UNIVERSITÀ DEGLI STUDI DI PAVIA

Facoltà di Scienze MM. FF. NN.
Dipartimento di Fisica “A. Volta”

Optimal estimation of quantum signals in the presence of symmetry

PhD Thesis
by
Giulio Chiribella

Supervisor:
Chiar.mo Prof. **Giacomo Mauro D’Ariano**

XIX Ciclo, A. A. 2003-2006

Contents

Introduction

Any physical system is by its very definition a carrier of information. When the nature of the system is quantum, however, the read-out and the processing of information acquire rather enigmatic features, such as the complementarity of different properties, the impossibility of perfectly determining the system's state and the impossibility of perfectly copying it. The presence of quantum limitations in the estimation of physical parameters and, consequently, in the transmission rate of communication channels, firstly suggested by Gabor and Brillouin in the fifties, came to the experimental domain in the sixties in the field of quantum optics, along with the possibility of sending and manipulating information via coherent laser radiation. These issues came together into the general framework of *Quantum Estimation Theory,* worked out in the seventies and beautifully presented in the books by Helstrom [1] and Holevo [2].

From the foundation of quantum estimation theory up to now, the research about the best measurements and the best strategies for estimating physical parameters has been a major focus. In particular, a common situation is the one in which the parameters of interest identify to the action of a symmetry group, which transforms the quantum state of the system. Such a situation arises in communication schemes where the signal states that are used to encode information are generated by applying a group transformation to a given input state, or, more generally, to a set of input states. A typical example is the encoding of information into the phase of a laser beam, where the signal states are generated from a coherent input state by applying elements of the group $U(1)$ of phase shifts. The interest in the problem of optimally estimating a group transformation received in the last fifteen years a strong motivation from the new field of Quantum Information [3], due to the broad spectrum of its applications, covering quantum communication and cryptography, the study of generalized uncertainty relations, and the design of high sensitivity measurements for quantum metrology.

Despite the long-dated attention to the problem, many new applications of quantum estimation involving symmetry groups still continue to come up, for example in the line of research about reference frames as communication resources [4, 5, 6, 7, 8] (for a very recent review on this topic, see Ref. [9]), or in the study of a quantum-enhanced methods for global positioning and clock synchronization [10, 11]. As a counterpart, some controversial issues have been clarified only in the very last years. In the variety of this scenario, it is somehow natural to look for a general point of view, suitable for capturing the main features of the problem without loosing the understanding in a labirinth of special cases. The aim of this presentation is to give a systematic tractation of the strategies for optimal estimation of group parameters. A clear understanding of the group theoretical structure involved in the problem provides a deep insight into the mechanisms regulating the performances of quantum estimation strategies, thus solving in a unified fashion a large number of problems appearing in concrete applications. In this spirit, the interplay between the mathematical tools and physical features will receive a constant attention. Accordingly, the method of presentation will be to first establish general properties of optimal estimation, and then to specialize them to concrete examples.

In the following, the contents of the various chapters are outlined:

Chapter 1 gives a brief introduction to quantum estimation. Following essentially the exposition of Holevo's book [2], we will introduce the statistical model of quantum mechanics, specified in terms of quantum states (density operators) and generalized observables (POVMs). The concept of POVM will provide the most general description of a strategy for the estimation of a physical parameter. Finally, three possible definitions of optimality are presented, i.e. the minimization of a cost function in the Bayesian and the minimax approaches, and the maximization of the mutual information.

Chapter 2 summarizes a number of fundamental notions of representation theory that will be extensively used in the rest of this presentation. While the material included in the Chapter is mainly taken from the standard references [12, 13, 14], its exposition is the result of an effort of symplifying the notation and the proofs in order to make the presented notions suitable for an immediate use in quantum information problems.

In **Chapter 3** the general problem of estimating a physical parameter is specialized to the covariant case, i.e. the case where the space of estimated parameters is invariant under the action of a symmetry group. In the covariant problem, the optimal estimation strategy can be restricted without loss of generality to the class of *covariant POVMs.* The general form of a covariant POVM, and the relation between covariance and optimality are presented.

Chapter 4, based on the papers [15, 16], opens the original part of this presentation. This Chapter is devoted to the derivation of the strategies that maximize the likelihood in the estimation of a group transformation, i.e. the probability (probability density for continuous groups) of estimating the correct value of the unknown group parameters. The maximization of the likelihood is a prototype of optimization problem, that admits a simple and general solution, holding for a large class of groups. The results of this Chapter emphasize to role of entanglement in tensor product structure induced by the group action in the Hilbert space, a feature that turns out to be crucial for further developments in the subsequent Chapters.

In **Chapter 5** we focus on the discrimination of a finite set of symmetry transformations, presenting a necessary and sufficient condition under which such a discrimination is perfect, i.e. error-free. It is shown that any unitary gate in a given group representation can be identified with zero error probability, provided that the gate is applied a suitable number of times to an entangled input state. The minimum number of iterations is quantified via a set of bounds that only depend on the algebraic properties the given group representation. This Chapter, which represents a kind of long example of application of the maximum likeihood approach, can be skipped by the reader which is is more intersted of the derivation of optimal estimation strategies with generic cost functions, which is given in Chapter 6. The material of Chapter 5 is original and unpublished [17]

Chapter 6 is the central part of this presentation, and refers to the papers [8, 18, 19]. In this Chapter we derive the optimal estimation strategy and the optimal input states for the estimation of a set of group transformations. Optimality is defined here as the minimization of a Bayes cost, for a large class of cost functions. The result, holding for any finite and compact Lie group, clarifies the role of entanglement in the optimal estimation. Two direct applications of the general method are the estimation of $\mathbb{SU}(2)$ transformations for the alignement of spatial reference frames, and the estimation of an unknown maximally entangled state with a finite ensemble of identical copies. The Chapter concludes with the optimal estimation of a squeezing parameter, an example where the validity of the main result is extended to a case of noncompact group.

Chapter 7 generalizes the results of Chapter 4 to the case of nonunimodular groups, i.e. groups whose left-invariant measure differs from the right-invariant one. This generalization makes possible to treat the joint estimation of a squeezing and a displacement parameter in the radiation field and to show that the product of the two uncertainties in the optimal measurement is asymptotically twice the Heisenberg limit, the same relation occurring in the joint measurement of position and momentum [20]. The presented materials are published in Ref.[21].

Chapter 8 contains with the characterization of the extremal POVMs in finite dimensional systems. Extremal POVMs represent estimation strategies that cannot by realized by mixing the statistics of different experiments, i.e. estimation strategies that are completely free from classical noise. We will prove that for finite dimensional systems an extremal POVM can have only a finite number of outcomes. In addiction, since in the presence of symmetry the optimization can be restricted to the subset of covariant POVMs, the characterization of extremal covariant POVMs provides a useful tool for optimization. An account about some properties of extremal POVMs in the solution of optimization problems is given at the end of the Chapter. The part of the Chapter regarding extremal covariant POVMs are based on the papers [22, 23], while the characterization of the extremal POVMs with continuous outcome space is an original unpublished result.

Chapter 9 concludes the thesis with a brief *excursus* in the field of quantum information processing, included to provide, after the derivation of the optimized strategies of the Chapters 4, 6, and 7, an additional *a posteriori* motivation for the interest in the estimation of quantum signals. We will show that the estimation of suitable physical parameters can be used to approximate efficiently any quantum channel that distributes quantum information to a large number of users in a permutationally invariant way. The estimation techniques illustrated in the preceeding Chapters appear then as a tool for the approximate realization of important quantum information processing tasks such as optimal cloning. The material of this concluding Chapter is taken from Ref. [24].

Notations

The beginning of each Chapter is provided with a short abstract that summarizes the aim and the main ideas that are going to be presented.

Some notations are extensively used as a standard throughout the whole presentation. For the reader's convenience, most of them are reported below:

$\mathcal{H}, \mathcal{K}$	Hilbert spaces
$\mathcal{L}(\mathcal{H})$	space of linear operators on $\mathcal{H}$
$\mathcal{B}(\mathcal{H})$	C^*-algebra of bounded operators on $\mathcal{H}$
$\mathcal{T}(\mathcal{H})$	Banach space of trace-class operators
$\mathcal{S}(\mathcal{H})$	convex set of quantum states on $\mathcal{H}$
$\mathbf{G}, \mathbf{H}$	groups
$\lvert\mathbf{G}\rvert$	cardinality of the finite group $\mathbf{G}$
$\mathrm{d}_L g, \mathrm{d}_R g$	left- and right-invariant Haar measure over a locally compact group
$\mathrm{d}g$	invariant Haar measure over a unimodular group
$\mathrm{Irr}(\mathbf{G}, \omega)$	collection of all IRREPs of the group $\mathbf{G}$ with cocycle ω
$\mathsf{R}(\mathbf{G}) = \{U_g \mid g \in \mathbf{G}\}$	projective representation of the group $\mathbf{G}$
S	collection of all the IRREPs contained in the Clebsch-Gordan decomposition of $\{U_g\}$
$\mathcal{A}_g : \rho \longmapsto U_g \rho U_g^\dagger$	automorphism of the set of states $\mathcal{S}(\mathcal{H})$
Ω, Θ	measure spaces
$\sigma(\Omega)$	$\sigma-$algebra of measurable subsets of Ω

• $P : \sigma(\Omega) \longrightarrow \mathcal{B}(\mathcal{H})$	POVM on the Hilbert space $\mathcal{H}$ with outcome space Ω
• $\doteq$	equal by definition
• $\cong$	isomorphic

Chapter 1

Introduction to quantum estimation theory

1.1 The statistical model of quantum mechanics

The notions of *states* and *observables* lie at the core of any probabilistic theory. States are the mathematical object describing the possible preparations of a given system, while observables describe the experimental procedures which produce an outcome—the measurement result—according to some probability distribution. The set of states and the set of observables are both convex, according to the possibility of randomizing both the preparations of the system and the different experimental procedures.

1.1.1 Quantum states

In standard quantum mechanics, any physical system is associated with a complex separable Hilbert space $\mathcal{H}$, of dimension $\dim \mathcal{H} \leq \infty$. The possible states of the system are described by density operators on $\mathcal{H}$, namely by trace-class operators $\rho \in \mathcal{T}(\mathcal{H})$ satisfying $\rho \geq 0$ and $\mathrm{Tr}[\rho] = 1$. Physically, the state ρ can be interpreted as the equivalence class of all the preparations of a system that give the same statistics for any possible experiment [25].

The set of quantum states for the Hilbert space $\mathcal{H}$ will be denoted as $\mathcal{S}(\mathcal{H})$. It is immediate to check that the set $\mathcal{S}(\mathcal{H})$ is convex, namely for any two density operators $\rho_1, \rho_2 \in \mathcal{S}(\mathcal{H})$ and for any $p \in [0, 1]$, the convex combination $\rho = p\rho_1 + (1 - p)\rho_2$ is a density operator. The convex combination can be interpreted as the randomized choice between two inequivalent preparation procedures.

The extreme points of the state space $\mathcal{S}(\mathcal{H})$ are called *pure states*, while the other states are called *mixed*. It is easy to see that the pure states are the rank-one projectors $\rho = |\phi\rangle\langle\phi|$. In fact, any state ρ can be diagonalized as[1]

$$\rho = \sum_{i=1}^{r} p_i |\phi_i\rangle\langle\phi_i| \ , \tag{1.1}$$

where $p_i \geq 0, \sum_i p_i = 1$, and $\{|\phi_i\rangle \mid i = 1, \ldots, r \leq \infty\}$ is a suitable orthonormal basis. The diagonalization yields one of the possible decompositions of a density operator as convex combination of extremal points. However, it is worth stressing that a mixed state admits many different convex decompositions, corresponding to many different realizations of it as a statistical mixture of pure states.

1.1.2 Quantum observables

Generally speaking, an experiment is a procedure performed on a physical system, producing as output an event—the measurement outcome—with some probability. The statistical description of the experiment is given by specifying the space of measurement outcomes, and by assigning a rule that, given the state of the system, predicts the probabilities of the possible events. The space of outcomes, here denoted by Ω, must be a measure space, i.e. a set equipped with a $\sigma-$algebra $\sigma(\Omega)$ of subsets[2]. The elements of the $\sigma-$algebra are the possible *events* in the experiment, namely the subset $B \subseteq \Omega$ corresponds to the event "the measurement outcome is an element of B".

The outcome space Ω can be either discrete or continuous. For example, in a Stern-Gerlach experiment with spin 1/2 particles the possible outcomes are "spin up" or "spin down", namely $\Omega = \{\text{up}, \text{down}\}$, while in the measurement of the position of a particle one has $\Omega = \mathbb{R}^3$. In this case, the Borel subset $B \subseteq \mathbb{R}^3$ corresponds to the event "the particle is observed in the region B".

The statistics of an experiment is given by an affine map from the state space $\mathcal{S}(\mathcal{H})$ to the convex set of probability distributions over Ω. Such a map associates any state $\rho \in \mathcal{S}(\mathcal{H})$ with a probability distribution $p(B|\rho)$, which

[1] It is known that a trace-class operator $A \in \mathcal{TH}$ has discrete spectrum, with the only exception, eventually, of accumulation points of the discrete spectrum (that belong to the continuous spectrum). As a consequence, A has a discrete spectral resolution.

[2] A σ-algebra $\sigma(\Omega)$ for the measure space Ω is a collection of subsets with the properties:

- $\emptyset \in \sigma(\Omega)$
- $\Omega - B \in \sigma(\Omega) \qquad \forall B \in \sigma(\Omega)$
- $(\bigcup_{i=1}^{n} B_i) \in \sigma(\Omega) \qquad \forall \{B_i \in \sigma(\Omega) \mid i = 1, \ldots, n \leq \infty\} \ , B_i \cap B_j = \emptyset.$

specifies the probability of obtaining the event $B \in \sigma(\Omega)$, provided that the system was prepared in the state ρ. Physically, the requirement of the map being affine means that if the preparation of the state is the randomization of two different procedures, i.e. $\rho = p\rho_1 + (1-p)\rho_2$, then the statistics of the experiment is the mixing of the corresponding probability distributions, i.e.

$$p(B|\rho) = p\ p(B|\rho_1) + (1-p)\ p(B|\rho_2)\ . \tag{1.2}$$

According to the proof given by Holevo [26] (see also [2]), any such map can be represented by a *positive operator valued measure (POVM).* The concept of POVM, extensively studied by Naĭmark, was firstly introduced in the theory of quantum measurement by Davies and Lewis [27] and Holevo [28].

Definition 1. *A POVM P with probability space Ω is a function $P : \sigma(\Omega) \to \mathcal{B}(\mathcal{H})$ that associates any event $B \in \sigma(\Omega)$ with an operator $P(B) \in \mathcal{B}(\mathcal{H})$, according to the requirements*

$$P(B) \geq 0 \qquad \forall B \tag{1.3}$$

$$P(\Omega) = 1 \tag{1.4}$$

$$P(\cup_{n=1}^{\infty} B_n) = \sum_{n=1}^{\infty} P(B_n) \qquad \forall \{B_n\} : B_m \cap B_n = \emptyset,\ m \neq n \tag{1.5}$$

where the series in the last equation converges strongly.

Once the POVM is given, the affine map from the state ρ to the probability distribution $p(B|\rho)$ is uniquely specified via the *Born rule*

$$p(B|\rho) = \mathrm{Tr}[P(B)\rho]\ . \tag{1.6}$$

The requirements of Eqs. (1.3), (1.4), (1.5) guarantee for any state $\rho \in \mathcal{S}(\mathcal{H})$ the positivity of probabilities, the normalization of the probability distribution, and the additivity of probabilities for disjoint events, respectively.

In the following we will refer to a POVM as to a (generalized) *quantum observable*, often omitting the term “generalized”. This terminology may look unusual, since commonly the term “observable” denotes a self-adjoint operator. Since a self-adjoint operator has a spectral resolution made of orthogonal projectors, one might prefer to call “observable” only the *projector valued measures (PVMs)* [29], namely the POVMs that enjoy the additional property

$$P(B_1)P(B_2) = P(B_1 \cup B_2)\ . \tag{1.7}$$

However, one has the Naĭmark extension theorem:

Theorem 1 (Naĭmark extension [30]). *Any POVM P for the Hilbert space $\mathcal{H}$ can be extended to a PVM E on a larger Hilbert space $\widetilde{\mathcal{H}} \supseteq \mathcal{H}$, namely*

$$P(B) = P_{\mathcal{H}} E(B) P_{\mathcal{H}} \qquad \forall B \in \sigma(\Omega) \ , \tag{1.8}$$

where $P_{\mathcal{H}}$ is the projector onto $\mathcal{H}$.

Here, the extended Hilbert space $\widetilde{\mathcal{H}}$ can be always considered as representing a compound system, i.e. $\widetilde{\mathcal{H}} = \mathcal{H} \otimes \mathcal{H}_A$, where $\mathcal{H}_A$ is the Hilbert space of an ancilla. Moreover, considering the isomorphism $\mathcal{H} \cong \mathcal{H} \otimes |0\rangle$, where $|0\rangle \in \mathcal{H}_A$ is a normalized vector, one can write $P_{\mathcal{H}} = \mathbb{1} \otimes |0\rangle\langle 0|$. Then, Naĭmark theorem implies

$$\mathrm{Tr}[\rho P(B)] = \mathrm{Tr}[\rho \otimes |0\rangle\langle 0| E(B)] \ , \tag{1.9}$$

for any state $\rho \in \mathcal{S}(\mathcal{H})$. This means that any POVM can be implemented by preparing an ancilla in a pure state $\rho_A = |0\rangle\langle 0|$, and by measuring a suitable PVM on the compound system. In conclusion, if any self-adjoint operator is observable, then any POVM is observable as well. A part from devotion to traditional terminology, there is no reason at all to call "observables" only the self-adjoint operators.

1.2 Quantum estimation strategies

When classical information is encoded into quantum systems, in general its read-out suffers the intrinsically quantum limitation of discriminating among nonorthogonal states (an error-free read-out is possible only if the information has been encoded into orthogonal states). It becomes then crucial to choose the best estimation strategy, namely the experimental procedure that optimizes the decoding of the signal with respect to a given optimality criterion. This is the central problem of *quantum estimation theory*[1, 2].

The paradigmatic situation in quantum estimation is the following: some classical information is encoded into a parametric family of *signal states*, denoted by

$$\mathcal{F}(\Theta) = \{\rho_\theta \in \mathcal{S}(\mathcal{H}) \mid \theta \in \Theta\} \ , \tag{1.10}$$

where $\theta \in \Theta$ a multidimensional parameter—that we call here the *signal parameter*—and Θ is some measure space. The aim of the estimation strategy is to extract from the unknown signal state ρ_θ the value of some classical parameter $\omega \in \Omega$, which is generally a function of θ. In particular, the most common case is that of *state estimation*, in which $\omega = \theta$, and the goal is to produce the best possible guess of the state of the system under the premise that the latter is prepared in a state of the parametric family $\mathcal{F}(\Theta)$.

