by the POVM formalism. This is to say, one can describe a measurement by a set of positive-semidefinite operators $\{E_i\}$ such that $\sum_i E_i = \mathbb{1}$. In contrast to the case of finite-dimensional systems, the set of outcomes of a measurement performed over a CV state ρ is often continuous ($i \in \mathbb{R}$), so that $p(i) = \operatorname{tr}(E_i\rho)$ is a probability density function. A measurement is said to be Gaussian if, when applied to a Gaussian state, it yields outcomes that are Gaussian distributed. A property of such measurements is the following: given a $(N{+}M)$-mode Gaussian state, a Gaussian measurement of N modes gives a Gaussian probability density function for the outcomes, and the remaining M modes are left in a Gaussian state. From a practical point of view, any Gaussian measurement can be accomplished by homodyne detection, linear optics and Gaussian ancillary modes.

C.3.1 Homodyne detection

The most common Gaussian measurement used in CV quantum information is homodyne detection. It consists in measuring one of the quadratures of a mode. Mathematically, this is done by projecting over the quadrature basis, i.e., if $\hat{q}$ ($\hat{p}$) is the quadrature to be measured, the POVM elements are $E_q = |q\rangle\langle q|$ ($E_p = |p\rangle\langle p|$), that is they are projectors onto infinitely squeezed states. Experimentally, the homodyne detection is implemented by combining the target quantum mode with a local oscillator (LO) in a balanced beam splitter and measuring the intensity of the two output modes with two photodetectors. The subtraction of the signal of both photodetectors gives a signal proportional to $\hat{q}$ ($\hat{p}$).

The LO provides the phase reference ϕ for the quadrature measurement, thus by shifting the phase to $\phi \to \phi + \pi/2$ the other quadrature can be measured. For an arbitrary phase ϕ, the POVM elements associated to the homodyne detection are

$$E_{x_\phi} = |x_\phi\rangle\langle x_\phi| \ , \tag{C.43}$$

where $\hat{x}_\phi = \hat{q}\cos\phi + \hat{p}\sin\phi$.

C.3.2 Heterodyne detection

The heterodyne detection consists in, roughly speaking, measuring simultaneously *both* quadratures. The target mode is mixed with the vacuum by means of a balanced beam splitter, then homodyne detection of the conjugate quadratures is performed over the outgoing signals. Note that, in this case, quantum mechanics does not raise any objections to the simultaneous measurement of conjugate quadratures. This can be understood by taking into account that the fluctuations of the vacuum field introduce extra noise in the signal, and, as a consequence, the precision in the measurement of each quadrature is diminished so that the Heisenberg's uncertainty principle is preserved.

The heterodyne measurement can be viewed as a POVM which elements are projectors onto coherent states, i.e., $E_\alpha = (1/\pi)\,|\alpha\rangle\langle\alpha|$ [Leonhardt, 1997]. This procedure can be generalized to any POVM composed of projectors over pure Gaussian states [Giedke and Cirac, 2002]. This means that the most general pure Gaussian measurement that yields information about both quadratures of a state, which may be called a *generalized heterodyne measurement*, is achieved by a POVM with elements

$$E_{\alpha,r,\phi} = \frac{1}{\pi} \left|\alpha, r, \phi\right\rangle\!\left\langle\alpha, r, \phi\right| \, , \tag{C.44}$$

Moreover, such POVMs can be decomposed into a Gaussian unitary operation applied to the target mode and the ancillary modes (vacuum), the action of linear optical elements (beam splitters) and homodyne measurements on all output modes.

C.3.3 Photon counting and photodetection

Despite being non-Gaussian measurements, photon counting and photodetection play an important role in certain quantum information tasks, such as discrimination of Gaussian states and entanglement distillation. The photon counting measurement consists in projecting onto the number-state basis, i.e.,

$$E_n = |n\rangle\!\langle n| \, . \tag{C.45}$$

The measurement device is simply an optical receiver that converts light into electric current. When a single mode is excited, the receiver measures the intensity of the generated current, which is proportional to the photon number.

The photodetection measurement is a variant that serves to discriminate between two possible states: the vacuum, and one or more photons. The associated POVM elements are thus $E_0 = |0\rangle\!\langle 0|$ and $E_1 = \mathbb{1} - |0\rangle\!\langle 0|$. In practice, photodetectors typically have a small efficiency, i.e., only a small fraction of photons is detected. Real photodetectors can be modelled by adding a beam splitter before an ideal photodetector, which transmissivity relates to the efficiency of the detector.