An estimation strategy consists of two stages, the first being a quantum measurement, which extracts from the system an array of experimental data, and the second being a classical data analysis, that produces the final estimate of the parameter ω. The concept of POVM is particularly useful for optimization, since it allows to represent with a single mathematical object both the quantum measurement and the classical data processing. An estimation strategy is completely specified by its POVM $P : \sigma(\Omega) \to \mathcal{B}(\mathcal{H}),\ B \longmapsto P(B)$, and the conditional probability of obtaining the event B if the state is ρ_θ is given by the Born rule

$$p(B|\theta) = \mathrm{Tr}[P(B)\rho_\theta] \ . \tag{1.11}$$

In the following, we will indicate a POVM also the differential notation $P(\mathrm{d}\omega)$, so that, by definition

$$P(B) = \int_B P(\mathrm{d}\omega) \qquad \forall B \in \sigma(\Omega) \ . \tag{1.12}$$

1.2.1 Possible definitions of optimality

The starting point of the program of quantum estimation theory is a rigorous definition of an optimality criterion which makes quantitatively precise intuitive expressions such as “most efficient”, or “most accurate” strategy. Suppose that the system is prepared in the unknown state ρ_θ, and, accordingly, the parameter of interest takes the value $\omega = \omega(\theta)$. Consider an estimation strategy $P(\mathrm{d}\hat{\omega})$ that produces the estimate $\hat{\omega}$ according to the probability distribution $p(\mathrm{d}\hat{\omega}|\theta) = \mathrm{Tr}[P(\mathrm{d}\hat{\omega})\rho_\theta]$. Of course, the estimation strategy is appreciated as much as the estimate it provides is near to the true value. This idea can be made precise by introducing a *cost function*[1, 2] $c(\hat{\omega}, \omega)$, which quantify the “cost” of estimating $\hat{\omega}$ if the true value of the parameter is ω. The cost function achieves its minimum if $\hat{\omega} = \omega$, and increases monotonically with the deviation between the estimate and the true parameter.

We define the *average cost* $\bar{c}(\theta)$ of the estimation strategy $P(\mathrm{d}\hat{\omega})$ in the state ρ_θ as

$$\bar{c}(\theta) \doteq \int_\Omega c(\hat{\omega}, \omega(\theta)) \ \mathrm{Tr}[P(\mathrm{d}\omega)\rho_\theta] \ . \tag{1.13}$$

Since the value of the parameter θ is unknown, one would like to minimize the average cost $c(\theta)$ for any possible value $\theta \in \Theta$. However, this is generally impossible in quantum mechanics, where the allowed probability distributions are only those which are induced by the Born rule $p(\mathrm{d}\hat{\omega}|\theta) = \mathrm{Tr}[P(\mathrm{d}\hat{\omega})\theta]$. Due to quantum noncommutativity, the minimization of the average cost for a certain value θ_1 of the signal parameter is incompatible with the minimization for a different value θ_2. Therefore, optimality has to be defined in a different way.

The remaining Subsections present three fundamental approaches to the definition of optimality.

1.2.2 Bayesian approach

In this approach, a *prior distribution* $\nu(\mathrm{d}\theta)$ for the unknown signal parameter is introduced. The meaning of the prior distribution can be either subjective (if it reflects the experimenter's ignorance about the value of the signal parameter), or objective (if it reflects the characteristic probability of emission of signals from a given source). In both cases, the optimal estimation strategy is defined as the one that that minimizes the *Bayes expected cost*

$$\langle c \rangle \doteq \int_{\Theta} \nu(\mathrm{d}\theta)\ \bar{c}(\theta)\ , \tag{1.14}$$

i.e. the expectation value of expression (1.13) over the unknown signal parameters.

Notice that the expected cost is a linear functional of the probability distribution $p_\theta(\mathrm{d}\hat{\omega})$, and, therefore it is a linear functional of the POVM $P(\mathrm{d}\hat{\omega})$, i.e. if $P(\mathrm{d}\omega) = pP_1(\mathrm{d}\omega)+(1-p)P_2(\mathrm{d}\omega)$ for some probability $p \in [0,1]$, then one has

$$\langle c \rangle = p\langle c \rangle_1 + (1-p)\langle c \rangle_2\ . \tag{1.15}$$

Since the minimization of the expected cost is a linear optimization problem, the characterization of the extreme points of the convex set of POVMs with outcome space Ω becomes particularly useful for optimization. This topic will be treated in Chapter 8.

1.2.3 Mimimax approach

No prior distribution is introduced in this case: the optimal estimation strategy is defined as the one that minimizes the *worst case cost*

$$c_{\max} \doteq \max_{\theta \in \Theta} \bar{c}(\theta) \tag{1.16}$$

(in the case of Θ being noncompact, the maximum over θ has to be replaced with a supremum). Contrarily to what happens in the Bayesian approach, the search for the minimax POVM is a not a linear optimization problem. The worst case cost is indeed a convex functional of the POVM, i.e. if

$P(\mathrm{d}\omega) = pP_1(\mathrm{d}\omega) + (1-p)P_2(\mathrm{d}\omega)$ for some probability $p \in [0,1]$, then one has

$$c_{\max} \leq p\ c_{\max,1} + (1-p)\ c_{\max,2}\ . \tag{1.17}$$

Unfortunately, in this case the optimal POVM cannot be searched among the set of extremal POVMs. The minimum of the worst-case cost can be indeed achieved by a POVM in the interior of the convex set of POVMs. For example, consider the two states

$$\rho_1 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \qquad \rho_2 = \frac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \tag{1.18}$$

and the cost function $c(i,j) = 1 - \delta_{ij}$, corresponding to the criterion of minimum error probability. It is simple to check that the unique minimax POVM for this problem is the one given by

$$P_1 = \begin{pmatrix} 2/3 & 0 \\ 0 & 0 \end{pmatrix} \qquad P_2 = \begin{pmatrix} 1/3 & 0 \\ 0 & 1 \end{pmatrix}\ ,$$

achieving the worst-case cost $c_{\max} = 1/3$. This POVM is clearly not extremal, its decomposition into extremals being $P_i = 1/3\ Q_i + 2/3\ R_i$, where

$$Q_1 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \quad Q_2 = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}\ , \qquad R_1 = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \quad R_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\ . \tag{1.19}$$

It is worth stressing that in general the optimal estimation strategy in the Bayesian approach does not coincide with the optimal strategy in the minimax. In the previous example, assuming an equal a priori probability for the two states of Eq. (1.18), the optimal POVM in the Bayesian approach is the POVM Q_i of Eq. (1.19). Such a POVM is extremal and has Bayes cost $\langle c \rangle = 1/4$. Moreover, we also stress that the optimal strategy for a given cost function may be no longer optimal if we change the choice of the cost function. The dependence on the choice of the cost function and the discrepancy between the Bayesian and the minimax approach are two somewhat natural facts, since different approaches and different cost functions reflect the different practical situations in which an estimation strategy is needed. Nevertheless, in the presence of group symmetry the Bayesian approach with uniform prior and the minimax do coincide (see Section 3.3), and, in addiction, in many relevant situations it is possible to show that a large number of different cost functions lead to the same optimal strategy (see Section 6.1).

1.2.4 Mutual information

Besides the definitions of optimality involving cost functions, there is another possible definition, which is important in information theoretic contexts, and is based on the concept of *mutual information*. In this case one needs to define a prior distribution $\nu(\mathrm{d}\theta)$, which is generally interpreted as the characteristic probability distribution of the source emitting the signals. Suppose that both $\nu(\mathrm{d}\theta)$ and $P(\mathrm{d}\omega)$ admit a density with respect to the Lebesgue measures $\mathrm{d}\theta$ and $\mathrm{d}\omega$, namely $\nu(\mathrm{d}\theta) = n(\theta)\mathrm{d}\theta$ and $P(\mathrm{d}\omega) = M(\omega)\mathrm{d}\omega$, for some operator-valued function $M(\omega) \in \mathcal{B}(\mathcal{H})$. In this way, one can introduce the joint probability density $p(\omega, \theta) = n(\theta)\mathrm{Tr}[\rho_\theta \; m(\omega)]$ and its marginals $p_1(\omega) = \int_\Theta \mathrm{d}\theta \; p(\theta, \omega)$ and $p_2(\theta) = \int_\Omega \mathrm{d}\omega \; p(\theta, \omega) \equiv n(\theta)$. The mutual information is then defined as

$$M[p] \doteq H[p] - H[p_1] - H[p_2] \; , \tag{1.20}$$

where $H[q]$ denotes the Shannon entropy of the probability density $q(x)$, $x \in X$ namely

$$H[q] \doteq - \int_X \mathrm{d}x \; q(x) \log q(x) \; . \tag{1.21}$$

In this context, the optimal POVM is defined as the one that maximizes the value of the mutual information.

Notice that, since the mutual information is nonlinear in the probabilities, it is also nonlinear in the POVM $M(\mathrm{d}\omega)$. However, it is possible to prove that the mutual information is a convex functional of the probability density, and, therefore, it is a convex functional of the POVM, as it was pointed out in a seminal paper by Davies [31]. As a consequence, the maximization of the mutual information can be performed on the extreme points of the convex set of POVMs.

Chapter 2

Fundamental notions of representation theory

Representation theory is an exceedingly vast subject. Here we will summarize the basic definitions and results that are fundamental for the purposes of this presentation, and, more generally, that turn out to be useful for many applications in Quantum Information Theory.

2.1 Generalities

Definition 2. *A* group $\mathbf{G}$ *is a set equipped with an internal composition law* $(g_1, g_2) \to g_1 g_2$ *that enjoys the properties:*

$$g_1(g_2 g_3) = (g_1 g_2) g_3 \qquad \forall g_1, g_2, g_3 \in \mathbf{G} \tag{2.1}$$

$$\exists e \in \mathbf{G} : ge = eg = g \qquad \forall g \in \mathbf{G} \tag{2.2}$$

$$\forall g \in \mathbf{G} \quad \exists g^{-1} \in \mathbf{G} : gg^{-1} = g^{-1} g = e \tag{2.3}$$

In this presentation we will consider finite groups and Lie groups. A finite group is a group with a finite number of elements. A Lie group is a group which is also a differentiable manifold, namely it is parameterized by a chart of local coordinates, with the property that the inverse g^{-1} and the composition hg with an arbitrary group element h are differentiable functions of g. Since Lie groups are manifolds, they can be either compact or non-compact. For example, the groups $U(1)$ and $\mathbb{SU}(2)$ are compact, as they are diffeomorphic to a circle and to a sphere, respectively. The additive group $\mathbb{R}$ and the group $\mathbb{SU}(1,1)$ are instead non-compact, being diffeomorphic to a line and to a three-dimensional hyperboloid, respectively. Compact Lie groups are the simplest examples of continuous groups, and for many respects they are closely analogous to finite groups.

2.2 Unitary and projective representations

Definition 3. *Let* $\mathbf{G}$ *be a group and* $\mathcal{H}$ *a Hilbert space. A unitary representation of* $\mathbf{G}$ *is a function* $U : \mathbf{G} \rightarrow \mathcal{B}(\mathcal{H}), \quad g \rightarrow U_g$ *where* $\{U_g\}$ *are unitary operators enjoying the properties*

$$\begin{aligned} U_{g_1 g_2} &= U_{g_1} U_{g_2} \qquad \forall g_1, g_2 \in \mathbf{G} && (2.4) \\ U_e &= \mathbb{1} && (2.5) \end{aligned}$$

Notice that the unitaries $\{U_g\}$ form themselves a group: if the correspondence $g \longleftrightarrow U_g$ is one-to-one such a group is isomorphic to $\mathbf{G}$.

The action of the group on the Hilbert space $\mathcal{H}$ induces a transformation on the set of quantum states $\mathcal{S}(\mathcal{H})$, which is given by the automorphism

$$\rho \longrightarrow \mathcal{A}_g(\rho) = U_g \rho U_g^\dagger \ . \tag{2.6}$$

The automorphism $\mathcal{A}_g$ represents a *symmetry* of the state space of the quantum system, in the sense of Wigner [38]. Actually, this is the precise definition of the term "symmetry" as it is used throughout this presentation (and even in the title).

From Definition 3 immediately follow the group properties of the state transformations, namely

$$\begin{aligned} \mathcal{A}_{g_1 g_2}(\rho) &= \mathcal{A}_{g_1}(\mathcal{A}_{g_2}(\rho)) && (2.7) \\ \mathcal{A}_e(\rho) &= \rho && (2.8) \\ \mathcal{A}_{g^{-1}}(\mathcal{A}_g(\rho)) &= \rho \ . && (2.9) \end{aligned}$$

The automorphisms $\{\mathcal{A}_g\}$ form indeed a representation of the symmetry group $\mathbf{G}$ acting on the state space of the quantum system.

To guarantee the group properties of state transformations, however, the representation $\{U_g\}$ has not necessarily to be unitary. In the most general case, $\{U_g\}$ can be a *projective representation*:

Definition 4. *A representation* $g \longrightarrow U_g$ *is called* projective *if* $\{U_g\}$ *are unitary operators such that*

$$\begin{aligned} U_{g_1} U_{g_2} &= \omega(g_1, g_2)\ U_{g_1 g_2} \quad \forall g_1, g_2 \in \mathbf{G} && (2.10) \\ U_e &= \mathbb{1}\ , && (2.11) \end{aligned}$$

where $\omega(g_1, g_2)$ *is a phase factor, i.e.* $|\omega(g_1, g_2)| = 1$.

The phase factor $\omega(g_1, g_2)$ is called *cocycle.* In order to have a consistent definition, it must satisfy the relations

$$\omega(g_1, g_2g_3)\omega(g_2, g_3) = \omega(g_1, g_2)\omega(g_1g_2, g_3) \qquad \forall g_1, g_2, g_3 \in \mathbf{G} \ , \tag{2.12}$$

and

$$\omega(e, g) = \omega(g, e) = 1 \qquad \forall g \in \mathbf{G} \ . \tag{2.13}$$

Notice that unitary representations are a special case of projective representations, corresponding to the trivial cocycle $\omega(g_1, g_2) \equiv 1 \quad \forall g_1, g_2 \in \mathbf{G}$.

It is important to stress that different representations of the same group may have different cocycles. Accordingly, we have the following

Definition 5. *We say that two projective representations $\{U_g\}$ and $\{V_g\}$ are in the same* factor system *if they have the same cocycle, i.e. $\omega_U(g_1, g_2) = \omega_V(g_1, g_2) \quad \forall g_1, g_2 \in \mathbf{G}$.*

2.3 Irreducible representations and Schur lemmas

Definition 6. *Let $\mathcal{H}$ be a Hilbert space and $\{U_g \mid g \in \mathbf{G}\}$ be a projective representation. An* invariant *subspace $\mathcal{W} \subseteq \mathcal{H}$ is a subspace with the property $U_g\mathcal{W} = \mathcal{W} \quad \forall g \in \mathbf{G}$.*

Definition 7. *Let $\mathcal{W}$ be an invariant subspace. The representation $\{U_g\}$ is called* irreducible in $\mathcal{W}$ *if there exists no proper subspace $\mathcal{V}$, $\{0\} \neq \mathcal{V} \subset \mathcal{W}$ which is invariant.*

Definition 8. *A subspace carrying an irreducible representation is called* irreducible subspace.

Irreducible representations will be called *irrepS* for short. They are the fundamental building blocks on which a group representation can be decomposed. In particular, for finite and compact Lie groups we have the following:

Theorem 2. *Let $\mathbf{G}$ be a finite group, or alternatively, a compact Lie group, and $\{U_g \mid g \in \mathbf{G}\}$ a projective representation of $\mathbf{G}$ on the Hilbert space $\mathcal{H}$. Then, any unitary U_g can be decomposed into the direct sum of a discrete number of irreps.*

All irreps in the decomposition of a projective representation $\{U_g\}$ have necessarily the same cocycle, as it can be seen by taking on both sides in Eq. (2.10) the restriction to the irreducible subspaces. In the case of noncompact groups, however, Theorem 2 is no longer valid. For example, the representation of the additive group $\mathbb{R}$ given by $U_x = e^{-ixP}$, where $P = -i\frac{d}{dx}$ is the momentum operator in the Hilbert space $L^2(\mathbb{R})$, cannot be decomposed in a discrete number of irreps. In fact, in terms of the Dirac-normalized eigenstates of P, one has

$$U_x = \int_{\mathbb{R}} \mathrm{d}p \; e^{-ipx} \; |p\rangle\langle p| \;, \tag{2.14}$$

namely the representation is decomposed as a direct integral of the one-dimensional irreps $\{U_x^p = e^{-ipx} \mid x \in \mathbb{R}\}$ labeled by the eigenvalues of P. In the general part of this presentation we will always consider the case of representations that can be decomposed in a direct sum of irreps, leaving the case of direct integrals to specific examples (Section 4.4 for the case of complex displacement and Subsection 6.2.3 for the case of squeezing).

Definition 9. *Two projective irrepS $\{U_g\}$ and $\{V_h\}$ acting in the Hilbert spaces $\mathcal{H}$ and $\mathcal{K}$, respectively, are called* equivalent *if there exist an isomorphism $T : \mathcal{H} \to \mathcal{K}$ such that $T^\dagger T = 1\!\!1_{\mathcal{H}}, \quad TT^\dagger = 1\!\!1_{\mathcal{K}}$, and $TU_g = V_g T \quad \forall g \in$ $\mathbf{G}$.*

Notice that two equivalent representations have necessarily the same cocycle, due to the relation $V_g = TU_g T^\dagger$. The isomorphism T is usually called *intertwiner*. Equivalent irreps can by grouped in the same equivalence class, the latter being labeled by the Greek index μ.

Now we present the two Schur lemmas, whose proof can be found in any standard textbook on group theory. The first Schur lemma regards operators that commute with two equivalent irreps:

Lemma 1. *Let $\{U_g\}$ and $\{V_g\}$ be to equivalent projective irreps. Then, any operator $O : \mathcal{H} \to \mathcal{K}$ such that $OU_g = V_g O \quad \forall g \in \mathbf{G}$ has the form $O = \lambda T$ where $\lambda \in \mathbb{C}$ is a constant, while $T : \mathcal{H} \to \mathcal{K}$ is the isomorphism connecting the two irreps.*

In the case of inequivalent irrepS one has the second Schur lemma:

Lemma 2. *Let $\{U_g\}$ and $\{V_g\}$ be two inequivalent projective irreps. Then, the only operator $O : \mathcal{H} \to \mathcal{K}$ such that $OU_g = V_g O \quad \forall g \in \mathbf{G}$ is the null operator $O = 0$.*

The two Schur lemmas are the basic building blocks to obtain the general form of an operator commuting with a group representation. Consider a group representation $\{U_g\}$ acting in the Hilbert space $\mathcal{H}$, which can be decomposed in a discrete number of irreps, each of them acts in a different subspace

$$U_g = \bigoplus_{\mu \in \mathsf{S}} \bigoplus_{i=1}^{m_\mu} U_g^{\mu,i} \ . \tag{2.15}$$

Here S is the set of equivalence classes of irrepS contained in the decomposition of $\{U_g\}$. All the irreps in the same equivalence class are associated with the same index μ, while different irreps in the same equivalence class are tagged with the index i. The number m_μ of equivalent irreps in the same equivalence class is called *multiplicity*. The decomposition of the representation $\{U_g\}$ into irreps is associated with the decomposition of the Hilbert space into irreducible orthogonal subspaces

$$\mathcal{H} = \bigoplus_{\mu \in \mathsf{S}} \bigoplus_{i=1}^{m_\mu} \mathcal{H}_i^\mu \ . \tag{2.16}$$

According to Definition 9, two subspaces $\mathcal{H}_i^\mu$ and $\mathcal{H}_j^\mu$ carrying equivalent irreps are connected by the intertwiner $T_{ij}^\mu : \mathcal{H}_j^\mu \to \mathcal{H}_i^\mu$. The intertwiners can be always chosen in such a way that $T_{ij}^{\mu\dagger} = T_{ji}^\mu \quad \forall \mu \in \mathsf{S}, \forall i,j = 1, \dots m_\mu$. According to this notation, the operator T_{ii}^μ will be the projector onto the subspace $\mathcal{H}_i^\mu$. By definition, any intertwiner commutes with U_g, namely $[T_{ij}^\mu, U_g] = 0 \ \forall g \in \mathbf{G}$

Consider now a generic operator O in the commutant of $\{U_g\}$, i.e. an operator such that $[O, U_g] = 0 \quad \forall g \in \mathbf{G}$. The Schur lemmas allows to obtain its general form in a straightforward way:

Corollary 1. *Let $O \in \mathcal{B}(\mathcal{H})$ be an operator such that $[O, U_g] = 0 \quad \forall g \in \mathbf{G}$. Then, O has the form*

$$O = \bigoplus_{\mu \in \mathsf{S}} \bigoplus_{i=1}^{m_\mu} \lambda_{ij}^\mu \ T_{ij}^\mu \ , \tag{2.17}$$

where $\lambda_{ij}^\mu \in \mathbb{C}$ are suitable constants.

Proof. Use the resolution of the identity $\mathbb{1} = \bigoplus_{\mu \in \mathsf{S}} \bigoplus_{i=1}^{m_\mu} \ T_{ii}^\mu$ to write $O = \bigoplus_{\mu,\nu} \bigoplus_{i=1}^{m_\mu} \bigoplus_{j=1}^{m_\nu} \ T_{ii}^\mu O T_{jj}^\nu$. Applying the first Schur lemma one obtain $T_{ii}^\mu O T_{jj}^\mu = \lambda_{ij}^\mu T_{ij}^\mu$ for some constant λ_{ij}^μ, while applying the second one obtains $T_{ii}^\mu O T_{jj}^\nu = 0$ for $\mu \neq \nu$. ■

2.4 The Clebsch-Gordan tensor product structure

The general expression of an operator O in the commutant, given by Corollary 1, can be restated in a more elegant form by introducing an abstract tensor product structure in the Hilbert space. To this purpose, consider the orthonormal bases $\mathcal{B}_i^\mu = \{|\mu, i, n\rangle \mid n = 1, \dots, d_\mu\}$ and $\mathcal{B}_j^\mu = \{|\mu, j, n\rangle \mid n = 1, \dots, d_\mu\}$ for the subspaces $\mathcal{H}_i^\mu$ and $\mathcal{H}_j^\mu$, chosen in such a way that

$$|\mu, i, n\rangle = T_{ij}^\mu \, |\mu, j, n\rangle \qquad \forall n = 1, \dots, d_\mu \ . \tag{2.18}$$

This choice of bases makes evident the isomorphism between the direct sum $\bigoplus_{i=1}^{m_\mu} \mathcal{H}_i^\mu$ and the tensor product $\mathcal{H}_\mu \otimes \mathbb{C}^{m_\mu}$, where $\mathcal{H}_\mu$ is an abstract space of dimension d_μ. In fact, we can make the identification

$$|\mu, i, n\rangle \cong |\mu, n\rangle \otimes |i\rangle \ , \tag{2.19}$$

where $\{|\mu, n\rangle \mid n = 1, \dots d_\mu\}$ and $\{|i\rangle \mid i = 1, \dots, m_\mu\}$ are orthonormal bases for $\mathcal{H}_\mu$ and $\mathbb{C}^{m_\mu}$, respectively. With this identification, the decomposition of the Hilbert space of Eq. (2.16) yields the *Clebsch-Gordan tensor product structure* (TPS for short)

$$\mathcal{H} = \bigoplus_{\mu \in \mathsf{S}} \mathcal{H}_\mu \otimes \mathbb{C}^{m_\mu} \ , \tag{2.20}$$

i.e. the direct sum over equivalent representations in Eq. (2.16) is replaced here by the tensor product with the space $\mathbb{C}^{m_\mu}$. In this notation the intertwiner T_{ij}^μ has the simple form

$$T_{ij}^\mu = \mathbb{1}_{d_\mu} \otimes |i\rangle\langle j| \ , \tag{2.21}$$

where $\mathbb{1}_{d_\mu}$ denotes the identity in the abstract d_μ-dimensional space $\mathcal{H}_\mu$. The decomposition of the representation $\{U_g\}$ given in Eq. (2.15) becomes now

$$U_g = \bigoplus_{\mu \in \mathsf{S}} U_g^\mu \otimes \mathbb{1}_{m_\mu} \ , \tag{2.22}$$

where $\{U_g^\mu\}$ is an irreducible representation acting in the abstract space $\mathcal{H}_\mu$, while $\mathbb{1}_{m_\mu}$ is the identity in $\mathbb{C}^{m_\mu}$. Due to this decomposition, the spaces $\mathcal{H}_\mu$ and $\mathbb{C}^{m_\mu}$ are usually called *representation space* and *multiplicity space*, respectively. Once the unitaries $\{U_g\}$ are decomposed as above, the form of an arbitrary operator O in the commutant becomes straightforward:

$$O = \bigoplus_{\mu \in \mathsf{S}} \mathbb{1}_{d_\mu} \otimes O_\mu \ , \tag{2.23}$$

where O_μ is an operator acting in the multiplicity space $\mathbb{C}^{m_\mu}$ (this form could also be obtained inserting the expression (2.21) into Eq. (2.17), and defining $O_\mu = \sum_{i,j} \lambda^\mu_{ij} \ |i\rangle\langle j|$).

In finite dimensions, the Clebsch-Gordan decomposition of the group representation $\{U_g\}$ given by Eq. (2.22) is a particular example of the so-called *Wedderburn decomposition* of matrix algebras [13], which has been extensively used in the field of quantum error correction to construct *decoherence free subspaces* and *noiseless subsystems* [32, 33, 34]. In our case, if we consider the unitaries $\{U_g\}$ as noise operators and take the matrix algebra generated by their linear combinations, we have that the multiplicity spaces $\mathbb{C}^{m_\mu}$ are the noiseless subsystems of the noise algebra, i.e. they represent degrees of freedom that are unaffected by noise. The decoherence free subspace is the multiplicity space associated to the trivial representation $U^{\mu_0}_g = 1 \quad \forall g \in \mathbf{G}$. The above identification is far from being formal: in many real situations a quantum system may be subjected to an unknown group transformation, and the presence of multiplicity spaces allows to preserve quantum information from the depolarizing effect of such a noise.

2.5 Invariant measures over a Lie group

Consider a function $f : \mathbf{G} \longmapsto \mathbb{C}, g \mapsto f(g)$. If the group $\mathbf{G}$ has a finite number of elements, then one can construct from f a constant function simply by taking the average:

$$\overline{f}(g) = \frac{1}{|\mathbf{G}|} \sum_{h \in \mathbf{G}} f(hg) \ . \tag{2.24}$$

In the same way, given a vector $|\psi\rangle \in \mathcal{H}$ and a unitary representation $\{U_g\}$, one can construct another vector which is invariant under the group action by defining

$$\overline{|\psi\rangle} = \frac{1}{|\mathbf{G}|} \sum_{h \in \mathbf{G}} U_h |\psi\rangle \ . \tag{2.25}$$

With this definition, one has indeed $U_g \overline{|\psi\rangle} = \overline{|\psi\rangle} \quad \forall g \in \mathbf{G}$.

Finally, starting from an operator O, the group average over the action of a projective representation $\{U_g\}$ yields the invariant operator $\overline{O}$ given by

$$\overline{O} = \frac{1}{\mathbf{G}} \sum_{h \in \mathbf{G}} U_h O U_h^\dagger \ . \tag{2.26}$$

It is immediate to verify that $U_g \overline{O} U_g^\dagger = \overline{O} \quad \forall g \in \mathbf{G}$, i.e. that $\overline{O}$ is in the commutant of $\{U_g\}$.

The three examples of group averages just mentioned appear often in applications. For example, group invariant vectors are relevant in quantum error correction, since they span the decoherence free subspaces of the Hilbert space, and also for quantum algorithms, since they generalize the Fourier transform [35]. In order to extend these examples to the general case of continuous groups one need to substitute the sums with integrals, and for this the concept of invariant measure is needed.

Let $\mathbf{G}$ be a Lie group. For any fixed group element $h \in \mathbf{G}$, the map $g \longmapsto hg$ is a diffeomorphism, and transforms the region $B \subseteq \mathbf{G}$ in the region $hB = \{hg \mid g \in B\}$.

Definition 10. *A measure $\mu_L(\mathrm{d}g)$ on the Lie group $\mathbf{G}$ is called* left-invariant *if $\mu_L(gB) = \mu_L(B)$, for any group element $g \in \mathbf{G}$ and for any region $B \subseteq \mathbf{G}$.*

In the same way, one can define a *right-invariant* measure $\mu_R(\mathrm{d}g)$:

Definition 11. *A measure $\mu_R(\mathrm{d}g)$ on the Lie group $\mathbf{G}$ is called* right-invariant *if $\mu_R(gB) = \mu_R(B)$, for any group element $g \in \mathbf{G}$ and for any region $B \subseteq \mathbf{G}$.*

Any Lie group admits a left-invariant measure and a right-invariant one, which are both unique up to a multiplicative constant. However, the two invariant measures does not coincide necessarily: if $\mu_L = \mu_R$ (up to a constant) the group is called *unimodular*, otherwise the group is called nonunimodular. All compact groups and almost all noncompact groups that are important for applications are unimodular. Nevertheless, there are also interesting groups that are nonunimodular, as for example the affine group "$ax + b$" of translations and dilations on the real line, which has be extensively studied in the research about wavelets [36]. The estimation of the signal generated by the action of a nonunimodular group on a quantum system will be the subject of Chapter 7. In the rest of this presentation, the group $\mathbf{G}$ will be always assumed to be unimodular, and the invariant measure $\mu_L(\mathrm{d}g) = \mu_R(\mathrm{d}g)$ will be denoted for short as $\mathrm{d}g$.

Using the invariant measure it is possible to introduce the group average of a function $f : \mathbf{G} \to \mathbb{C}$

$$\overline{f}(g) = \int_{\mathbf{G}} \mathrm{d}h \; f(gh) \;, \tag{2.27}$$

of a vector $|\psi\rangle \in \mathcal{H}$

$$\overline{|\psi\rangle} = \int_{\mathbf{G}} \mathrm{d}h \; U_h|\psi\rangle \;, \tag{2.28}$$

and of an operator $O \in \mathcal{B}(\mathcal{H})$

$$\overline{O} = \int_{\mathbf{G}} \mathrm{d}h \; U_h O U_h^\dagger \;. \tag{2.29}$$

The convergence of the above integrals is always guaranteed in the case of compact groups, while for non-compact groups it needs additional hypotheses, which will be considered in the next section. At any rate, it is immediate to see that, when the integrals do converge, they yield group invariant functions, vectors[1] (for unitary representations), and operators (for generally projective representations).

2.6 Group average of an operator

The aim of this section is to give a useful formula for the group average of an operator. We start with the simplest cases of finite and compact groups, then generalizing the formula to the case of noncompact Lie groups, in the case of representations that can be decomposed in a discrete direct sum of irreps.

When dealing with compact groups we will always assume the invariant measure $\mathrm{d}g$ to be normalized in such a way that

$$\int_{\mathbf{G}} \mathrm{d}g = 1 \ . \tag{2.30}$$

This allows to treat at the same time both compact and finite groups, just by making the substitution

$$\int_{\mathbf{G}} \mathrm{d}g \longleftrightarrow \frac{1}{\mathbf{G}} \sum_{g \in \mathbf{G}} \ . \tag{2.31}$$

Proposition 1. *Let* $\mathbf{G}$ *be a compact (finite) group, with projective representation* $\{U_g = \bigoplus_{\mu \in \mathsf{S}} U_g^\mu \otimes \mathbb{1}_{m_\mu}\}$ *on the Hilbert space* $\mathcal{H}$*. Let* O *be an operator on* $\mathcal{H}$*, and* $\overline{O} = \int_{\mathbf{G}} \mathrm{d}h \; U_h O U_h^\dagger$ *its group average (with substitution* (2.31) *in the case of* $\mathbf{G}$ *being finite). Then,*

$$\overline{O} = \bigoplus_{\mu \in \mathsf{S}} \mathbb{1}_{d_\mu} \otimes \frac{\mathrm{Tr}_{\mathcal{H}_\mu}[O]}{d_\mu} \ , \tag{2.32}$$

where $\mathrm{Tr}_{\mathcal{H}_\mu}$ *denotes the partial trace over the first factor in the tensor product* $\mathcal{H}_\mu \otimes \mathbb{C}^{m_\mu}$ *of the decomposition* (2.20)*.*

Proof. Since the average $\overline{O}$ is in the commutant of the representation $\{U_g\}$, it has the form (2.23), i.e. $\overline{O} = \bigoplus_{\mu \in \mathsf{S}} \mathbb{1}_{d_\mu} \otimes \overline{O}_\mu$. Projecting onto the subspace

[1] Notice that, if there are no group invariant vectors in the Hilbert space, the group average necessarily outputs the zero vector, i.e. $\overline{|\psi\rangle} = 0, \quad \forall |\psi\rangle \in \mathcal{H}$.

$\mathcal{H}_\mu \otimes \mathbb{C}^{m_\mu}$ and taking the partial trace over $\mathcal{H}_\mu$, we get $\mathrm{Tr}_{\mathcal{H}_\mu}[\ \overline{O}\ ] = d_\mu \overline{O}_\mu$. On the other hand, we have

$$\mathrm{Tr}_{\mathcal{H}_\mu}[\overline{O}_\mu] = \int_{\mathbf{G}} \mathrm{d}g\ \mathrm{Tr}_{\mathcal{H}_\mu}[U_g^\mu \otimes \mathbb{1}_{m_\mu}\ O\ U_g^{\mu\dagger} \otimes \mathbb{1}_{m_\mu}] = \mathrm{Tr}_{\mathcal{H}_\mu}[O]\ ,$$

the last equality following from the cyclic property of trace and the normalization of the Haar measure. Hence, formula (2.32) holds. ■

In order to generalize the above formula to the case of non-compact Lie groups, we have to consider the convergence of the integral defining the group average. To this purpose, we first introduce the notion of *square-summable* irrep:

Definition 12. *Let* $\mathbf{G}$ *be a locally compact Lie group. A projective irrep* $\{U_g\}$ *acting in the Hilbert space* $\mathcal{H}$ *is called* square-summable *if there is a non-zero vector* $|\psi\rangle \in \mathcal{H}$ *such that*

$$\int_{\mathbf{G}} \mathrm{d}g\ |\langle\psi|U_g|\psi\rangle|^2 < \infty\ . \tag{2.33}$$

A vector $|\psi\rangle$ satisfying the property (2.33) is called *admissible.* An important property of unimodular groups is given by the following

Theorem 3. *Let* $\{U_g\}$ *be a square-summable irrep. Then, any vector in* $\mathcal{H}$ *is admissible.*

The proof of this result can be found for example in Ref. [37] as a special case of a general theorem about locally compact groups.

Suppose now that the projective representation $\{U_g\}$ acting in the Hilbert space $\mathcal{H}$ is reducible, and in addition that it can decomposed as a direct sum of square-summable irreps, i.e. $U_g = \bigoplus_{\mu\in\mathsf{S}} U_g^\mu \otimes \mathbb{1}_{m_\mu}$. Then, the formula for the group average becomes

Proposition 2. *Let* $\mathbf{G}$ *be a unimodular Lie group and* $U_g = \bigoplus_{\mu\in\mathsf{S}} U_g^\mu \otimes \mathbb{1}_{m_\mu}$ *a projective representation. Then the group average* $\overline{O}$ *of a given operator* O *is given by*

$$\overline{O} = \bigoplus_{\mu\in\mathsf{S}}\ \mathbb{1}_{\mathcal{H}_\mu} \otimes \frac{\mathrm{Tr}_{\mathcal{H}_\mu}[O]}{d_\mu}\ , \tag{2.34}$$

where d_μ *is the* formal dimension *defined as*

$$d_\mu = \left(\int \mathrm{d}g\ |\langle\psi_\mu|\ U_g^\mu\ |\phi_\mu\rangle|^2\right)^{-1}\ , \tag{2.35}$$

$|\psi_\mu\rangle$ *and* $|\phi_\mu\rangle$ *being any two normalized states in* $\mathcal{H}_\mu$.

Proof. Since the group average $\overline{O}$ of an operator is in the commutant of the representation $\{U_g\}$ it has the form $\overline{O} = \bigoplus_\mu \ \mathbb{1}_{\mathcal{H}_\mu} \otimes O_\mu$, for some suitable operators O_μ acting in the multiplicity space. Taking the expectation value with respect to a normalized vector $|\psi_\mu\rangle \in \mathcal{H}_\mu$, we obtain $O_\mu = \langle\psi_\mu| \, \overline{O} \, |\psi_\mu\rangle = \mathrm{Tr}_{\mathcal{H}_\mu}[O \ \overline{B_\mu}]$, where $B_\mu = |\psi_\mu\rangle\langle\psi_\mu| \otimes \mathbb{1}_{\mathcal{M}_\mu}$. Now, since the group average $\overline{B}_\mu$ is in the commutant of $\{U_g\}$, and since $|\psi_\mu\rangle \in \mathcal{H}_\mu$, we have $\overline{B}_\mu = 1/d_\mu \mathbb{1}_{\mathcal{H}_\mu} \otimes \mathbb{1}_{\mathcal{M}_\mu}$ for some constant d_μ. The constant d_μ is simply evaluated by taking the expectation value of $\overline{B}_\mu$ with respect to any normalized vector $|\phi_\mu\rangle|\alpha_\mu\rangle \in \mathcal{H}_\mu \otimes \mathcal{M}_\mu$. ■

Remark 1. The *formal dimension* of Eq. (2.35) is not a property of the sole Hilbert space $\mathcal{H}_\mu$, but also of the irreducible representation acting on it. Depending on the particular irreducible representations, the same Hilbert space may have different formal dimensions. Moreover, we stress that the formal dimension is always finite, even if the dimension of the Hilbert space is infinite.