APPENDIX D

Technical details of Chapter 6

D.1 Gaussian integrals

At many points in Chapter 6, we integrate complex-valued functions over the complex plane, weighted by the bidimensional Gaussian probability distribution $G(u)$. This Section gathers the integrals that we need. Recall that $G(u)$ is defined as

$$G(u) = \frac{1}{\pi\mu^2} e^{-u^2/\mu^2} \,, \quad u \in \mathbb{C}\,. \tag{D.1}$$

Expressing u either in polar or Cartesian coordinates in the complex plane, i.e., $u = re^{i\theta} = u_1 + iu_2$, one can readily check that $G(u)$ is normalized:

$$\int G(u) d^2u \;=\; \int_0^\infty \int_0^{2\pi} \frac{1}{\pi\mu^2} e^{-r^2/\mu^2} r dr d\theta = 1\,, \tag{D.2}$$

$$\int G(u) d^2u \;=\; \int_{-\infty}^\infty \int_{-\infty}^\infty \frac{1}{\pi\mu^2} e^{(-u_1^2-u_2^2)/\mu^2} du_1 du_2 = 1\,. \tag{D.3}$$

The average of a coherent state $|u\rangle\langle u|$ over the probability distribution $G(u)$ can be computed by expressing $|u\rangle$ in terms of Fock states, as in Eq. (C.28). It gives

$$\int G(u)\, |u\rangle\langle u|\, d^2u = \sum_{k=0}^{\infty} c_k\, |k\rangle\langle k| \,, \quad c_k = \frac{\mu^{2k}}{(\mu^2+1)^{k+1}}\,, \tag{D.4}$$

where $\{|k\rangle\}$ is the Fock basis. Note that the result of averaging a coherent state over $G(u)$ is nothing more than a thermal state with average photon number μ^2.

Variations of Eq. (D.4) with different complex functions that we use are

$$\begin{aligned}
\int G(u)u\,|u\rangle\langle u|\,d^2u &= \sum_{k=0}^{\infty} c_{k+1}\sqrt{k+1}\,|k\rangle\langle k+1|\;, && \text{(D.5)}\\
\int G(u)u^*\,|u\rangle\langle u|\,d^2u &= \sum_{k=0}^{\infty} c_{k}\sqrt{k}\,|k\rangle\langle k-1|\;, && \text{(D.6)}\\
\int G(u)|u|^2\,|u\rangle\langle u|\,d^2u &= \sum_{k=0}^{\infty} c_{k+1}(k+1)\,|k\rangle\langle k|\;, && \text{(D.7)}\\
\int G(u)u^2\,|u\rangle\langle u|\,d^2u &= \sum_{k=0}^{\infty} c_{k+2}\sqrt{k+2}\sqrt{k+1}\,|k\rangle\langle k+2|\;, && \text{(D.8)}\\
\int G(u)\,(u^*)^2\,|u\rangle\langle u|\,d^2u &= \sum_{k=0}^{\infty} c_{k}\sqrt{k}\sqrt{k-1}\,|k\rangle\langle k-2|\;, && \text{(D.9)}
\end{aligned}$$

and

$$\begin{aligned}
\int G(u)(u+u^*)d^2u &= 0 && \text{(D.10)}\\
\int G(u)(u+u^*)^2d^2u &= 2\mu^2 && \text{(D.11)}\\
\int G(u)|u|^2d^2u &= \mu^2 && \text{(D.12)}
\end{aligned}$$

D.2 Conditional probability $p(\beta|\alpha)$, Eq. (6.43)

Given two arbitrary Gaussian states ρ_A, ρ_B, the trace of their product is

$$\mathrm{tr}\,(\rho_A\rho_B) = \frac{2}{\sqrt{\det(V_A+V_B)}} e^{-\delta^T(V_A+V_B)^{-1}\delta}\,, \tag{D.13}$$

where V_A and V_B are their covariance matrices and δ is the difference of their displacement vectors. For the states $\rho_A \equiv |\sqrt{n}\alpha\rangle\langle\sqrt{n}\alpha|$ and $\rho_B \equiv E_{\bar\beta}$, we have