Remark 2. According to the formula (2.34), the integral in the group average converges for operators O such that the partial traces $\mathrm{Tr}_{\mathcal{H}_\mu}[O]$ are finite.

Remark 3. Proposition 2 also shows that the square-summable representations of a non-compact group are necessarily infinite-dimensional. Consider indeed the case of $U_g = U_g^\mu$ for some μ and take the trace on both sides of Eq. (2.34): one obtains $vol(\mathbf{G}) \ \mathrm{Tr}[O] = \mathrm{Tr}[\mathbb{1}_{\mathcal{H}_\mu}] \frac{\mathrm{Tr}[O]}{d_\mu}$, where $vol(\mathbf{G})$ is the volume of $\mathbf{G}$, defined as $vol(\mathbf{G}) = \int_{\mathbf{G}} \mathrm{d}g$. Since the group is non-compact, its volume is infinite, whence, necessarily $\mathrm{Tr}[\mathbb{1}_{\mathcal{H}_\mu}] = \infty$, i.e. the representation space $\mathcal{H}_\mu$ is infinite dimensional.

2.7 Orthogonality relations

Consider two projective irreps $\{U_g^\mu\}$ and $\{U_g^\nu\}$ in the same factor system, acting in the representation spaces $\mathcal{H}_\mu$ and $\mathcal{H}_\nu$, respectively. Let O be an operator from $\mathcal{H}_\nu$ to $\mathcal{H}_\mu$, and $\overline{O}$ be its group average $\overline{O} = \int_{\mathbf{G}} \mathrm{d}g \ U_g^\mu O U_g^{\nu\dagger}$. It is immediate to see that $\overline{O}$ commutes with the two representations, namely

$$\begin{aligned} U_g^\mu \, \overline{O} &= \int_{\mathbf{G}} \mathrm{d}h \ \ (U_g^\mu U_h^\mu) \ O \ (U_h^{\nu\dagger} U_g^{\nu\dagger}) \ U_g^\nu \\ &= \int_{\mathbf{G}} \mathrm{d}k \ U_k^\mu O U_k^{\nu\dagger} \ U_g^\nu \\ &= \overline{O} \ U_g^\nu \qquad \forall g \in \mathbf{G} \ . \end{aligned} \tag{2.36}$$

Therefore, if $\mu \neq \nu$, $\overline{O}$ must be zero, due to the second Schur lemma 2. Otherwise, if $\mu = \nu$, the formula for the group average (2.32) ((2.34) in the non-compact case) gives $\overline{O} = \mathbb{1}\ \frac{Tr[O]}{d_\mu}$. Choosing $O = |\mu, m\rangle\langle\nu, n|$ and taking the matrix elements $\langle\mu, l|\overline{O}|\nu, k\rangle$ we easily obtain the following

Theorem 4. *Let* $\mathbf{G}$ *be a compact (finite) group, and* U_g^μ *and* U_g^ν *two projective irreps in the same factor system. Consider the matrix elements* $u_{ml}^\mu(g) = \langle\mu, m|U_g^\mu|\mu, l\rangle$ *and* $u_{nk}^\nu(g) = \langle\nu, n|U_g^\nu|\nu, k\rangle$. *Then, if* $\mathbf{G}$ *is compact*

$$\langle u_{ml}^\mu | u_{nk}^\nu \rangle \doteq \int_{\mathbf{G}} \mathrm{d}g\ u_{ml}^{\mu *}(g) u_{nk}^\nu(g) = \frac{\delta_{\mu,\nu}\delta_{m,n}\delta_{l,k}}{d_\mu}\ , \qquad (2.37)$$

$\delta_{i,j}$ *denoting the usual Kronecker delta. If* $\mathbf{G}$ *is finite, Eq.* (2.37) *holds with the substitution* (2.31).

This Theorem shows that the matrix elements of the irreps of a compact group are a set of orthogonal functions in the Hilbert space $L^2(\mathbf{G})$. In the case of finite groups, the matrix elements are a set of orthogonal vectors in the space $\mathbb{C}^{|\mathbf{G}|}$, the vector $|u_{ij}^\mu\rangle$ being defined by

$$|u_{ij}^\mu\rangle = \sum_{g \in \mathbf{G}} u_{ij}^\mu(g)\ |g\rangle\ , \qquad (2.38)$$

where $\{|g\rangle \mid g \in \mathbf{G}\}$ is an orthonormal basis for $\mathbb{C}^{|\mathbf{G}|}$.

As an immediate corollary of Theorem 4 we have the integral formula

$$\int_{\mathbf{G}} \mathrm{d}g\ u_{ij}^{\mu *}(g)\ U_g^\nu = \frac{\delta_{\mu\nu}}{d_\mu}\ |\mu, i\rangle\langle\mu, j|\ , \qquad (2.39)$$

holding for irreps in the same factor system. More generally, for a reducible representation $U_g = \bigoplus_{\mu \in \mathsf{S}} U_g^\mu \otimes \mathbb{1}_{m_\mu}$ we have

$$\int_{\mathbf{G}} \mathrm{d}g\ u_{ij}^{\mu *}(g)\ U_g = \frac{\alpha_{\mathsf{S}}(\mu)}{d_\mu}\ |\mu, i\rangle\langle\mu, j| \otimes \mathbb{1}_{m_\mu}\ , \qquad (2.40)$$

α_{S} being the characteristic function of the set S ($\alpha_{\mathsf{S}}(\mu) = 1$ for $\mu \in \mathsf{S}$, and $\alpha_{\mathsf{S}}(\mu) = 0$ otherwise).

Finally, introducing the characters $\chi_\mu(g) = \mathrm{Tr}[U_g^\mu]$ we obtain the integral representation of the projector onto the subspace $\mathcal{H}_\mu \otimes \mathbb{C}^{m_\mu}$:

$$\int_{\mathbf{G}} \mathrm{d}g\ \chi_\mu^*(g)\ U_g = \frac{\alpha_{\mathsf{S}}(\mu)}{d_\mu}\ \ \mathbb{1}_{\mathcal{H}_\mu} \otimes \mathbb{1}_{m_\mu}\ . \qquad (2.41)$$

In particular, defining the function $\lambda_{\mathsf{S}}(g)$ as the linear combination

$$\lambda_{\mathsf{S}}(g) = \sum_{\mu \in \mathsf{S}} d_\mu \ \chi_\mu(g) \ , \tag{2.42}$$

we obtain the resolution of the identity

$$\mathbb{1} = \int_{\mathbf{G}} \mathrm{d}g \ \lambda^*_{\mathsf{S}}(g) \ U_g \ . \tag{2.43}$$

Of course, all the above formulas hold for finite groups, with the substitution (2.31).

2.8 Orthogonality of characters

In the case of finite or compact groups, the orthogonality of matrix elements gives immediately the orthogonality of characters

Theorem 5. *Let* $\mathbf{G}$ *be a compact (finite) group, and let* $\{U^\mu_g\}$ *and* $\{U^\nu_g\}$ *be two irreps in the same factor system. If* $\mathbf{G}$ *is compact, the following relation holds*

$$\langle \chi_\mu | \chi_\nu \rangle \doteq \int_{\mathbf{G}} \mathrm{d}g \ \chi^*_\mu(g) \ \chi_\nu(g) = \delta_{\mu,\nu} \tag{2.44}$$

If $\mathbf{G}$ *is finite, the same relation holds with the substitution* (2.31)*.*

Proof. By definition, $\chi_\mu = \sum_{i=1}^{d_\mu} u^\mu_{ii}$. Then, applying Eq. (2.37), one get Eq. (2.44). ■

This Theorem shows that the irreducible characters are orthonormal functions in $L^2(\mathbf{G})$. For finite groups, the characters are orthonormal vectors in $\mathbb{C}^{|\mathbf{G}|}$.

As a simple consequence of the orthogonality (2.44), for a reducible representation $U_g = \bigoplus_{\mu \in \mathsf{S}} U^\mu_g \otimes \mathbb{1}_{m_\mu}$ with character $\chi(g) \doteq \mathrm{Tr}[U_g]$, and for any irrep $\{U^\mu_g\}$ in the same factor system of $\{U_g\}$, we have the relation

$$\langle \chi_\mu | \chi \rangle = \int_{\mathbf{G}} \mathrm{d}g \ \chi^*_\mu(g) \chi(g) = m_\mu \ , \tag{2.45}$$

with $m_\mu = 0$ if μ does not appear in the decomposition of $\{U_g\}$. In other words, the scalar product with the irreducible characters in the factor system allows to count the multiplicity of irreps. This will be useful for applications in Chapter 5.

2.9 The regular representation

Let $\mathbf{G}$ be a finite group and $\omega(g_1, g_2)$ a cocycle. Consider the Hilbert space $\mathbb{C}^{|\mathbf{G}|}$ with the orthonormal basis $\{|g\rangle \mid g \in \mathbf{G}\}$.

Definition 13. *The projective representation $\{U_g^{reg} \mid g \in \mathbf{G}\}$ defined by the relation*

$$U_h|g\rangle = \omega(h,g)\ |hg\rangle\ , \tag{2.46}$$

is called the regular representation *in the factor system specified by the cocycle ω.*

It is easy to verify that the regular representation U_g^{reg} is a projective representation with cocycle ω. Writing the unitaries $\{U_h\}$ as

$$U_h = \sum_{g\in\mathbf{G}} \omega(h,g)\ |hg\rangle\langle g|\ , \tag{2.47}$$

one can immediately evaluate the characters of the regular representation:

$$\chi_{reg}(g) = \mathrm{Tr}[U_g^{reg}] = |\mathbf{G}|\ \delta_{g,e}\ . \tag{2.48}$$

The regular decomposition is reducible. Its decomposition in irreps can be easily obtained using Eq. (2.45), namely the multiplicity of the irrep μ is

$$m_\mu = \langle \chi_\mu | \chi_{reg} \rangle = d_\mu\ . \tag{2.49}$$

In other words, the regular representation contains all the irreps in the factor system specified by the cocycle ω, each of them with multiplicity $m_\mu = d_\mu$, i.e. the Clebsch-Gordan decomposition of $\{U_g^{reg}\}$ is

$$U_g^{reg} = \bigoplus_{\mu\in\mathrm{Irr}(\mathbf{G},\omega)} U_g^\mu \otimes \mathbb{1}_{d_\mu}\ , \tag{2.50}$$

where $\mathrm{Irr}(\mathbf{G},\omega)$ denotes the set of all irreps contained in the factor system with cocycle ω. The corresponding decomposition of the Hilbert space is then given by the Clebsch-Gordan TPS

$$\mathbb{C}^{|\mathbf{G}|} = \bigoplus_{\mu\in\mathrm{Irr}(\mathbf{G},\omega)} \mathcal{H}_\mu \otimes \mathbb{C}^{d_\mu}\ . \tag{2.51}$$

Counting the dimensions on both sides of the equality we also obtain the remarkable relation

$$|\mathbf{G}| = \sum_{\mu\in\mathrm{Irr}(\mathbf{G},\omega)} d_\mu^2\ . \tag{2.52}$$

Finally, taking the trace on both sides of Eq. (2.50) and keeping in mind that $\chi_{reg}(g) = |\mathbf{G}|\delta_{g,e}$, we obtain the decomposition of the Kronecker delta:

$$\delta_{g,e} = \frac{1}{|\mathbf{G}|} \sum_{\mu \in \mathrm{Irr}(\mathbf{G},\omega)} d_\mu \; \chi_\mu(g) \; . \tag{2.53}$$

The definition and the properties of the regular representation can be rather simply extended to the case of a compact Lie group $\mathbf{G}$. In this case one has to consider the Hilbert space $L^2(\mathbf{G})$, and the vectors $\{|g\rangle \mid g \in \mathbf{G}\}$, with Dirac normalization condition

$$\langle g|h\rangle = \delta(g^{-1}h) \; . \tag{2.54}$$

Here $\delta(g)$ denotes the Dirac-delta on the group, defined by the relation

$$\langle \delta | f \rangle = \int_{\mathbf{G}} \mathrm{d}g \; \delta(g) f(g) = f(e) \; , \tag{2.55}$$

for any function $f(g)$ which is regular in a neighborhood of the identity e. Again, given a cocycle ω, we can define the regular representation by the relation

$$U^{reg}_h |g\rangle = \omega(g,h) \; |hg\rangle \; . \tag{2.56}$$

In this case, the characters of the regular representation are given by $\chi_{reg}(g) = \mathrm{Tr}[U^{reg}_g] = \delta(g)$. In fact, since

$$U^{reg}_g = \int_{\mathbf{G}} \mathrm{d}h \; \omega(g,h) \; |gh\rangle\langle h| \; , \tag{2.57}$$

we have indeed

$$\begin{aligned} \langle \chi_{reg}|f\rangle &= \int_{\mathbf{G}} \mathrm{d}g \int_{\mathbf{G}} \mathrm{d}h \; f(g) \; \omega(g,h) \delta(h^{-1}gh) && (2.58) \\ &= \int_{\mathbf{G}} \mathrm{d}h \int_{\mathbf{G}} \mathrm{d}g \; f(hgh^{-1}) \; \omega(hgh^{-1},h) \delta(g) && (2.59) \\ &= f(e) = \langle \delta | f \rangle \; , && (2.60) \end{aligned}$$

whence $\chi_{reg} = \delta$.

According to Theorem 2, the regular representation of a compact group $\mathbf{G}$ can be decomposed in a discrete sum of irreps. In such a decomposition, the multiplicity m_μ of the irrep $\{U^\mu_g\}$ is given by Eq. (2.45), therefore one has

$$m_\mu = \langle \chi_\mu | \chi_{reg} \rangle = \int_{\mathbf{G}} \mathrm{d}g \; \delta(g) \; \mathrm{Tr}[U^\mu_g] = d_\mu \; . \tag{2.61}$$

Again, the regular representation has the decomposition

$$U_g^{reg} = \bigoplus_{\mu \in \mathrm{Irr}(\mathbf{G},\omega)} U_g^{\mu} \otimes \mathbb{1}_{d_\mu} \ , \tag{2.62}$$

which induces the Clebsch-Gordan TPS for the Hilbert space:

$$L^2(\mathbf{G}) = \bigoplus_{\mu \in \mathrm{Irr}(\mathbf{G},\omega)} \mathcal{H}_\mu \otimes \mathbb{C}^{d_\mu} \ , \tag{2.63}$$

as in Eq. (2.20).

Finally, taking the trace on both sides of Eq. (2.62) and keeping in mind that $\chi_{reg}(g) = \delta(g)$, we obtain the decomposition of the Dirac delta on the group:

$$\delta(g) = \sum_{\mu \in \mathrm{Irr}(\mathbf{G},\omega)} d_\mu \ \chi_\mu(g) \ . \tag{2.64}$$

2.10 Fourier analysis

It is well known that any function in the Hilbert space $L^2([0,2\pi])$ can be expanded in Fourier series, namely for any $\psi(\theta) \in L^2([0,2\pi])$ one has

$$\psi(\theta) = \sum_{k \in \mathbb{Z}} \psi_k \ e^{ik\theta} \ . \tag{2.65}$$

It is immediate to recognize in the exponentials $\{e^{ik\theta} \mid k \in \mathbb{Z}\}$ the matrix elements of the one-dimensional irreps of the Abelian group $U(1) \cong [0,2\pi]$. Remarkably, the possibility of a Fourier expansion is not an exclusive feature of the group $U(1)$: as we will see in the following it is possible to expand any function $\psi(g)$ in the Hilbert space $L^2(\mathbf{G})$ as a linear combination of irreducible matrix elements. A similar results holds of course also for finite groups, where any vector $|\psi\rangle$ in $\mathbb{C}^{|\mathbf{G}|}$ can be expanded on the orthonormal basis defined in Eq. (2.38).

Let us start with the case of finite groups. As it was shown in the previous Section, in this case the Hilbert space $\mathbb{C}^{|\mathbf{G}|}$ can be decomposed as

$$\mathbb{C}^{|\mathbf{G}|} = \bigoplus_{\mu \in \mathrm{Irr}(\mathbf{G},\omega)} \mathcal{H}_\mu \otimes \mathbb{C}^{d_\mu} \ . \tag{2.66}$$

Here we will use the irreducible matrix elements $\{u_{ij}^{\mu}(g)\}$ to construct explicitly an orthonormal basis for the invariant subspaces $\mathcal{H}_\mu \otimes \mathbb{C}^{d_\mu}$. To this purpose, define the orthonormal vectors

$$|\tilde{u}_{ij}^{\mu *}\rangle \doteq \sqrt{\frac{d_\mu}{|\mathbf{G}|}} \sum_{g \in \mathbf{G}} u_{ij}^{\mu *}(g) \ |g\rangle \ , \tag{2.67}$$

where $*$ denotes the complex conjugation. The normalization and the orthogonality are ensured by Theorem 4.

It is simple to show that the vectors $\{|\tilde{u}_{ij}^{\mu}\rangle \mid i = 1, \ldots, d_\mu\}$ (for fixed μ and j) are a basis for an irreducible subspace carrying the irrep $\{U_g^\mu\}$. One has indeed

$$U_g^{reg}\ |\tilde{u}_{ij}^{\mu*}\rangle = \sum_{k=1}^{d_\mu} u_{ki}^{\mu}(g)\ |\tilde{u}_{kj}^{\mu*}\rangle\ , \tag{2.68}$$

this relation following from the use of the cocycle relation $\omega(g,h)\omega(gh,k) = \omega(g,hk)\omega(h,k)$. Notice that the degree of freedom specified by the index j is not affected by the action of the group. Accordingly, we can make the identification

$$|u_{ij}^{\mu*}\rangle \equiv |\mu, i\rangle \otimes |j\rangle \in \mathcal{H}_\mu \otimes \mathbb{C}^{d_\mu}\ , \tag{2.69}$$

thus obtaining the desired basis for $\mathcal{H}_\mu \otimes \mathbb{C}^{d_\mu}$.

Using the fact that $\{|\tilde{u}_{ij}^{\mu*}\rangle\}$ is an orthonormal basis, we can expand any vector in the Hilbert space $\mathbb{C}^{|\mathbf{G}|}$:

Theorem 6. *Let* $\mathbf{G}$ *be a finite group. Then, for any vector* $|\psi\rangle \in \mathbb{C}^{|\mathbf{G}|}$*, one has the expansion*

$$|\psi\rangle = \bigoplus_{\mu\in\mathrm{Irr}(\mathbf{G},\omega)} \sum_{i,j=1}^{d_\mu} \langle\tilde{u}_{ij}^{\mu*}|\psi\rangle\ |\tilde{u}_{ij}^{\mu*}\rangle\ . \tag{2.70}$$

The above results can be simply extended to the case of compact groups. In this case, the matrix elements $\{u_{ij}^{\mu*}(g)\}$ provide a basis for the Hilbert space $L^2(\mathbf{G})$. It is enough to define the orthonormal vectors $\{|\tilde{u}_{ij}^{\mu*}\rangle\}$ as

$$|\tilde{u}_{ij}^{\mu*}\rangle \doteq \sqrt{d_\mu}\ \int_{g\in\mathbf{G}} \mathrm{d}g\ u_{ij}^{\mu*}(g)\ |g\rangle\ . \tag{2.71}$$

Again, it can be proved that the vectors $\{|\tilde{u}_{ij}^{\mu*}\rangle \mid i = 1, \ldots, d_\mu\}$ (for fixed μ and j) are a basis for an irreducible subspace carrying the irrep $\{U_g^\mu\}$. Accordingly, we can make the identification (2.69) to obtain a basis for $\mathcal{H}_\mu \otimes \mathbb{C}^{d_\mu}$.

The generalization of Theorem 6 in the case of compact groups is then

Theorem 7. *Let* $\mathbf{G}$ *be a compact group. Then, for any function* $\psi(g) \in L^2(\mathbf{G})$*, one has the expansion*

$$|\psi\rangle = \bigoplus_{\mu\in\mathrm{Irr}(\mathbf{G},\omega)} \sum_{i,j=1}^{d_\mu} \langle\tilde{u}_{ij}^{\mu*}|\psi\rangle\ |\tilde{u}_{ij}^{\mu*}\rangle\ . \tag{2.72}$$

In other words, for any possible cocycle ω of the compact group $\mathbf{G}$ one has a generalized Fourier analysis, in which the function $\psi(g)$ in $L^2(\mathbf{G})$ has the series expansion:

$$\psi(g) = \sum_{\mu \in \mathrm{Irr}(\mathbf{G},\omega)} \sum_{i,j=1}^{d_\mu} \psi^\mu_{ij} \; u^{\mu *}_{ij}(g) \; , \tag{2.73}$$

where

$$\psi^\mu_{ij} \doteq d_\mu \int_{\mathbf{G}} \mathrm{d}g \; u^\mu_{ij}(g) \; \psi(g) \; . \tag{2.74}$$

A generalization of the previous results to noncompact groups is possible, but since the Clebsch-Gordan decomposition of the regular representation of a noncompact group may require a direct integral of irreps, the Fourier series considered so far should be substituted with Fourier integrals. As an example, for the additive group $\mathbf{G} = \mathbb{R}$ one gets the well known Fourier transform of a function $\psi(x) \in L^2(\mathbb{R})$:

$$\psi(x) = \int_{\mathbb{R}} \mathrm{d}k \; \widetilde{\psi}(k) \; e^{-ikx} \; , \tag{2.75}$$

where

$$\widetilde{\psi}(x) = \frac{1}{2\pi} \int_{\mathbb{R}} \mathrm{d}x \; e^{ikx} \; \psi(x) \; . \tag{2.76}$$

However, the theory of generalized Fourier analysis for noncompact groups is beyond the scope of this presentation.

Chapter 3

Covariant quantum estimation

A wide class of topics in quantum estimation theory involves the presence of symmetry with respect to some group of physical transformations. The elements of the symmetry group transform the state of the system, thus inducing a group action in the space of signal parameters Θ.

The covariant quantum estimation problem regards the optimal extraction of information from families of signal states that are invariant under the action of a given symmetry group. The group action has to be considered as a special kind of encoding, in which the signal states carry the information about the parameters of a group transformation imprinted into the system. The covariant encoding requires appropriate decoding strategies, which are given by the class of *covariant POVMs*.

3.1 Invariant families of states

Consider a symmetry group $\mathbf{G}$, which can be either finite or continuous. In the case of $\mathbf{G}$ being continuous, we will assume it to be a locally compact Lie group. According to the well-known theorem by Wigner[38], the action of the group on the state space $\mathcal{S}(\mathcal{H})$ of a quantum system is given by the automorphism

$$\rho \longmapsto \mathcal{A}_g(\rho) = U_g \rho U_g^\dagger \tag{3.1}$$

where U_g can be either a unitary or an antiunitary operator. Here we will always consider the case of unitary operators. Moreover, to guarantee the group properties of the automorphism $\mathcal{A}_g$, the unitary U_g must be an element of a projective representation $\{U_g \in \mathcal{B}(\mathcal{H}) \mid g \in \mathbf{G}\}$.

The covariant estimation problem is characterized by the fact that the family of signal states $\mathcal{F}(\Theta) = \{\rho_\theta \mid \theta \in \Theta\}$ is invariant under the action of the group, namely for any element $g \in \mathbf{G}$ one has $g\mathcal{F}(\Theta) = \mathcal{F}(\Theta) \quad \forall g \in \mathbf{G}$, where

$$g\mathcal{F}(\Theta) \doteq \{\mathcal{A}_g(\rho_\theta) \mid \theta \in \Theta\} \ . \tag{3.2}$$

Since an invariant family contains the group orbit of any state $\rho_\theta \in \mathcal{F}(\Theta)$, i.e.

$$\mathcal{O}_\theta \doteq \{\mathcal{A}_g(\rho_\theta) \mid g \in \mathbf{G}\} \ , \tag{3.3}$$

the family itself can be considered as the union of a certain number of disjoint orbits, namely

$$\mathcal{F}(\theta) = \bigcup_{x \in \ X} \mathcal{O}_x \ . \tag{3.4}$$

Notice that the relation of being in the same orbit is an equivalence relation, whence each orbit can be considered as an equivalence class $x \in X$ with $X = \Theta/\mathbf{G}$. Moreover, the action of the group restricted to a given orbit is *transitive*, namely any two states in the same orbit are connected by some group transformation. Choosing a fixed state $\rho_{x,0} \in \mathcal{O}_x$, it is possible to parametrize the elements of the orbit $\mathcal{O}_x$ by points of the coset space $\mathbf{G}/\mathbf{G}_{x,0}$, where $\mathbf{G}_{x,0}$ is the *stability group* of the state $\rho_{x,0}$, defined by

$$\mathbf{G}_{x,0} \doteq \{g \in \mathbf{G} \mid \mathcal{A}_g(\rho_{x,0}) = \rho_{x,0} \ \} \tag{3.5}$$

Introducing the identification $\mathcal{O}_x \equiv \mathbf{G}/\mathbf{G}_{x,0}$ in Eq. (3.4), the space of signal parameters can be decomposed as

$$\Theta = \bigcup_{x \in X} \mathbf{G}/\mathbf{G}_{x,0} \ , \tag{3.6}$$

and any signal parameter $\theta \in \Theta$ can be represented as a couple $\theta = (x, y)$, where $x \in X$ labels different orbits, while $y \in \mathbf{G}/\mathbf{G}_{0,x}$ labels different elements of the same orbit.

The action of the group on states naturally induces an action on the space of signal parameters, which we simply indicate by $g : \theta \longmapsto g\theta$. The parameter $g\theta$ is defined by the relation $\rho_{g\theta} = \mathcal{A}_g(\rho_\theta)$. In the parametrization $\theta = (x, y)$, one has $g\theta = (x, gy)$, where the gy denotes the action of the group on the equivalence class y.

The most common case considered in the literature is that of a transitive space of signal parameters [1, 2], which is the case where the family of signal states is a single group orbit:

$$\Theta = \mathbf{G}/\mathbf{G}_0 \ . \tag{3.7}$$

An even more particular case, but still very relevant, is that of *trivial stability group*, where $\mathbf{G}_0 = \{e\}$ (e denotes the identity element). In this case, the signal states are in one-to-one correspondence with the elements of the group **G**, namely

$$\Theta = \mathbf{G} \ . \tag{3.8}$$

This is the case, for example, of the coherent states of the radiation field $\{|\alpha\rangle \mid \alpha \in \mathbb{C}\}$, which are in one to one correspondence with translations in the complex plane $\mathbb{C}$.

3.2 Covariant POVMs

Consider a measurement devised to estimate the value of a parameter $\omega \in \Omega$, in the case where the outcome space supports a group action $g : \omega \mapsto g\omega$. In the same way as it was done in the previous Section, the outcome space can be decomposed into orbits, and, similarly to Eq. (3.6), we have

$$\Omega = \bigcup_{z \in Z} \mathbf{G}/\mathbf{H}_{0,z} \ , \tag{3.9}$$

where z labels different orbits and $\mathbf{H}_{0,z}$ is the stability group of the element $\omega_{z,0}$ chosen in the $z-$th orbit.

Definition 14 (Covariant POVMs). *A POVM $P(\mathrm{d}\omega)$ is* covariant *iff for any state $\rho \in \mathcal{S}(\mathcal{H})$, the probability distribution $p(B|\rho) = \mathrm{Tr}[P(B)\rho]$ is group invariant, namely*

$$p(B|\rho) = p(gB|\mathcal{A}_g(\rho)) \qquad \forall B \in \sigma(\Omega) \ , \tag{3.10}$$

where $gB \doteq \{g\omega \mid \omega \in B\}$ and $\mathcal{A}_g(\rho) = U_g \rho U_g^\dagger$.

The structure of covariant POVMs has been studied in detail by Davies and Holevo in the case of transitive group action in the outcome space [39, 2], namely if Ω is a single group orbit

$$\Omega = \mathbf{G}/\mathbf{H}_0 \ , \tag{3.11}$$

and the possible outcomes are equivalence classes, namely $\Omega \ni \omega = [g(\omega)]$ for some $g(\omega) \in \mathbf{G}$. In this case, covariant POVMs have a particularly simple form, expressed by the following

Theorem 8. *Let **G** be a locally compact unimodular group, and $\{U_g\}$ a projective representation of **G** in the Hilbert space $\mathcal{H}$. Let $\mathsf{H}_0 \subseteq \mathbf{G}$ be a compact subgroup. Then, a POVM $P(\mathrm{d}\omega)$ with outcome space $\Omega = \mathbf{G}/\mathbf{H}_0$ is covariant if and only if it has the form*

$$P(\mathrm{d}\omega) = U_{g(\omega)} \ \Xi \ U_{g(\omega)}^\dagger \ \nu(\mathrm{d}\omega) \ , \tag{3.12}$$

where $g(\omega) \in \mathbf{G}$ is any representative element of the equivalence class $\omega \in \Omega$, $\nu(\mathrm{d}\omega)$ is the group invariant measure over ω, and Ξ is an operator satisfying the properties

$$\Xi \geq 0 \tag{3.13}$$

$$[\Xi, U_h] = 0 \qquad \forall h \in \mathbf{H}_0 \tag{3.14}$$

$$\int_{\mathbf{G}} \mathrm{d}g \; U_g \Xi U_g^\dagger = \mathbb{1} \ , \tag{3.15}$$

$\mathrm{d}g$ *being the invariant Haar measure over the group* $\mathbf{G}$.

Under the hypotheses of the above theorem, any covariant POVM with outcome space $\Omega = \mathbf{G}/\mathbf{H}_0$ is in one-to-one correspondence with a single operator Ξ, called the *seed* of the POVM. This simple structure becomes very useful in the solution of optimization problems.

For group representations that have a discrete Clebsch-Gordan series $U_g = \bigoplus_{\mu \in \mathsf{S}} U_g^\mu \otimes \mathbb{1}_{m_\mu}$, exploiting the formulas (2.32) and (2.34) for the group average of an operator, the normalization condition (3.15) can be put in a simple form:

Proposition 3. *Let ξ be an operator on $\mathcal{H}$, and decompose $\mathcal{H}$ as*

$$\mathcal{H} = \bigoplus_{\mu \in \mathsf{S}} \mathcal{H}_\mu \otimes \mathbb{C}^{m_\mu} \ . \tag{3.16}$$

Then, the normalization condition (3.15) *is equivalent to*

$$\mathrm{Tr}_{\mathcal{H}_\mu}[\Xi] = d_\mu \ \mathbb{1}_{m_\mu} \ . \tag{3.17}$$

Theorem 8 can be generalized to the case of non transitive group action, in which the outcome space is the union of a certain number of group orbits. In this case, a covariant POVM is in one to one correspondence with an operator-valued function over the set of different orbits.

Theorem 9. *Let $P(\mathrm{d}\omega)$ be a POVM with outcome space $\Omega = \bigcup_{z \in Z} \mathbf{G}/\mathbf{H}_{0,z}$, where $\mathbf{G}$ is a locally compact unimodular group and $\mathbf{H}_{0,z} \subseteq \mathbf{G}$ are compact subgroups. Parametrize the elements of Ω as $\omega = (z,t)$ where $z \in Z$ labels the orbit and $t \in \mathbf{G}/\mathbf{H}_{0,z}$ labels the point of the orbit. The POVM $P(\mathrm{d}\omega)$ is covariant, if an only if it has has the general form*

$$P(\mathrm{d}\omega) = U_{g(\omega)} \ A(z) \ U_{g(\omega)}^\dagger \ \alpha(\mathrm{d}z)\nu_z(\mathrm{d}t) \tag{3.18}$$

where $g(\omega) \in \mathbf{G}$ is any representative element of the equivalence class $t \in \mathbf{G}/\mathbf{H}_{0,z}$, $\alpha(\mathrm{d}z)$ is a scalar measure over Z, $\nu_z(\mathrm{d}t)$ is the group invariant *measure over the $z-$th orbit, and $A(z)$ is an operator-valued density which is uniquely defined $\alpha-$almost everywhere and satisfies the properties*

$$A(z) \geq 0 \qquad \alpha - a.e. \tag{3.19}$$

$$[A_z, U_h] = 0 \qquad \alpha - a.e. \ , \forall h \in H_{0,z} \tag{3.20}$$

$$\int_{\mathbf{G}} \mathrm{d}g \int_Z \alpha(\mathrm{d}z) \ U_g \ A(z) \ U_g^\dagger = 1\!\!1 \ . \tag{3.21}$$

Sketch of proof. We outline the proof for POVMs in finite dimensional Hilbert spaces, nevertheless the result holds also in the infinite dimensional case. Consider the scalar measure over Ω, defined by $\mu(B) \doteq \mathrm{Tr}[P(B)]$. Such a measure dominates $P(\mathrm{d}\omega)$, i.e. $P(B) \leq \mu(B) 1\!\!1 \quad \forall B \in \sigma(\Omega)$. This implies that $P(\mathrm{d}\omega)$ admits a density with respect to $\mu(\mathrm{d}\omega)$, namely there exists an operator-valued function $M(\omega)$ such that $P(B) = \int_B \mu(\mathrm{d}\omega) \ M(\omega)$. Moreover, the measure $\mu(\mathrm{d}\omega)$ is invariant. In fact, $\mu(B) = d\mathrm{Tr}[\rho_c M(B)]$, with $\rho_c = 1\!\!1/d$ being the maximally chaotic state, and

$$\begin{aligned} \mu(gB) &= d\mathrm{Tr}[\rho_c \ P(gB)] \\ &= d\mathrm{Tr}[\mathcal{A}_{g^{-1}}(\rho_c) \ P(B)] \\ &= d\mathrm{Tr}[\rho_c \ P(B)] = \mu(B) \qquad \forall B \in \sigma(\Omega) \ . \end{aligned} \tag{3.22}$$

Since $\mu(\mathrm{d}\omega)$ is an invariant measure over $\Omega = \bigcup_{z \in Z} \mathbf{G}/\mathbf{H}_{0,z}$, where any element ω is parametrized as $\omega = (z,t)$, $\mu(\mathrm{d}\omega)$ must have the form $\mu(\mathrm{d}\omega) = \alpha(\mathrm{d}z)\nu_z(\mathrm{d}t)$, where $\alpha(\mathrm{d}z)$ is a suitable measure over Z, and $\nu_z(\mathrm{d}t)$ is the invariant measure over $\mathbf{G}/\mathbf{H}_{0,z}$. Finally, the covariance condition (3.10) requires $U_g^\dagger M(g\omega) U_g = M(\omega)$ for any $g \in \mathbf{G}$, and, equivalently, $M(z, gt) = U_g M(z,t) U_g^\dagger$. This condition implies the, for fixed z, the operator $M(z,t)$ has the form $M(z,t) = U_{g(t)} \ A(z) \ U_{g(t)}^\dagger$ (this can be proved in the same way as in the original theorem by Holevo, since for fixed z the group action becomes transitive). ■

In the case of nontransitive group actions, the normalization of the POVM, given by Eq. (3.21) can be turned a simple condition:

Proposition 4. *Let $A : Z \in \mathcal{B}(\mathcal{H}) \ , z \longmapsto A(z)$ be an operator-valued function, and $\mathcal{H} = \bigoplus_{\mu \in \mathsf{S}} \mathcal{H}_\mu \otimes \mathbb{C}^{m_\mu}$ the Clebsch-Gordan TPS induced by the action of the representation $\{U_g\}$. Then, the normalization condition* (3.21) *is equivalent to*

$$\int_Z \alpha(\mathrm{d}z) \ \mathrm{Tr}_{\mathcal{H}_\mu}[A(z)] = d_\mu \ 1\!\!1_{m_\mu} \ . \tag{3.23}$$

3.3 Covariant POVMs and optimization

In the presence of group symmetry, covariant POVMs play a major role in the search of the optimal estimation strategy. This is a common feature, holding for anyone of the possible definitions of optimality examined in Sec. 1.2.1. We will illustrate in the following some results about the optimality of covariant POVMs for the minimization of a cost function (Bayesian and minimax approach), and for the maximization of the mutual information.

In the minimization of a cost function we will refer for simplicity to the case of *state estimation*, where one produces an estimate $\hat{\theta}$ of the signal parameter θ encoded in ρ_θ. Accordingly, the cost function will be of the form $c(\hat{\theta},\theta)$. Since we are considering invariant families of signal states, it is natural to require the cost function $c(\hat{\theta},\theta)$ to be invariant, namely

$$c(g\hat{\theta}, g\theta) = c(\hat{\theta},\theta) \qquad \forall g \in \mathbf{G},\ \forall \hat{\theta}, \theta \in \Theta \ . \tag{3.24}$$

For example, if the signal states are coherent states of the radiation field $\mathcal{F}(\mathbb{C}) = \{|\alpha\rangle \mid \alpha \in \mathbb{C}\}$, it is natural consider cost functions of the form $c(\hat{\alpha},\alpha) = f(|\hat{\alpha}-\alpha|)$, f being an increasing function of the distance in phase space. Any such function is invariant under complex shifts, namely $c(\hat{\alpha}-\beta, \alpha-\beta) = c(\hat{\alpha},\alpha) \quad \forall \hat{\alpha}, \alpha, \beta \in \mathbb{C}$. The invariance of the cost function means that there are no privileged points in the orbits that parametrize the family of signal states. .

Always assuming the cost function to be invariant, in the following we will discuss the minimization of a cost function in the Bayesian and minimax approach. Finally, we will conclude the Chapter by considering the maximization of the mutual information in the presence of group symmetry.