$$\begin{aligned}
V_A &= \begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}\;,\quad V_B = R\begin{pmatrix}e^{-2r} & 0\\ 0 & e^{2r}\end{pmatrix}R^T\;, && \text{(D.14)}\\
R &= \begin{pmatrix}\cos\phi & -\sin\phi\\ \sin\phi & \cos\phi\end{pmatrix}\;, && \text{(D.15)}\\
\delta &= (\sqrt{n}a_1-\bar b_1, \sqrt{n}a_2-\bar b_2)\,, && \text{(D.16)}
\end{aligned}$$

where $\alpha = a_1+ia_2$, $\bar\beta = \bar b_1 + i\bar b_2$, r is the squeezing parameter, and ϕ indicates the direction of squeezing in the phase space. In terms of α and $\bar\beta$, Eq. (D.13) reads

$$\mathrm{tr}\,(\rho_A\rho_B) = \frac{1}{\pi\cosh r} e^{-|\sqrt{n}\alpha-\bar{\beta}|^2-\mathrm{Re}[(\sqrt{n}\alpha-\bar{\beta})^2 e^{-i2\phi}]\tanh r}\,. \tag{D.17}$$

D.3 Integrals in Section 6.3.1

With the probability distribution $p(v)$, defined in Eq. (6.48), we make use of the integrals

$$\int p(v) I_1 d^2v = \int p(v) I_1^* d^2v = 0\,, \tag{D.18}$$

$$\int p(v) I_3 d^2v = \int p(v) I_3^* d^2v = 0\,, \tag{D.19}$$

$$\int p(v) I_2 d^2v = \mu^2\,, \tag{D.20}$$

$$\begin{aligned}\int p(v) I_1^2 d^2v &= \int p(v) (I_1^*)^2 d^2v \\ &= \frac{\mu^4\sinh(2r)}{(2\mu^2+1)\cosh(2r)+2\mu^2(\mu^2+1)+1}\,,\end{aligned} \tag{D.21}$$

$$\int p(v) |I_1|^2 d^2v = \frac{\mu^4(\cosh(2r)+2\mu^2+1)}{(2\mu^2+1)\cosh(2r)+2\mu^2(\mu^2+1)+1}\,. \tag{D.22}$$

Bibliography

M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions* (Dover Publications, New York, 1972).

a. Acín, E. Bagan, M. Baig, L. Masanes, and R. Muñoz Tapia, "Multiple-copy two-state discrimination with individual measurements", *Physical Review A* **71**, 032338 (2005).

E. Aïmeur, G. Brassard, and S. Gambs, "Machine Learning in a Quantum World", in I. L. Lamontagne and M. Marchand, editors, "Advances in Artificial Intelligence, Volume 4013 of Lecture Notes in Computer Science", pp. 431–442 (Springer, Berlin/Heidelberg, 2006).

D. Akimoto and M. Hayashi, "Discrimination of the change point in a quantum setting", *Physical Review A* **83**, 052328 (2011), 1102.2555.

K. M. R. Audenaert, L. Masanes, A. Acin, and F. Verstraete, "Discriminating States: The Quantum Chernoff Bound", *Physical Review Letters* **98**, 160501 (2007).

K. M. R. Audenaert, M. Mosonyi, and F. Verstraete, "Quantum state discrimination bounds for finite sample size", *Journal of Mathematical Physics* **53**, 122205 (2012), 1204.0711.

D. Avis and K. Fukuda, "A pivoting algorithm for convex hulls and vertex enumeration of arrangements and polyhedra", *Discrete & Computational Geometry* **8**, 295 (1992).

E. Bagan, M. Baig, and R. Muñoz Tapia, "Aligning Reference Frames with Quantum States", *Physical Review Letters* **87**, 257903 (2001).

E. Bagan, M. Ballester, R. Gill, A. Monras, and R. Muñoz Tapia, "Optimal full estimation of qubit mixed states", *Physical Review A* **73**, 19 (2006), 0510158.

E. Bagan, R. Muñoz Tapia, G. A. Olivares-Rentería, and J. A. Bergou, "Optimal discrimination of quantum states with a fixed rate of inconclusive outcomes", *Physical Review A* **86**, 040303 (2012).

K. Banaszek, "Optimal receiver for quantum cryptography with two coherent states", *Physics Letters A* **253**, 12 (1999).

S. M. Barnett, A. Chefles, and I. Jex, "Comparison of two unknown pure quantum states", *Physics Letters A* **307**, 189 (2003).

L. Bartůšková, A. Černoch, J. Soubusta, and M. Dušek, "Programmable discriminator of coherent states: Experimental realization", *Physical Review A* **77**, 034306 (2008).