3.3.1 Bayesian approach

Consider the case of a finite or compact Lie group $\mathbf{G}$, and consider the Bayesian optimization problem with cost function $c(\hat{\theta},\theta)$, and with a group invariant probability distribution over the signal parameters, namely $\nu(gB) = \nu(B)$ for any $B \in \sigma(\Theta)$. From a subjective point of view, such a uniform prior distribution represents a complete lack of knowledge about the parameters of the group transformation imprinted into the system. In this case, we have the following

Proposition 5. *For any generic POVM $Q(\mathrm{d}\theta)$ there is a covariant POVM $P(\theta)$ with the same Bayes expected cost, i.e. $\langle c\rangle_M = \langle c\rangle_N$.*

Proof. A straightforward generalization of the theorem given in Ref.[2] for transitive group actions. Consider the invariant measure on the group $\mathbf{G}$, denoted as $\mathrm{d}g$, and normalized as $\int_{\mathbf{G}} \mathrm{d}g = 1$. Then, it is easy the check that the POVM

$$P(\mathrm{d}\theta) \doteq \int_{\mathbf{G}} \mathrm{d}g \; U_g^\dagger \; Q(\mathrm{d}(g\theta)) \; U_g \tag{3.25}$$

is covariant, and has the same average cost as $Q(\mathrm{d}\theta)$. ■

As a consequence of the above proposition, in the minimization of the Bayes cost there is no need of considering all possible POVMs, since the minimization can be restricted without loss of generality to the set of covariant POVMs.

It is worth stressing that this result holds only for finite and compact groups: If the group is noncompact it is not possible to define a uniform probability distribution on it (a uniform measure on a noncompact group cannot be normalized).

3.3.2 Minimax approach

Also in the minimax approach, the optimality of covariant measurements can be easily established for compact groups.

Proposition 6. *For any generic POVM $Q(\theta)$ there is a covariant POVM $P(\mathrm{d}\theta)$ such that $c^{(P)}_{\max} \leq c^{(Q)}_{\max}$.*

Proof. Consider the covariant POVM defined by Eq. (3.25). Its worst case cost is

$$\begin{aligned} c^{(P)}_{\max} &= \max_{\theta\in\Theta} \int_{\mathbf{G}} \mathrm{d}g \int_{\Theta} \; c(\hat\theta,\theta)\mathrm{Tr}[\rho_{g\theta} Q(\mathrm{d}(g\hat\theta))] \\ &= \max_{\theta\in\Theta} \int_{\mathbf{G}} \mathrm{d}g \int_{\Theta} \; c(\hat\theta,g\theta)\mathrm{Tr}[\rho_{g\theta} Q(\mathrm{d}\hat\theta)] \\ &\geq \int_{\mathbf{G}} \mathrm{d}g \;\; \max_{\theta\in\Theta} \overline{c}^{(Q)}(g\theta) \\ &= c^{(Q)}_{\max} \; . \end{aligned} \tag{3.26}$$

In the minimax approach the optimality of covariant measurements holds also for noncompact groups under suitable assumptions. The proof becomes quite technical, and, to the best of our knowledge, it is given only in Ref.[40].

Remarkably, for finite and compact groups, and for transitive group actions, the minimax and the Bayesian approach do coincide, namely both approaches lead to the same optimal covariant POVM. In fact, in that particular case one has the following

Proposition 7. *[2] If the group action is transitive on the parameter space Θ, i.e. $\Theta = \mathbf{G}/\mathbf{G}_0$ for some subgroup $\mathbf{G}_0$, then for a covariant POVM $P(\mathrm{d}\theta)$ the average cost $\bar{c}(\theta)$, as defined in Eq.* (1.13)*, does not depend on θ. In particular,*

$$\bar{c}(\theta) \equiv c_{\max} \equiv \langle c \rangle \qquad \forall \theta \in \Theta \ . \tag{3.27}$$

As a consequence, a covariant POVM that minimizes the Bayes expected cost also minimizes the worst case cost, and viceversa. However, we stress that this result does not hold if the group action is nontransitive. In fact, any POVM with outcome space Ω can be regarded as a POVM which is covariant with respect to the nontransitive action of the trivial group $\mathbf{G} = \{e\}$, and it is well known that in general the Bayesian approach and the minimax do not coincide (see the example in Subsection 1.2.3).

3.3.3 Mutual information

The maximization of the mutual information (1.20) is a completely different problem with respect to the minimization of a given cost function. First of all, the mutual information is a nonlinear functional of the POVM, this making the optimization much harder. Moreover, the space of outcomes Ω has not necessarily to coincide with the space of signal parameters Θ, and, therefore, not only the POVM, but also the choice of its space of outcomes has to be optimized. Despite to these remarkable differences, in the presence of group symmetry, the optimality of covariant POVMs also holds in the context of mutual information. In fact, one has the following result

Proposition 8. *Let $\mathbf{G}$ be a compact group, $\mathcal{F}(\Theta)$ an invariant family of signal states, and $Q(\mathrm{d}x)$ an arbitrary POVM with outcome space X. Consider the parameter $\omega \doteq (x, g)$, belonging to the space $\Omega = X \times \mathbf{G}$. Then the covariant POVM*

$$P(\mathrm{d}\omega) = U_g \ Q(\mathrm{d}x) \ U_g^\dagger \ \mathrm{d}g \ , \tag{3.28}$$

achieves the same value of the mutual information as $Q(\mathrm{d}x)$.

For discrete families of signals, discrete outcome spaces, and finite groups, this is a classic result by Davies[31]. The extension of the standard proof to continuous families of states only requires the compactness of the group.

Chapter 4

Maximum likelihood estimation strategies

The maximum likelihood approach to the estimation of group parameters is a prototype of optimization in which the optimal measurements, the optimal signal states, and their relations with other information theoretical quantities can be derived in a simple and general fashion. These results provide a deep insight into the general structure underlying optimal quantum estimation strategies, in particular by evidencing the role of entanglement between representation spaces and multiplicity spaces in the tensor product structure induced by the group action.

4.1 The maximum likelihood criterion

In this chapter we will consider the simplest case of optimal state estimation in the presence of symmetry. The family of signal states will be the orbit of a pure input state

$$\rho_0 = |\Psi\rangle\langle\Psi| \ , \qquad |\Psi\rangle \in \mathcal{H} \ , \tag{4.1}$$

under the action of a projective representation $\{U_g \mid g \in \mathbf{G}\}$ of the locally compact unimodular group $\mathbf{G}$. For clarity, we will first present the results in the case where the signal states are in one-to-one correspondence with the elements of the group, namely in the case where the stability group $\mathbf{H}_0 = \{g \in \mathbf{G} \mid U_g \rho_0 U_g^\dagger = \rho_0 \ \}$ contains only the identity element e. The extension to the case of nontrivial stability group is rather simple and does not add any new feature to the results presented in this Chapter.

In the following, the optimality will be defined in the *maximum likelihood criterion*[26, 1, 2], which corresponds to the maximization of the probability (probability density in the case of continuous groups) that the estimated value of the unknown parameter coincides with its true value. This case is the simplest example of minimization of the average value of a cost function, with the cost function being a Dirac-delta

$$c(\hat{g}, g) = -\delta(\hat{g}^{-1} g) \tag{4.2}$$

(Kronecker-delta for finite groups). Notice that the delta function is an invariant cost function, i.e. $c(h\hat{g}, hg) = c(\hat{g}, g) \quad \forall h, \hat{g}, g \in \mathbf{G}$, therefore the optimal POVM can be assumed without loss of generality to be covariant (see Sec. 3.3). Moreover, for finite and compact groups, the minimax and the Bayesian approach with uniform prior lead to the same optimal POVM. For non-compact groups, instead, the Bayesian approach with uniform prior is no longer possible: the covariant POVMs presented in the following are optimal in the minimax sense.

For finite groups the maximum likelihood is in some sense the most natural criterion. In fact, if we are trying to decide among a finite set of alternatives, of course we would like to make the correct decision with maximum probability of success[26]. On the other hand, in the case of continuous groups, the maximization of the likelihood might appear as a weak criterion of optimality, since it takes into account only a point in the probability distribution resulting from the estimation strategy. Nevertheless, there is a wide class of examples in which the maximum likelihood criterion singles out a POVM which is optimal also for other apparently more meaningful criteria. A rigorous explanation of the reason of this feature will be given in Chapter 6 by introducing a class of cost functions that lead to the same optimal POVM. This will give a partial account of the optimality of the maximum likelihood POVMs, but does not cover all the lucky coincidences in which such POVMs are optimal for different criteria. An intuitive explanation of this “universality” of the maximum likelihood approach is that, when the estimation becomes asymptotically precise, i.e. in cases in which many copies of the same system are available for estimation or, typically, when the average energy of the signal states is asymptotically large, the probability distribution resulting from the estimation strategy tends to become Gaussian, and, therefore the maximization of the peak in the probability distribution becomes equivalent to the minimization of the variance. Since in general for an unbiased and asymptotically precise estimation the average cost depends only on the variance, the maximization of the likelihood is intuitively linked with the minimization of different cost functions[1].

[1] More precisely, we recall that an estimation is *unbiased* if the average value of the

4.2 Optimal POVMs

Here we derive for any pure input state $|\Psi\rangle \in \mathcal{H}$ the measurement that maximizes the probability (probability density for continuous groups) of estimating the true value of the unknown group parameter $g \in \mathbf{G}$. To this purpose we work in the decomposition (2.20) of Chapter 1, introducing the tensor product structure

$$\mathcal{H} = \bigoplus_{\mu \in \mathsf{S}} \mathcal{H}_\mu \otimes \mathbb{C}^{m_\mu} \ . \tag{4.6}$$

According to Theorem 2, this decomposition is always possible for finite and compact groups. For noncompact groups, the existence of a discrete Clebsch-Gordan decomposition has to be considered as an additional requirement for the representation $\{U_g\}$. In the following, we will always refer to such a case when deriving the general results about maximum likelihood POVMs. The generalization to the case of a direct integral of irreps will be given in a the specific example of Section 4.4.

Exploiting the Clebsch-Gordan TPS, the input state $|\Psi\rangle$ can be written as

$$|\Psi\rangle = \bigoplus_{\mu \in \mathsf{S}} c_\mu |\Psi_\mu\rangle\!\rangle \ , \tag{4.7}$$

where c_μ are amplitudes, and $|\Psi_\mu\rangle\!\rangle$ are bipartite states[2] in the tensor product $\mathcal{H}_\mu \otimes \mathbb{C}^{m_\mu}$. According to Theorem 8, the covariant POVM representing the

estimated parameters is equal to the true value, i.e.

$$\bar{\omega}(\omega) \doteq \int_\Omega \mathrm{d}\omega \ \hat{\omega} \ p(\mathrm{d}\hat{\omega}|\omega) = \omega \ . \tag{4.3}$$

For a cost function with nonzero Hessian $f(\hat{\omega}, \omega)$, $H_{ij} \doteq \left(\frac{\partial^2 f}{\partial \omega_i \partial \omega_j}\right) \neq 0$, and for a concentrated probability distribution, one has the average cost

$$\begin{aligned} \bar{f}(\omega) &\approx f(\omega) + \frac{1}{2} \sum_{i,j} \left[\frac{\partial^2 f}{\partial \omega_i \partial \omega_j}\right]_\omega \overline{(\hat{\omega}_i - \omega_i)(\hat{\omega}_j - \omega_j)} && (4.4) \\ &= f(\omega) + \frac{1}{2} \ \mathrm{Tr}[H(\omega) C(\omega)] \ , && (4.5) \end{aligned}$$

i.e. the average cost depends only on the trace of the covariance matrix $C(\omega)$ with the Hessian of the cost function. On the other hand, for multivariate Gaussian distributions, the maximization of the likelihood is equivalent to the minimization of the product of the variances, i.e. to the determinant of the covariance matrix. Since in asymptotic cases the probability distribution of the optimal estimation tends to a Gaussian, the link between the maximization of the likelihood and the minimization of different cost functions can be easily understood.

[2] Here and all throughout the presentation we use the expression "bipartite state" for

quantum measurement has the form

$$P(\mathrm{d}g) = U_g \Xi U_g^\dagger \ \mathrm{d}g \ , \tag{4.8}$$

where $\mathrm{d}g$ is the uniform Haar measure and the seed Ξ is a nonnegative operator satisfying the normalization constraint $\mathrm{Tr}_{\mathcal{H}_\mu}[\xi] = d_\mu \ \mathbb{1}_{m_\mu}$, as stated in Proposition 3. Note that, due to covariance, the probability density of correct estimation has the same value for any group element:

$$p(g|g) = \mathrm{Tr}\left[\left(U_g|\Psi\rangle\langle\Psi|U_g^\dagger\right)\left(U_g\xi U_g^\dagger\right)\right] = \langle\Psi|\xi|\Psi\rangle \ . \tag{4.9}$$

Once the state $|\Psi\rangle$ has been expressed as in Eq. (4.7), it is suitable to write each bipartite state $|\Psi_\mu\rangle\!\rangle$ in the Schmidt form:

$$|\Psi_\mu\rangle\!\rangle = \sum_{m=1}^{r_\mu} \sqrt{\lambda_m^\mu} \ |\psi_m^\mu\rangle|\phi_m^\mu\rangle \ , \tag{4.10}$$

where $r_\mu \le k_\mu = \min\{d_\mu, m_\mu\}$ is the Schmidt number, and $\lambda_m^\mu > 0 \quad \forall m = 1, \dots, r_\mu$. We can now define the projection

$$P_\Psi = \bigoplus_{\mu \in \mathsf{S}} \ \sum_{m=1}^{r_\mu} \mathbb{1}_{d_\mu} \otimes |\phi_m^\mu\rangle\langle\phi_m^\mu| \ . \tag{4.11}$$

It projects onto the subspace $\mathcal{H}_\Psi$ spanned by the orbit of the input state, i.e. the smallest invariant subspace containing the input state. Clearly, the probability distribution of the outcomes of a covariant measurement $P(\mathrm{d}\hat{g}) = U_{\hat{g}} \ \Xi \ U_{\hat{g}}^\dagger \ \mathrm{d}\hat{g}$ performed on any state in the orbit depends only on the projection $P_\Psi \ \Xi \ P_\Psi$. Therefore, to specify an optimal covariant POVM for the state $|\Psi\rangle$, we need only to specify the operator $P_\Psi \Xi P_\Psi$. All covariant POVM's corresponding to the same operator will be equally optimal.

Theorem 10 (optimal POVM). *For a pure input state $|\Psi\rangle$, the optimal covariant POVM in the maximum likelihood approach is given by*

$$P_\Psi \Xi P_\Psi = |\eta\rangle\langle\eta| \ , \tag{4.12}$$

where

$$|\eta\rangle = \bigoplus_{\mu \in \mathsf{S}} \ \sqrt{d_\mu} e^{i \arg(c_\mu)} \ \sum_{m=1}^{r_\mu} |\psi_m^\mu\rangle|\phi_m^\mu\rangle \ . \tag{4.13}$$

any quantum state on the tensor product $\mathcal{H}_A \otimes \mathcal{H}_B$ of two Hilbert spaces $\mathcal{H}_A$ and $\mathcal{H}_B$. Notice that this notion is not necessarily related to the presence of two spatially separated subsystems, but only to the presence of two independent degrees of freedom, represented by the commuting algebras $\mathcal{B}(\mathcal{H}_A)$ and $\mathcal{B}(\mathcal{H}_B)$ embedded in the algebra $\mathcal{B}(\mathcal{H}_A \otimes \mathcal{H}_B)$.

$\arg(c_\mu)$ *denoting the argument of the complex number* c_μ, *i.e.* $c_\mu = |c_\mu|\, e^{i\arg(c_\mu)}$. *The value of the likelihood for the optimal POVM is*

$$p^{\text{opt}}(g|g) = \left(\sum_{\mu\in\mathsf{S}} |c_\mu| \sum_{m=1}^{r_\mu} \sqrt{\lambda_m^\mu d_\mu}\right)^2 \qquad \forall g \in \mathbf{G}\ . \tag{4.14}$$

Proof. Using Schwartz inequality, the likelihood can be bounded as follows:

$$\begin{aligned} p(g|g) &= \langle\Psi|\Xi|\Psi\rangle \\ &\leq \sum_{\mu,\nu} \ |c_\mu c_\nu| \ |\langle\!\langle\Psi_\mu|\Xi|\Psi_\nu\rangle\!\rangle| \\ &\leq \left(\sum_\mu |c_\mu| \ \sqrt{\langle\!\langle\Psi_\mu|\Xi|\Psi_\mu\rangle\!\rangle}\right)^2 \ . \end{aligned}$$

Moreover, exploiting the Schmidt form (4.10) and applying a second Schwartz inequality, we obtain

$$\begin{aligned} \langle\!\langle\Psi_\mu|\Xi|\Psi_\mu\rangle\!\rangle &= \sum_{m,n=1}^{r_\mu} \ \sqrt{\lambda_m^\mu \lambda_n^\mu}\ \langle\psi_m^\mu|\langle\phi_m^\mu|\ \Xi\ |\psi_n^\mu\rangle|\phi_n^\mu\rangle \\ &\leq \left(\sum_{m=1}^{r_\mu} \ \sqrt{\lambda_m^\mu \ \ \langle\psi_m^\mu|\langle\phi_m^\mu|\ \Xi\ |\psi_m^\mu\rangle|\phi_m^\mu\rangle}\right)^2 \ . \end{aligned}$$

Finally, the positivity of Ξ implies

$$\begin{aligned} \langle\psi_m^\mu|\langle\phi_m^\mu|\ \Xi\ |\psi_m^\mu\rangle|\phi_m^\mu\rangle &\leq \langle\phi_m^\mu|\ \mathrm{Tr}_{\mathcal{H}_\mu}[\Xi]\ |\phi_m^\mu\rangle \\ &= d_\mu \ , \end{aligned}$$

where the last equality is due to the normalization condition (3.17). By putting together these inequalities, we obtain the bound

$$p(g|g) \leq \left(\sum_{\mu\in\mathsf{S}} |c_\mu| \sum_{m=1}^{r_\mu} \ \sqrt{\lambda_m^\mu d_\mu}\right)^2 \equiv p^{\text{opt}}(g|g)\ ,$$

holding for any possible POVM. It is immediate to see that the covariant POVM given by Eqs. (4.12) and (4.13) achieves the bound, hence it is optimal. ■

Remark 1: *Uniqueness of the optimal POVM.*
Since the Theorem specifies the optimal POVM only in the subspace $\mathcal{H}_\Psi$

spanned by the orbit of the input state, it follows that the optimal POVM is unique if and only if the orbit spans the whole Hilbert space[3]. If it is not the case, one can arbitrarily complete the POVM given by (4.12) to the entire Hilbert space, without affecting the probability distribution resulting from the input state $|\Psi\rangle$.

Remark 2: *Square-root-measurements and maximum likelihood.* According with Theorem 10, the optimal POVMs in the maximum likelihood approach coincide with the so-called "square-root-measurement", introduced by Hausladen and Wootters [41]. For a discrete set of signal states $\{\rho_i\}$ with probabilities $\{p_i\}$ the square-root-measurement is indeed given by the POVM $P_i^{sq} = F^{-1/2} p_i \rho_i F^{-1/2}$, where $F = \sum_i \; p_i \rho_i$. When the set of states is a group orbit we have

$$P^{sq}(\mathrm{d}g) = F^{-1/2} \left(U_g |\Psi\rangle\langle\Psi| U_g^\dagger \right) F^{-1/2} \ \mathrm{d}g \ , \tag{4.15}$$

where $F = \int \mathrm{d}g \ U_g |\Psi\rangle\langle\Psi| U_g^\dagger$. Using the formula for the group average (2.32), one can easily compute $F = \bigoplus_\mu |c_\mu|^2 \sum_m \frac{\lambda_m^\mu}{d_\mu} \mathbb{1}_\mu \otimes |\phi_m^\mu\rangle\langle\phi_m^\mu|$, thus obtaining

$$F^{-1/2} U_g |\Psi\rangle = U_g F^{-1/2} |\Psi\rangle = U_g |\eta\rangle \ , \tag{4.16}$$

with $|\eta\rangle$ as in Theorem 10. Therefore, one has $P^{sq}(\mathrm{d}g) = U_g |\eta\rangle\langle\eta| U_g^\dagger \ \mathrm{d}g$, namely the square-root measurement-coincides with the maximum likelihood POVM.

Remark 3: *Signal states with nontrivial stability group.* The result Theorem 10 can be easily generalized to the case where the input state $\rho_0 = |\Psi\rangle\langle\Psi|$ is invariant under the action of a nontrivial stability group $\mathbf{H}_0$, i.e. $U_g \rho_0 U_g^\dagger = \rho_0 \quad \forall g \in \mathbf{H}_0$. In this case, the signal states are in one-to-one correspondence with the coset space $\Theta = \mathbf{G}/\mathbf{H}_0$, and the seed Ξ must satisfy the commutation relation of Eq. (3.14). Consider a seed Ξ which is optimal according to Theorem 10, then, if the stability group $\mathbf{H}_0$ is compact, one can define the averaged seed

$$\widetilde{\Xi} \doteq \int_{\mathbf{H}_0} \mu(\mathrm{d}h) \ U_h \Xi U_h^\dagger \ , \tag{4.17}$$

where $\mu(\mathrm{d}h)$ is the invariant Haar measure over $\mathbf{H}_0$, with normalization $\int_{\mathbf{H}_0} \mu(\mathrm{d}h) = 1$. The operator $\widetilde{\Xi}$ is obviously nonnegative and satisfies the

[3] Alternatively, one could use the expression "whole Hilbert space" for $\mathcal{H}_\Psi$ itself, i.e. for the span of the signal states, with the effect that the optimal POVM is always unique. This point of view is very reasonable since, for a given the family of states, this is the only Hilbert space relevant for the estimation. However, especially in view of the optimization of the choice of the signal states, in the following we will use expressions such as "whole Hilbert space" and "entire Hilbert space" to denote the Hilbert space of the physical system under attention.

normalization constraint (3.17). Moreover, it is easy to check that the covariant POVM $P(\mathrm{d}\theta) = U_{g(\theta)}\widetilde{\Xi}U^\dagger_{g(\theta)}\ \mathrm{d}\theta$ has the maximum likelihood

$$p(\theta|\theta) = \left(\sum_{\mu\in\mathsf{S}} |c_\mu| \sum_{m=1}^{r_\mu} \sqrt{\lambda^\mu_m d_\mu}\right)^2 \qquad \forall\theta\in\Theta\ , \tag{4.18}$$

with c_μ and λ^μ_m as in Theorem 10.

Remark 4: *Maximum likelihood for finite groups.*
In the case of finite groups, the maximum likelihood approach is equivalent to the minimization of the error probability in the discrimination of the signal states $\{\rho_g = U_g|\Psi\rangle\langle\Psi|U_g^\dagger \mid g\in\mathbf{G}\}$. The optimal POVM is then given by

$$P(g) = \frac{1}{|\mathbf{G}|} U_g \Xi U_g^\dagger\ , \tag{4.19}$$

with Ξ as in Theorem 10. The probability of successful discrimination is then

$$p(g|g) = \frac{1}{|\mathbf{G}|}\left(\sum_{\mu\in\mathsf{S}} |c_\mu| \sum_{m=1}^{r_\mu} \sqrt{\lambda^\mu_m d_\mu}\right)^2 \qquad \forall g\in\mathbf{G}\ . \tag{4.20}$$

An interesting example of maximum likelihood POVM for a finite group can be found in Ref.[42], where the estimation of an unknown permutation of distinguishable particles is considered.

4.3 Optimal input states

While in the previous Section we assumed the input state to be given, and we were interested in estimating the states in its orbit, here we focus our attention on the problem of estimating the unknown symmetry transformation U_g. From this point of view, it is important to determine which are the states in the Hilbert space that allow to maximize the probability of correct estimation.

We will first show that the dimension of the subspace spanned by the orbit of the input state is always an upper bound for the likelihood, and also that this bound can be always achieved by using suitable input states. Then, the optimal input states will be the ones that maximize the dimension of the subspace spanned by the orbit.

Lemma 3. *Let* $|\Psi\rangle = \bigoplus_{\mu\in\mathsf{S}}\ c_\mu|\Psi_\mu\rangle\rangle$ *be an input state, and*

$$d_\Psi \doteq \dim \mathrm{Span}\{U_g|\Psi\rangle \mid g\in\mathbf{G}\} \tag{4.21}$$

the dimension of the subspace spanned by its orbit. Then

$$d_\Psi = \sum_{\mu \in \mathsf{S}} d_\mu r_\mu \ , \tag{4.22}$$

where r_μ is the Schmidt number of the bipartite state $|\Psi_\mu\rangle\!\rangle$ (we define $r_\mu = 0$ if $|c_\mu| = 0$).

Proof. The subspace spanned by the orbit is the support of the frame operator

$$F = \int_{\mathbf{G}} \mathrm{d}g \; U_g |\Psi\rangle\langle\Psi| U_g^\dagger = \bigoplus_{\mu \in \mathsf{S}} |c_\mu|^2 \mathbb{1}_{d_\mu} \otimes \frac{\mathrm{Tr}_{\mathcal{H}_\mu}[|\Psi_\mu\rangle\!\rangle\langle\!\langle\Psi_\mu|]}{d_\mu} \ ,$$

the r.h.s. coming from Eq. (2.32). Using the Schmidt form (4.10) of each bipartite state $|\Psi_\mu\rangle\!\rangle$, it follows that the dimension of the support is $d_\Psi = \sum_{\mu \in \mathsf{S}} d_\mu r_\mu$. ■

Theorem 11 (relation between likelihood and dimension). *For any pure input state $|\Psi\rangle \in \mathcal{H}$, the following bound holds:*

$$p(g|g) \le d_\Psi \ . \tag{4.23}$$

The bound is achieved if and only if the state has the form

$$|\Psi\rangle = \frac{1}{\sqrt{d_\Psi}} \bigoplus_{\mu \in \mathsf{S}} \sqrt{d_\mu r_\mu} \; e^{i\theta_\mu} \; |\Psi_\mu\rangle\!\rangle \ , \tag{4.24}$$

where $e^{i\theta_\mu}$ are arbitrary phase factors and $|\Psi_\mu\rangle\!\rangle \in \mathcal{H}_\mu \otimes \mathbb{C}^{m_\mu}$ is a bipartite state with Schmidt number r_μ and equal Schmidt coefficients, i.e. in Eq. (4.10) $\lambda^\mu_m = 1/r_\mu$ for any $m = 1, \dots, r_\mu$.

Proof. Exploiting Eq.(4.14), we have

$$\begin{aligned} p(g|g) &\le p^{\mathrm{Opt}}(g|g) && (4.25)\\ &= \left(\sum_\mu |c_\mu| \sum_{m=1}^{r_\mu} \sqrt{\lambda^\mu_m d_\mu} \right)^2 && (4.26)\\ &\le \left(\sum_\mu |c_\mu| \sqrt{r_\mu d_\mu} \right)^2 && (4.27)\\ &\le \sum_\mu r_\mu d_\mu = d_\Psi \ , && (4.28) \end{aligned}$$

the inequalities (4.27) and (4.28) coming from Schwartz inequality and from the normalizations $\sum_{m=1}^{r_\mu} \lambda_m^\mu = 1$ and $\sum_\mu |c_\mu|^2 = 1$. Let us see when this bound is attained. Clearly, the equality in (4.25) holds if we use the optimal POVM of Theorem 10. On the other hand, the Schwartz inequality (4.27) becomes equality if and only if $\lambda_m^\mu = 1/r_\mu$ for any $m = 1, \dots, r_\mu$. Finally, the last Schwartz inequality (4.28) becomes equality if and only if $|c_\mu| = \sqrt{\frac{r_\mu d_\mu}{d_\Psi}}$. The requirements $|c_\mu| = \sqrt{\frac{r_\mu d_\mu}{d_\Psi}}$ and $\lambda_m^\mu = 1/r_\mu$ are satisfied only by states of the form (4.24). ■

We can now answer to the question which are the best input states for estimating a group of unitaries: since the likelihood for a pure state is bounded by the dimension of the orbit, the best input states are the pure states with the largest orbits.

Theorem 12 (optimal input states). *For any state ρ on $\mathcal{H}$ and for any POVM, the likelihood $p_\rho(g|g)$ is bounded as follows*

$$p_\rho(g|g) \leq L = \sum_{\mu \in S} d_\mu k_\mu \ , \tag{4.29}$$

where $k_\mu \equiv \min\{d_\mu, m_\mu\}$. The bound is achieved using input pure states $\rho = |\Psi\rangle\langle\Psi|$ with $|\Psi\rangle$ of the form

$$|\Psi\rangle = \frac{1}{\sqrt{L}} \bigoplus_{\mu \in \mathsf{S}} \sqrt{d_\mu k_\mu}\ e^{i\theta_\mu}\ |E_\mu\rangle\!\rangle \ , \tag{4.30}$$

where $e^{i\theta_\mu}$ are arbitrary phase factors and $|E_\mu\rangle\!\rangle \in \mathcal{H}_\mu \otimes \mathbb{C}^{m_\mu}$ are arbitrary maximally entangled states, i.e.

$$|E_\mu\rangle\!\rangle = \frac{1}{\sqrt{k_\mu}} \sum_{m=1}^{k_\mu} |\psi_m^\mu\rangle|\phi_m^\mu\rangle \ . \tag{4.31}$$

Proof. Since the likelihood $p_\rho(g|g) = \mathrm{Tr}[\rho\Xi]$ is a linear functional of the input state, it is clear that the maximum likelihood over all possible states is achieved by a pure state. Therefore, according to Eq. (4.23), the maximum likelihood is given by the maximum of d_Ψ over all pure states. On the other hand, the value of d_Ψ for a given input state depends only on the Schmidt numbers r_μ via Eq. (4.22). Since the Schmidt number r_μ cannot exceed $k_\mu = \max\{d_\mu, m_\mu\}$, we obtain that the maximum value is

$$L \ = \ \max\{d_\Psi \mid |\Psi\rangle \in \mathcal{H}\} \ = \ \sum_{\mu \in \mathsf{S}} d_\mu k_\mu \ .$$

According to Theorem 11, such a maximum is achieved by pure states of the form (4.30). ■

The results of Theorems 10, 11, and 12 have some important consequences.

Consequence I: *Each irreducible subspace contributes to the likelihood with its dimension.*
According to Eq. (4.30), the probability of successful discrimination is maximized by exploiting in the input state all the irreducible representations appearing in the Clebsch-Gordan decomposition of U_g. Moreover, the contribution of each irreducible subspace to the likelihood is given by its dimension d_μ by Eqs. (4.22), (4.23), and (4.29). The maximum likelihood approach allows to give a general quantitative formulation to the common heuristic argument that relates the quality of the estimation to the dimension of the subspace spanned by the orbit of the input state.

Consequence II: *Role of equivalent representations.*
The repeated use of equivalent representations is crucial for attaining the maximum probability of successful discrimination. In fact, in order to achieve the upper bound (4.29) one necessarily needs to use the maximal amount of entanglement between representation spaces and multiplicity spaces, corresponding to the maximum number of irreducible subspaces carrying the same representation μ, for any μ in the Clebsch-Gordan decomposition.

Consequence III: *Maximization of the Holevo χ-quantity.*
For finite and compact groups, the optimal states in the maximum likelihood approach are those which maximize the Holevo χ-quantity[43], which in the group covariant case is defined as

$$\chi_{\mathbf{G}}(\rho) = S\left(\int \mathrm{d}g\ U_g \rho U_g^\dagger\right) - \int \mathrm{d}g\ S(U_g \rho U_g^\dagger)\ , \tag{4.32}$$

$S(\rho) = -\mathrm{Tr}[\rho \log(\rho)]$ being the von Neumann entropy. In fact, for pure input states $\rho = |\Psi\rangle\langle\Psi|$, the χ-quantity is the entropy of the average state: $\chi_{\mathbf{G}}(\rho) = S\left(\overline{\rho}\right)$. Using Eq. (2.32), we have

$$\overline{\rho} = \oplus_{\mu \in \mathsf{S}}\ |c_\mu|^2 \frac{1\!\!1_{d_\mu}}{d_\mu} \otimes \mathrm{Tr}_{\mathcal{H}_\mu}\left[|\Psi_\mu\rangle\!\rangle\langle\!\langle\Psi_\mu|\right]\ .$$

It is then easy to see that, for any pure state $\rho = |\Psi\rangle\langle\Psi|$,

$$\chi_{\mathbf{G}}(\rho) \leq \log d_\Psi\ , \tag{4.33}$$

and that the bound is attained by states of the form (4.24). Finally, the maximum over all pure states is

$$\chi_{\mathbf{G}}(\rho) = \log L\ , \tag{4.34}$$

achieved by states of the form (4.30). In this way, the likelihood of the optimal input state is directly related to the maximum χ-quantity, providing an upper bound to the amount of classical information that can be extracted from the orbit of the input state.

Consequence IV: *The case of finite groups.* According to Eq. (4.29), for any possible input state the probability of correct discrimination is bounded as follows

$$p(g|g) \leq \frac{L}{|\mathbf{G}|} \ , \tag{4.35}$$

where $L = \sum_{\mu \in \mathsf{S}} d_\mu k_\mu \quad k_\mu = \min\{d_\mu, m_\mu\}$. The bound is achieved by the optimal input states of the form (4.30), which are maximally entangled in the Clebsch-Gordan TPS. Notice that, for the optimal input states one has

$$p(g|g) = \frac{L}{|\mathbf{G}|} \leq \frac{\sum_{\mu \in \mathsf{S}} \ d_\mu k_\mu}{|\mathbf{G}|} \leq \frac{\sum_{\mu \in \mathsf{S}} d_\mu^2}{|\mathbf{G}|} \leq \frac{\sum_{\mu \in \mathrm{Irr}(\mathbf{G},\omega)} d_\mu^2}{|\mathbf{G}|} = 1 \ , \tag{4.36}$$

the last equality coming from Eq. (2.52), where ω is the cocycle of the representation $\{U_g\}$ and $\mathrm{Irr}(\mathbf{G}, \omega)$ is the collection of all irreps of $\mathbf{G}$ with cocycle ω. The above relation shows that a correct discrimination can be performed with certainty if and only if in the decomposition $U_g = \bigoplus_{\mu \in \mathsf{S}} \ U_g^\mu \otimes \mathbb{1}_{m_\mu}$ the multiplicity m_μ is at least equal to the dimension d_μ and the set S contains all the irreps with cocycle ω. The possibility of perfect discrimination for finite groups of unitaries will be analyzed extensively in the Chapter 5.

4.4 Example: estimating coherent states of the radiation field

Here we give two examples about the estimation of coherent states of a harmonic oscillator. Both cases involve the Abelian group of displacements in the complex plane, with projective representation on infinite dimensional Hilbert space $\mathcal{H}$ given by the Weyl-Heisenberg unitary operators $\{D(\alpha) = e^{\alpha a^\dagger - \alpha^* a} \mid \alpha \in \mathbb{C}\}$, where $a^\dagger$ and a are creation and annihilation operators respectively. Since the group is Abelian, it is obviously unimodular, a translation-invariant measure being $\frac{\mathrm{d}^2\alpha}{\pi}$ (here we put the constant π just for later convenience).

In the first example (two identical coherent states) we will consider two identical copies of an unknown coherent state, while in the second (conjugated coherent states) we will consider two coherent states with the same displacement in position and opposite displacement in momentum. Exploiting the method of maximum likelihood we will find in both cases the optimal POVM for the estimation of the unknown displacement. Recall that, since the group $\mathbb{C}$ is noncompact, a Bayesian approach with uniform prior is not possible, and, therefore the optimality of the presented POVMs is defined in the minimax sense. From the comparison between the sensitivities of the optimal measurements in the two cases, a close analogy will emerge with the well known example by Gisin and Popescu about the information carried by parallel and anti-parallel spins[44]. This analogy, already noticed in the study of the optimal "phase conjugation map" by Cerf and Iblisdir[98], will be analyzed here in detail from the general point of view of group parameter estimation.

Two identical coherent states

Here we consider two harmonic oscillators prepared in the same unknown coherent state $|\alpha\rangle$, $\alpha \in \mathbb{C}$. In this case, the family of signal states is

$$\mathcal{S} = \{ \ |\alpha\rangle|\alpha\rangle \in \mathcal{H}^{\otimes 2} \mid \alpha \in \mathbb{C}\} \ , \tag{4.37}$$

and is obtained from the ground state $|\Psi\rangle = |0\rangle|0\rangle$ by the action of the two-fold tensor representation $\{D(\alpha)^{\otimes 2} \mid \alpha \in \mathbb{C}\}$. The Clebsch-Gordan decomposition of such a representation can be easily obtained by using the relation

$$D(\alpha)^{\otimes 2} = V^\dagger \ D(\sqrt{2}\alpha) \otimes \mathbb{1} \ V \ , \tag{4.38}$$

where V is the unitary operator $V = \exp\left[-\frac{\pi}{4}(a_1^\dagger a_2 - a_1 a_2^\dagger)\right]$ (a_1 and a_2 being annihilation operators for the two oscillators), which in quantum optics describes the effect of a 50/50 beamsplitter. This relation shows that—modulo a non-local change of basis in the Hilbert space—the two-fold tensor representation is unitarily equivalent to a direct sum where the irreducible representation $\{D(\sqrt{2}\alpha) \mid \alpha \in \mathbb{C}\}$ appears with infinite multiplicity. Such a representation is square-summable, and has the formal dimension

$$d = \left(\int_{\mathbb{C}} \frac{\mathrm{d}^2\alpha}{\pi} \ |\langle 0| \ D(\sqrt{2}\alpha) \ |0\rangle|^2 \right)^{-1} = 2 \ , \tag{4.39}$$

calculated via Eq. (2.35). Moreover, according to Eq. (2.20), a possible decomposition of the tensor product Hilbert space into irreducible subspaces is given by any set of the form

$$\mathcal{H}_n = V^\dagger \ \mathcal{H} \otimes |\phi_n\rangle \ , \tag{4.40}$$

where $\{|\phi_n\rangle \mid n \in 0, 1, \dots\}$ is an orthonormal basis for $\mathcal{H}$. By taking the basis of eigenvectors of the number operator $a^\dagger a$, we immediately see that the input state $|\Psi\rangle = |0\rangle|0\rangle$ completely lies in the irreducible subspace $\mathcal{H}_0$. Denoting by $P_0 = V^\dagger(\mathbb{1}\otimes|0\rangle\langle 0|)V$ the projection onto $\mathcal{H}_0$, we have indeed $P_0|\Psi\rangle = |\Psi\rangle$. Using Theorem 10, we have that for the state $|\Psi\rangle$ the maximum-likelihood covariant POVM must have a seed Ξ such that $P_0\Xi P_0 = |\eta\rangle\langle\eta|$ with $|\eta\rangle = \sqrt{2}|0\rangle|0\rangle$, since here $r_\mu = 1$ (see Eq. (4.10)). Then, we have that any covariant POVM with $P_0\ \Xi\ P_0 = 2\,(|0\rangle\langle 0|)^{\otimes 2}$ is optimal for estimation of α. For example, we can take the POVM

$$P(\mathrm{d}^2\alpha) = 2\ D(\alpha)^{\otimes 2}\ \left(V^\dagger\ |\mathbb{1}\rangle\!\rangle\langle\!\langle\mathbb{1}|\ V\right)\ D(\alpha)^{\dagger\otimes 2}\ \frac{\mathrm{d}^2\alpha}{\pi}\ , \tag{4.41}$$

where the unitary V is defined as above, and $|\mathbb{1}\rangle\!\rangle$ is the vector

$$|\mathbb{1}\rangle\!\rangle\ =\ \sum_{n=0}^{\infty}\ |n\rangle|n\rangle\ . \tag{4.42}$$

It can be shown that this POVM corresponds to measuring the two commuting observables corresponding to the position of the first oscillator and the momentum of the second one. In this scheme, if the outcomes of the two measurements are q_1 and p_2 respectively, we simply declare that our estimate of the displacement is $\alpha = q_1 + ip_2$.