A. Barvinok, "A bound for the number of vertices of a polytope with applications", (2012), 1108.2871.

I. Bengtsson and K. Zyczkowski, *Geometry of Quantum States* (Cambridge University Press, Cambridge, UK, 2006).

J. A. Bergou, V. Bužek, E. Feldman, U. Herzog, and M. Hillery, "Programmable quantum-state discriminators with simple programs", *Physical Review A* **73**, 062334 (2006a).

J. A. Bergou, E. Feldman, and M. Hillery, "Optimal unambiguous discrimination of two subspaces as a case in mixed-state discrimination", *Physical Review A* **73**, 032107 (2006b).

J. A. Bergou, U. Herzog, and M. Hillery, "Discrimination of Quantum States", *Lecture Notes in Physics* **649**, 417 (2004).

J. A. Bergou and M. Hillery, "Universal Programmable Quantum State Discriminator that is Optimal for Unambiguously Distinguishing between Unknown States", *Physical Review Letters* **94**, 160501 (2005).

J. M. Bernardo and A. F. Smith, *Bayesian Theory* (Wiley, Chichester, 1994).

C. M. Bishop, *Pattern Recognition and Machine Learning* (Springer, Berlin, 2006).

A. Bisio, G. Chiribella, G. M. D'Ariano, S. Facchini, and P. Perinotti, "Optimal quantum learning of a unitary transformation", *Physical Review A* **81**, 032324 (2010).

K. Blum, *Density Matrix Theory and Applications* (Plenum Press, New York, 1996), 2nd edition.

S. Boyd and L. Vandenberghe, *Convex Optimization* (Cambridge University Press, 2004).

S. L. Braunstein and P. van Loock, "Quantum information with continuous variables", *Reviews of Modern Physics* **77**, 513 (2005).

V. Bužek, M. Hillery, M. Ziman, and M. Roško, "Programmable Quantum Processors", *Quantum Information Processing* **5**, 313 (2006).

J. Calsamiglia, J. I. de Vicente, R. Muñoz Tapia, and E. Bagan, "Local Discrimination of Mixed States", *Physical Review Letters* **105**, 080504 (2010), 1004.5522.

J. Calsamiglia, R. Muñoz Tapia, A. Acin, and E. Bagan, "Quantum Chernoff bound as a measure of distinguishability between density matrices: Application to qubit and Gaussian states", *Physical Review A* **77**, 032311 (2008), 0708.2343.

C. Caves, C. Fuchs, and R. Schack, "Quantum probabilities as Bayesian probabilities", *Physical Review A* **65**, 022305 (2002).

N. J. Cerf, *Quantum Information with Continuous Variables of Atoms and Light* (Imperial College Press, London, 2007).

A. Chefles, "Quantum state discrimination", *Contemporary Physics* **41**, 401 (2000), 0010114.

A. Chefles, "Unambiguous discrimination between linearly dependent states with multiple copies", *Physical Review A* **64**, 062305 (2001).

A. Chefles and S. M. Barnett, "Optimum unambiguous discrimination between linearly independent symmetric states", *Physics Letters A* **250**, 223 (1998a).

A. Chefles and S. M. Barnett, "Strategies for discriminating between non-orthogonal quantum states", *Journal of Modern Optics* **45**, 1295 (1998b).

H. Chernoff, “A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations”, *Annals of Mathematical Statistics* **23**, 493 (1952).

G. Chiribella, G. M. D’Ariano, and D. Schlingemann, “How continuous quantum measurements in finite dimension are actually discrete”, *Physical Review Letters* **98**, 4 (2007).

G. Chiribella, G. M. D’Ariano, and D. Schlingemann, “Barycentric decomposition of quantum measurements in finite dimensions”, *Journal of Mathematical Physics* **51**, 022111 (2010).

J. Cirac, A. Ekert, and C. Macchiavello, “Optimal Purification of Single Qubits”, *Physical Review Letters* **82**, 4344 (1999).

A. J. T. Colin, “Programmed discrimination of multiple sets of qbits with added classical information”, *The European Physical Journal D* **66**, 185 (2012).

T. M. Cover and J. A. Thomas, *Elements of Information Theory* (John Wiley & Sons, New York, 2006), 2nd edition.

G. M. D’Ariano, P. L. Presti, and P. Perinotti, “Classical randomness in quantum measurements”, *Journal of Physics A: Mathematical and General* **38**, 5979 (2005).

B. de Finetti, “Sul significato soggettivo della probabilità”, *Fundamenta Mathematicae* **17**, 298 (1931).

R. Derka, V. Bužek, and A. Ekert, “Universal Algorithm for Optimal Estimation of Quantum States from Finite Ensembles via Realizable Generalized Measurement”, *Physical Review Letters* **80**, 1571 (1998).

D. Deutsch, “Quantum Theory of Probability and Decisions”, *Proc. R. Soc. Lond. A* **455**, 3129 (1999), 9906015.

D. Dieks, “Overlap and distinguishability of quantum states”, *Physics Letters A* **126**, 303 (1988).

S. J. Dolinar, “Processing and Transmission of Information”, *Research Laboratory of Electronics, Massachusetts Institute of Technology (MIT), Quarterly Process Report* **111**, 115 (1973).

M. Dušek and V. Bužek, “Quantum-controlled measurement device for quantum-state discrimination”, *Physical Review A* **66**, 022112 (2002).

A. R. Edmonds, *Angular Momentum in Quantum Mechanics* (Princeton University Press, Princeton, New Jersey, 1960).

J. Eisert and M. B. Plenio, “Introduction to the basics of entanglement theory in continuous-variable systems”, *International Journal of Quantum Information* **1**, 14 (2003), 0312071.

Y. Eldar, “Mixed-quantum-state detection with inconclusive results”, *Physical Review A* **67**, 042309 (2003).

W. Feller, *An Introduction to Probability Theory and Its Applications*, volume 14 (John Wiley & Sons, Inc., New York, 1950).

J. Fiurášek and M. Dušek, “Probabilistic quantum multimeters”, *Physical Review A* **69**, 032302 (2004).

J. Fiurášek, M. Dušek, and R. Filip, “Universal Measurement Apparatus Controlled by Quantum Software”, *Physical Review Letters* **89**, 190401 (2002).

J. Fiurášek and M. Ježek, “Optimal discrimination of mixed quantum states involving inconclusive results”, *Physical Review A* **67**, 012321 (2003), 0208126.

D. Fortin and I. Tseveendorj, “Piece adding technique for convex maximization problems”, *Journal of Global Optimization* **48**, 583 (2010).

C. A. Fuchs, *Distinguishability and Accessible Information in Quantum Theory*, Ph.D. thesis (1996), 9601020.

C. A. Fuchs, “QBism, the Perimeter of Quantum Bayesianism”, p. 30 (2010), 1003.5209.

C. A. Fuchs and R. Schack, “Unkwown quantum states and operations, a Bayesian view”, *Lecture Notes in Physics* **649**, 147 (2004).

A. Furusawa, J. L. Sø rensen, S. L. Braunstein, C. A. Fuchs, H. J. Kimble, and E. S. Polzik, “Unconditional Quantum Teleportation”, *Science* **282**, 706 (1998).

B. Gendra, E. Ronco-Bonvehi, J. Calsamiglia, R. Muñoz Tapia, and E. Bagan, “Beating noise with abstention in state estimation”, *New Journal of Physics* **14**, 105015 (2012), 1205.5479.

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

G. Giedke and J. I. Cirac, “Characterization of Gaussian operations and distillation of Gaussian states”, *Physical Review A* **66**, 032316 (2002), 0204085.

R. D. Gill and M. Guţă, “On Asymptotic Quantum Statistical Inference”, in M. Banerjee, F. Bunea, J. Huang, V. Koltchinskii, and M. H. Maathuis, editors, “From Probability to Statistics and Back: High-Dimensional Models and Processes – A Festschrift in Honor of Jon A. Wellner”, pp. 105–127 (Institute of Mathematical Statistics, Beachwood, Ohio, 2013), 1112.2078.

R. D. Gill and B. Y. Levit, “Applications of the Van Trees inequality: A Bayesian Cramér-Rao bound”, *Bernoulli* **1**, 59 (1995).

D. A. Gillies, *Philosophical Theories of Probability* (Routledge, London, 2000).

F. Grosshans, G. Van Assche, J. Wenger, R. Brouri, N. J. Cerf, and P. Grangier, “Quantum key distribution using gaussian-modulated coherent states.”, *Nature* **421**, 238 (2003).

M. Guţă and W. Kotłowski, “Quantum learning: asymptotically optimal classification of qubit states”, *New Journal of Physics* **12**, 123032 (2010).

E. Haapasalo, T. Heinosaari, and J.-P. Pellonpää, “Quantum measurements on finite dimensional systems: relabeling and mixing”, *Quantum Information Processing* **11**, 1751 (2011).

M. J. Hall, “Random quantum correlations and density operator distributions”, *Physics Letters A* **242**, 123 (1998).

a. Hayashi, T. Hashimoto, and M. Horibe, “State discrimination with error margin and its locality”, *Physical Review A* **78**, 1 (2008).

a. Hayashi, M. Horibe, and T. Hashimoto, “Quantum pure-state identification”, *Physical Review A* **72**, 1 (2005).

a. Hayashi, M. Horibe, and T. Hashimoto, “Unambiguous pure-state identification without classical knowledge”, *Physical Review A* **73**, 1 (2006).

B. He and J. A. Bergou, “Programmable unknown quantum-state discriminators with multiple copies of program and data: A Jordan-basis approach”, *Physical Review A* **75**, 1 (2007).

T. Heinosaari and J.-P. Pellonpää, “Generalized coherent states and extremal positive operator valued measures”, *Journal of Physics A: Mathematical and Theoretical* **45**, 244019 (2012).

C. W. Helstrom, *Quantum Detection and Estimation Theory* (Academic Press, New York, 1976).

A. Hentschel and B. C. Sanders, “Machine Learning for Precise Quantum Measurement”, *Physical Review Letters* **104**, 063603 (2010).

U. Herzog and J. A. Bergou, “Optimum unambiguous discrimination of two mixed quantum states”, *Physical Review A* **71**, 050301 (2005).

U. Herzog and J. A. Bergou, “Erratum: Optimum unambiguous identification of d unknown pure qudit states [Phys. Rev. A 78, 032320 (2008)]”, *Physical Review A* **78**, 069902 (2008a).

U. Herzog and J. A. Bergou, “Optimum unambiguous identification of d unknown pure qudit states”, *Physical Review A* **78**, 032320 (2008b).

B. Higgins, A. C. Doherty, S. Bartlett, G. Pryde, and H. Wiseman, “Multiple-copy state discrimination: Thinking globally, acting locally”, *Physical Review A* **83**, 11 (2011), 1012.3525.

P. G. Hoel, S. C. Port, and C. J. Stone, “Testing Hypotheses”, in Houghton-Mifflin, editor, “Introduction to Statistical Theory”, chapter 3 (University of Minnesota, 1971).

A. Holevo, *Probabilistic and Statistical Aspects of Quantum Theory* (North-Holland, Amsterdam, 1982).

A. S. Holevo, “Bounds for the Quantity of Information Transmitted by a Quantum Communication Channel [Probl. Inf. Transm. 9 (1973) 110]”, *Probl. Peredachi Inf.* **9**, 3 (1973).

C. Howson and P. Urbach, *Scientific Reasoning: The Bayesian Approach.*, volume 36 (Open Court, La Salle, Illinois, 2006), 3rd edition.

Y. Ishida, T. Hashimoto, M. Horibe, and A. Hayashi, “Locality and nonlocality in quantum pure-state identification problems”, *Physical Review A* **78**, 1 (2008).

I. Ivanovic, “How to differentiate between non-orthogonal states”, *Physics Letters A* **123**, 257 (1987).

G. Jaeger and A. Shimony, “Optimal distinction between two non-orthogonal quantum states”, *Physics Letters A* **197**, 83 (1995).

E. T. Jaynes, *Probability Theory: The Logic of Science* (Cambridge University Press, 2003).

R. Jeffrey, *Subjective Probability: The Real Thing* (Cambridge University Press, 2004).

I. Jex, E. Andersson, and A. Chefles, “Comparing the states of many quantum systems”, *Journal of Modern Optics* **51**, 505 (2004), `0305120`.

B. Julsgaard, A. Kozhekin, and E. S. Polzik, “Experimental long-lived entanglement of two macroscopic objects.”, *Nature* **413**, 400 (2001).

R. S. Kennedy, E. V. Hoversten, P. Elias, and V. Chan, “Processing and Transmission of Information”, *Research Laboratory of Electronics, Massachusetts Institute of Technology (MIT), Quarterly Process Report* **108**, 219 (1973).

J. M. Keynes, *A Treatise on Probability* (Macmillan and Company, 1921).

L. Khachiyan, E. Boros, K. Borys, K. Elbassioni, and V. Gurvich, “Generating All Vertices of a Polyhedron Is Hard”, *Discrete & Computational Geometry* **39**, 174 (2008).

U. Leonhardt, *Measuring the Quantum State of Light* (Cambridge University Press, New York, 1997).

D. Lewis, *A Subjectivist’s Guide to Objective Chance (in Studies in Inductive Logic and Probability)* (University of California Press, Berkeley and Los Angeles, 1980).

X. Li, Q. Pan, J. Jing, J. Zhang, C. Xie, and K. Peng, “Quantum Dense Coding Exploiting a Bright Einstein-Podolsky-Rosen Beam”, *Physical Review Letters* **88**, 047904 (2002).

S. Lloyd, “Enhanced sensitivity of photodetection via quantum illumination”, *Science (New York, N.Y.)* **321**, 1463 (2008).

S. Lloyd, M. Mohseni, and P. Rebentrost, “Quantum algorithms for supervised and unsupervised machine learning”, p. 11 (2013), `1307.0411`.

D. J. MacKay, *Information Theory, Inference, and Learning Algorithms* (Cambridge University Press, 2003).

R. Nair, “Discriminating quantum-optical beam-splitter channels with number-diagonal signal states: Applications to quantum reading and target detection”, *Physical Review A* **84**, 032312 (2011).

H. Neven, V. S. Denchev, G. Rose, and W. G. Macready, “Training a Large Scale Classifier with the Quantum Adiabatic Algorithm”, p. 14 (2009), `0912.0779`.

M. Nielsen and I. Chuang, “Programmable Quantum Gate Arrays”, *Physical Review Letters* **79**, 321 (1997).

M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, Cambridge, England, 2000).

J. Niset, A. Acín, U. Andersen, N. Cerf, R. García-Patrón, M. Navascués, and M. Sabuncu, “Superiority of Entangled Measurements over All Local Strategies for the Estimation of Product Coherent States”, *Physical Review Letters* **98**, 260404 (2007).

M. Nussbaum and A. Szkoła, “The Chernoff lower bound for symmetric quantum hypothesis testing”, *The Annals of Statistics* **37**, 1040 (2009).

J.-P. Pellonpää, “Complete characterization of extreme quantum observables in infinite dimensions”, *Journal of Physics A: Mathematical and Theoretical* **44**, 085304 (2011).

A. Peres, “How to differentiate between non-orthogonal states”, *Physics Letters A* **128**, 19 (1988).

A. Peres, “Neumark’s theorem and quantum inseparability”, *Foundations of Physics* **20**, 1441 (1990).

A. Peres and W. Wootters, “Optimal detection of quantum information”, *Physical Review Letters* **66**, 1119 (1991).

D. Petz, “Monotone metrics on matrix spaces”, *Linear Algebra and its Applications* **244**, 81 (1996).

D. Petz and C. Sudár, “Geometries of quantum states”, *Journal of Mathematical Physics* **37**, 2662 (1996).

S. Pirandola, "Quantum Reading of a Classical Digital Memory", *Physical Review Letters* **106**, 090504 (2011).

K. R. Popper, *Quantum Theory and the Schism in Physics* (Rowman & Littlefield, New Jersey, 1982).

K. L. Pudenz and D. A. Lidar, "Quantum adiabatic machine learning", *Quantum Information Processing* **12**, 2027 (2013), 1109.0325.

M. F. Pusey, J. Barrett, and T. Rudolph, "On the reality of the quantum state", *Nature Physics* **8**, 476 (2012), 1111.3328.

F. P. Ramsey, *The Foundations of Mathematics and Other Logical Essays* (Routledge & Kegan Paul, London, 1931), r. b. brai edition.

P. Rapčan, J. Calsamiglia, R. Muñoz Tapia, E. Bagan, and V. Bužek, "Scavenging quantum information: Multiple observations of quantum systems", *Physical Review A* **84**, 032326 (2011), 1105.5326.

P. Raynal, *Unambiguous State Discrimination of two density matrices in Quantum Information Theory*, Ph.D. thesis (2006), 0611133.

T. Rudolph, R. Spekkens, and P. Turner, "Unambiguous discrimination of mixed states", *Physical Review A* **68**, 010301 (2003), 0303071.

R. Ruggles and H. Brodie, "An Empirical Approach to Economic Intelligence in World War II", *Journal of the American Statistical Association* **42**, 72 (1947).

M. Sasaki and A. Carlini, "Quantum learning and universal quantum matching machine", *Physical Review A* **66**, 022303 (2002).

M. Sedlák, M. Ziman, V. Bužek, and M. Hillery, "Unambiguous comparison of ensembles of quantum states", *Physical Review A* **77**, 042304 (2008).

M. Sedlák, M. Ziman, V. Bužek, and M. Hillery, "Unambiguous identification of coherent states. II. Multiple resources", *Physical Review A* **79**, 062305 (2009).

M. Sedlák, M. Ziman, O. Přibyla, V. Bužek, and M. Hillery, "Unambiguous identification of coherent states: Searching a quantum database", *Physical Review A* **76**, 022326 (2007).

G. Sentís, E. Bagan, J. Calsamiglia, and R. Muñoz Tapia, "Programmable discrimination with an error margin", *Physical Review A* **88**, 052304 (2013).

R. A. Servedio and S. J. Gortler, "Equivalences and Separations Between Quantum and Classical Learnability", *SIAM Journal on Computing* **33**, 1067 (2004).

G. Spedalieri, C. Lupo, S. Mancini, S. L. Braunstein, and S. Pirandola, "Quantum reading under a local energy constraint", *Physical Review A* **86**, 012315 (2012).

H. Sugimoto, T. Hashimoto, M. Horibe, and A. Hayashi, "Discrimination with error margin between two states: Case of general occurrence probabilities", *Physical Review A* **80**, 1 (2009).

Y. Sun, J. A. Bergou, and M. Hillery, "Optimum unambiguous discrimination between subsets of nonorthogonal quantum states", *Physical Review A* **66**, 032315 (2002).

M. Takeoka and M. Sasaki, "Discrimination of the binary coherent signal: Gaussian-operation limit and simple non-Gaussian near-optimal receivers", *Physical Review A* **78**, 022320 (2008).

S.-H. Tan, B. Erkmen, V. Giovannetti, S. Guha, S. Lloyd, L. Maccone, S. Pirandola, and J. Shapiro, "Quantum Illumination with Gaussian States", *Physical Review Letters* **101**, 253601 (2008).

J. P. Tej, a. R. U. Devi, and a. K. Rajagopal, "Quantum reading of digital memory with non-Gaussian entangled light", *Physical Review A* **87**, 052308 (2013).

M. J. Todd, "The many facets of linear programming", *Mathematical Programming* **91**, 417 (2002).

M. Touzel, R. Adamson, and A. Steinberg, "Optimal bounded-error strategies for projective measurements in nonorthogonal-state discrimination", *Physical Review A* **76**, 062314 (2007).

L. Vandenberghe and S. Boyd, "Semidefinite Programming", *SIAM Review* **38**, 49 (1996).

G. Vidal, J. Latorre, P. Pascual, and R. Tarrach, "Optimal minimal measurements of mixed states", *Physical Review A* **60**, 126 (1999).

D. Wallace, “Quantum Probability from Subjective Likelihood: improving on Deutsch’s proof of the probability rule”, *Stud. Hist. Phil. Mod. Phys.2* **38**, 311 (2007), 0312157v2.

C. Weedbrook, S. Pirandola, R. García-Patrón, N. J. Cerf, T. C. Ralph, J. H. Shapiro, and S. Lloyd, “Gaussian quantum information”, *Reviews of Modern Physics* **84**, 621 (2012).

J. A. Wheeler, “Information, Physics, Quantum: The Search for Links”, in S. Kobayashi, H. Ezawa, Y. Murayama, and S. Nomura, editors, “Proceedings of the 3rd International Symposium on Foundations of Quantum Mechanics in the Light of New Technology”, pp. 354–368 (Physical Society of Japan, Tokyo, 1990).

E. Wigner, “On the Quantum Correction For Thermodynamic Equilibrium”, *Physical Review* **40**, 749 (1932).

C. Zhang, M. Ying, and B. Qiao, “Universal programmable devices for unambiguous discrimination”, *Physical Review A* **74**, 042308 (2006).

C.-w. Zhang, C.-f. Li, and G.-c. Guo, “General strategies for discrimination of quantum states”, *Physics Letters A* **261**, 25 (1999), 9908001.

T. Zhou, “Unambiguous discrimination between two unknown qudit states”, *Quantum Information Processing* **11**, 1669 (2011), 1105.3004.

T. Zhou, “Success probabilities for universal unambiguous discriminators between unknown pure states”, *Physical Review A* **89**, 014301 (2014), 1308.0707.

K. Życzkowski and H.-J. Sommers, “Average fidelity between random quantum states”, *Physical Review A* **71**, 032313 (2005).