A different POVM which is equally optimal is

$$\widetilde{P}(\mathrm{d}^2\alpha) = 2\ D(\alpha)^{\otimes 2}\ \left(V^\dagger\ |0\rangle\langle 0| \otimes \mathbb{1}\ V\right)\ D(\alpha)^{\dagger\otimes 2}\ \frac{\mathrm{d}^2\alpha}{\pi}. \tag{4.43}$$

In a quantum optical setup, this POVM corresponds to performing firstly an heterodyne measurement on each oscillator, thus obtaining two different estimates α_1 and α_2 for the displacement, and then averaging them with equal weights. The final estimate is $\alpha = \frac{(\alpha_1+\alpha_2)}{2}$.

Although the two POVM's are different and correspond to two different experimental setups, they give rise to the same probability distribution when applied to coherent states. It is indeed straightforward to see that the probability density of estimating $\hat{\alpha}$ when the true displacement is α is given in both cases by the Gaussian

$$p(\hat{\alpha}|\alpha)\ \frac{\mathrm{d}^2\hat{\alpha}}{\pi} = 2\ e^{-2|\hat{\alpha}-\alpha|^2}\ \frac{\mathrm{d}^2\hat{\alpha}}{\pi}\ . \tag{4.44}$$

The value of the likelihood is $p(\alpha|\alpha) = 2$, according to our general formula (4.14).

Observation : *Improving the estimation with squeezing.*
The probability density of correct estimation can be improved using the

doubly-squeezed state

$$|\Psi_x\rangle\!\rangle = V^\dagger \sqrt{1-x^2} \sum_{n=0}^{\infty} x^n |n\rangle |n\rangle \ , \tag{4.45}$$

where without loss of generality we choose $x > 0$ ($x < 1$ for normalization). Then, by applying Theorem 10, it is immediate to show that $|\eta\rangle = \sqrt{2} V^\dagger |I\rangle\!\rangle$ and to evaluate the likelihood of the optimal POVM as

$$p(\alpha|\alpha) = 2 \ \frac{1+x}{1-x} \ . \tag{4.46}$$

Notice that for zero squeezing ($x = 0$) we retrieve the case of two identical coherent states, while for infinite squeezing ($x \to 1^-$) the likelihood becomes infinite, according to the fact that the displaces states $D(\alpha)^{\otimes 2} |\Psi_{x \to 1^-}\rangle\!\rangle$ become orthogonal in the Dirac sense, allowing for an ideal estimation.

Conjugated coherent states

Now the family of signal states is

$$\mathcal{S} = \{ |\alpha\rangle |\alpha^*\rangle \mid \alpha \in \mathbb{C} \} \ , \tag{4.47}$$

where complex conjugation is defined with respect to the Fock basis $\{|n\rangle \mid n = 0, 1, \dots\}$. These states are generated from the input state $|\Psi\rangle = |0\rangle|0\rangle$ by the action of the unitary representation $\{D(\alpha) \otimes D(\alpha^*) \mid \alpha \in \mathbb{C}\}$. Such a representation cannot be decomposed into a discrete Clebsch-Gordan series, due to the fact that all the unitaries in the representation can be simultaneously diagonalized on a continuous set of non normalizable eigenvectors. In fact, for any vector of the form $|D(\beta)\rangle\!\rangle = D(\beta) \otimes \mathbb{1} |\mathbb{1}\rangle\!\rangle$, where $|\mathbb{1}\rangle\!\rangle$ is the vector $|\mathbb{1}\rangle\!\rangle = \sum_n \ |n\rangle|n\rangle$, we have

$$D(\alpha) \otimes D(\alpha^*) \ |D(\beta)\rangle\!\rangle = e^{\alpha\beta^* - \alpha^*\beta} \ |D(\beta)\rangle\!\rangle \ . \tag{4.48}$$

These vectors are orthogonal in the Dirac sense, namely $\langle\!\langle D(\alpha)|D(\beta)\rangle\!\rangle = \pi \delta^2(\alpha - \beta)$. Therefore, any such vector can be regarded the basis of a one-dimensional irreducible subspace $\mathcal{H}_\beta$. The multiplicity of any irreducible representation is one, and the Hilbert space can be decomposed as a direct integral

$$\mathcal{H} \otimes \mathcal{H} = \int_{\mathbb{C}}^{\oplus} \frac{\mathrm{d}^2\beta}{\pi} \ \mathcal{H}_\beta \ . \tag{4.49}$$

This decomposition is a continuous version of the Clebsch-Gordan TPS of Eq. (2.20), with one-dimensional representation spaces $\mathcal{H}_\beta$ an one-dimensional multiplicity spaces (omitted in the integral). In the same way as in (2.23), an operator $O \in \mathcal{B}(\mathcal{H} \otimes \mathcal{H})$ in the commutant of the representation can be written as

$$O = \int_{\mathbb{C}}^{\oplus} \frac{\mathrm{d}^2\beta}{\pi} \; \mathbb{1}_\beta \; \lambda(\beta) \; , \tag{4.50}$$

where $\mathbb{1}_\beta = |D(\beta)\rangle\rangle\langle\langle D(\beta)|$ is the identity in $\mathcal{H}_\beta$, and $\lambda(\beta)$ is some scalar function.

In this particular example it is easy to extend the results of Section 4.2 to the case of a direct integral. In fact, using functional calculus we can generalize the formula (2.32) for the group average:

Proposition 9. *Let O be an operator on $\mathcal{H} \otimes \mathcal{H}$ and let*

$$\overline{O} = \int_{\mathbb{C}} \frac{\mathrm{d}^2\alpha}{\pi} \; D(\alpha) \otimes D(\alpha^*) \; O \; D(\alpha)^\dagger \otimes D(\alpha^*)^\dagger \tag{4.51}$$

be its group average. Then,

$$\overline{O} = \int_{\mathbb{C}}^{\oplus} \frac{\mathrm{d}^2\beta}{\pi} \; \mathbb{1}_\beta \; \mathrm{Tr}_{\mathcal{H}_\beta}[\; O \;] \; , \tag{4.52}$$

where $\mathbb{1}_\beta = |D(\beta)\rangle\rangle\langle\langle D(\beta)|$ and $\mathrm{Tr}_{\mathcal{H}_\beta}[\; A \;] = \langle\langle D(\beta)| \; A \; |D(\beta)\rangle\rangle$.

This expression for the group average is equivalent to that of Eq.(2.32) modulo the substitutions:

$$\begin{cases} \bigoplus_{\mu \in \mathsf{S}} & \to \quad \int_{\mathbb{C}}^{\oplus} \frac{\mathrm{d}^2\beta}{\pi} \\ \\ d_\mu & \to \quad d_\beta = 1 \qquad \forall \beta \in \mathbb{C} \end{cases} \tag{4.53}$$

The optimal POVM can be obtained from Theorem 10 by making these substitutions. We just need to decompose the input state on the irreducible subspaces, i.e.

$$|0\rangle|0\rangle = \int_{\mathbb{C}}^{\oplus} \frac{\mathrm{d}^2\beta}{\pi} \; e^{-|\beta|^2/2} \; |D(\beta)\rangle\rangle \; , \tag{4.54}$$

and then take the optimal POVM given by the operator $\Xi = |\eta\rangle\langle\eta|$ in Eqs. (4.12) and (4.13), which in the present case becomes

$$|\eta\rangle = \int_{\mathbb{C}}^{\oplus} \frac{\mathrm{d}^2\beta}{\pi} \; |D(\beta)\rangle\rangle \; . \tag{4.55}$$

Notice that, since the input state $|0\rangle|0\rangle$ has nonzero components in all the irreducible subspaces, the optimal covariant POVM is now unique. Using such optimal POVM,

$$M(\alpha) = D(\alpha) \otimes D(\alpha^*) \; |\eta\rangle\langle\eta| \; D(\alpha)^\dagger \otimes D(\alpha^*)^\dagger \; , \tag{4.56}$$

the probability density of estimating $\hat{\alpha}$ when the true displacement is α can be calculated to be the Gaussian

$$p(\hat{\alpha}|\alpha)\ \frac{\mathrm{d}^2\hat{\alpha}}{\pi} = 4\ e^{-4|\hat{\alpha}-\alpha|^2}\ \frac{\mathrm{d}^2\alpha}{\pi}\ . \tag{4.57}$$

Notice that the value of the likelihood $p(\alpha|\alpha) = 4$ could also be calculated directly using the formula (4.14), which now reads

$$p(\alpha|\alpha) = \left(\int_{\mathbb{C}} \frac{\mathrm{d}^2\beta}{\pi}\ e^{-|\beta|^2/2}\right)^2 = 4. \tag{4.58}$$

Comparing the optimal distribution (4.57) for two conjugated coherent states with the corresponding one for two identical coherent states (4.44) we can observe that the variance has been reduced by one half, while the likelihood has become twice. It is interesting to note the remarkable analogy between this example in continuous variables and the example by Gisin and Popescu [44] about the directional information encoded into a pair of parallel and anti-parallel spins. In fact, in the case of spins the authors stressed that, quite counter-intuitively, while two classical arrows pointing in opposite direction carry the same information, in a quantum mechanical setup two anti-parallel spins carry more information than two parallel ones. In the same way, in the continuous variables context, while classically two conjugated points α and α^* in the phase space carry the same information (such information being the couple of real numbers (x,p)), quantum mechanically two conjugated coherent states carry more information than two identical ones. The analogy is even closer, since for spin-$\frac{1}{2}$ particles the "spin-flip" operation is unitarily equivalent to the complex conjugation, whence we can regard also the example of spins as a comparison between pairs of identical states and pairs of conjugated states.

It is important to stress that the group theoretical analysis and the maximum likelihood approach provide in both cases also a clear explanation of the mechanism generating the asymmetry between pairs of identical and conjugated states. In fact, the whole orbit of a pair of identical coherent states (or spins states) lies just in *one* irreducible subspace of the Hilbert space, while the orbit of a pair of conjugated coherent states (spin states) covers *all* irreducible subspaces. According to formula (4.14), the likelihood in the case of conjugated states is higher than the likelihood for identical states both in the case of coherent and spin states, corresponding to an enhancement of the probability of successful discrimination.

Chapter 5

Perfect discrimination for finite symmetry groups

The unitary transformations in a finite group can be perfectly discriminated among themselves, provided that the unknown gate that we want to identify is applied a finite number of times on an entangled input state. In this Chapter we will give general upper bounds on the number of uses needed to achieve perfect discrimination. This part of the presentation, which represents a kind of long example of application of the maximum likeihood approach, can be skipped by the reader which is is more intersted of the derivation of optimal estimation strategies with generic cost functions, which is given in the following Chapter 6.

5.1 Discriminating a set of unitaries

Distinguishing states and distinguishing channels are two very different issues. While in the first case one has only to optimize the measurement used for the discrimination, in the latter one has also to optimize the choice of the input state which is fed into the channel. In particular, if the same channel is used many times, one can take advantage of the use of an entangled input state to enhance the probability of a correct discrimination. Remarkably, while two nonorthogonal pure states $|\psi_1\rangle$ and $|\phi_2\rangle$ cannot be perfectly distinguished, even if a finite number of identical copies is given, i.e. if N systems are prepared in the unknown state $\rho_i^{\otimes N} = |\psi_i\rangle\langle\psi_i|^{\otimes N}$ $i = 1, 2$, any two unitary channels U_1 and U_2 can perfectly distinguished with a finite number of uses, i.e. if the unknown gate $U_i^{\otimes N}$ $i = 1, 2$ in applied to an N-particle system [46, 47]. Recently, it has been also shown that the two unitaries U and V can be also perfectly distinguished without the use of entanglement, just by applying them many times on the same physical system[48]. This is a typical example of the common equivalence in quantum information between parallel and sequential schemes. A similar result has been recently obtained also for the estimation of an unknown unitary belonging to a one-parameter group for applications in quantum metrology[49, 50].

The possibility of perfectly distinguishing between any two unitaries also enables one to discriminate the elements of any finite set of unitaries $\{U_i \mid i \in \mathcal{I}\}$. Of course, the set must not contain a couple of unitaries U_i and U_j that are proportional up to a phase, otherwise these unitaries would induce exactly the same transformation on density matrices, and would be therefore completely indistinguishable. In the following we will refer to this as to the *minimal distinguishability requirement*. Once such a requirement is fulfilled, there is a simple method of identifying an unknown unitary in the set with zero error probability[46]. This method is based on the possibility of perfect pairwise discrimination: Let us choose two unitaries in the set (say U_1 and U_2, for example), and apply the strategy that perfectly discriminates among them, using the unknown gate a suitable number of times. If the result is "1", we can infer with certainty that the unknown transformation is not U_2. In this way, iterating the procedure of pairwise discrimination it is possible to eliminate all the wrong alternatives, thus remaining only with the correct answer.

The method based on pairwise discrimination leads to the possibility of perfect discrimination, but it is rather inefficient: to identify an unknown unitary one must eliminate $|\mathcal{I}| - 1$ wrong alternatives, that is one has to perform $|\mathcal{I}| - 1$ rounds of pairwise discrimination, each round requiring a certain number of uses of the unknown black box. Therefore, the number of uses needed to achieve perfect discrimination is at least $|\mathcal{I}| - 1$, this value corresponding to the most favorable case where the unitaries in the set are perfectly pairwise distinguishable already with a single use. It is natural to suspect that there might exist more efficient methods to identify an unknown unitary in the given set: in particular, a global discrimination strategy using an entangled input state should require less uses than the pairwise method. While the discrimination problem for arbitrary sets of unitaries in general cannot be solved analytically, for sets that have a group structure the problem becomes much simpler, allowing one to show more efficient discrimination strategies in a simple close form.

5.1.1 Necessary and sufficient conditions for perfect discrimination

Here we consider the case in which where the unitaries to be discriminated form a group representation $\{U_g \mid g \in \mathbf{G}\}$, which can be unitary or, more generally, projective. Our aim is to present a simple criterion to decide whether or not a perfect discrimination is possible performing the unknown unitary *only once*. Notice that the results about the distinguishability of two unitaries cannot be exploited for this purpose: the unitaries in a set might not be perfectly discriminated in a single run even though any two of them are perfectly distinguishable. An example of this situation is given by the four Pauli matrices, which are perfectly distinguishable pairwise, even though they are not distinguishable in a single run.

By definition, a perfect discrimination is possible if and only if there is an input state $|\psi\rangle \in \mathcal{H}$ such that the output states $\{U_g|\psi\rangle \mid g \in \mathbf{G}\}$ are mutually orthogonal, i.e. $\langle\psi|U_{g_1}^\dagger U_{g_2}|\psi\rangle = \delta_{g_1,g_2}$. Obviously, this condition implies that the dimension of the Hilbert space is at least $|\mathbf{G}|$.

If perfect discrimination is possible, the states $\{|\psi_g\rangle = U_g|\psi\rangle \mid g \in \mathbf{G}\}$ form an orthonormal basis for an invariant subspace $\mathcal{H}_\psi$. In this case, we can label the basis elements with the elements of the group, by defining $|g\rangle \doteq U_g|\psi\rangle$. The action of the representation $\{U_g\}$ on this basis is

$$U_h|g\rangle = \omega(h,g)|hg\rangle \ , \tag{5.1}$$

where ω is the cocycle defined by the relation $U_g U_h = \omega(g,h)U_{gh}$. Comparing this relation with the definition of the regular representation (Def. 13 of Chapter 2) we obtain the following necessary and sufficient condition for perfect discrimination:

Proposition 10. *Let $\mathbf{G}$ be a finite group and $\{U_g \mid g \in \mathbf{G}\}$ be a projective representation. Then the unitaries $\{U_g\}$ can be perfectly discriminated if and only if the representation $\{U_g\}$ contains the regular representation $\{U_g^{reg}\}$.*

As it was shown in Sec. 2.9, the Clebsch-Gordan decomposition of the regular representation contains all the irreps with cocycle ω, each of them with multiplicity $m_\mu^{reg} = d_\mu$. This leads immediately to another necessary and sufficient condition:

Proposition 11. *Let $\mathbf{G}$ be a finite group and $\{U_g \mid g \in \mathbf{G}\}$ be a projective representation, with Clebsch-Gordan decomposition*

$$U_g = \bigoplus_{\mu \in \mathsf{S}} U_g^\mu \otimes \mathbb{C}^{m_\mu} \ . \tag{5.2}$$

The unitaries $\{U_g\}$ can be perfectly discriminated if and only if

1. $\mathsf{S} = \mathrm{Irr}(\mathbf{G}, \omega)$

2. $m_\mu \geq d_\mu \quad \forall \mu \in \mathrm{Irr}(\mathbf{G}, \omega)$.

In other words, perfect discrimination is possible if and only if the Clebsch-Gordan decomposition contains *all the irreps* in a factor system, each of them with multiplicity m_μ greater than the dimension d_μ of the representation space. This condition is very easy to be checked: Using the orthogonality of characters, it is indeed easy to compute the multiplicities of the irreps contained in $\{U_g\}$ (Eq. (2.45)). With this simple test, it is possible to establish whether or not there is a way to discriminate the unitaries in a single run.

Remark: *Optimal discrimination and maximum likelihood.*
If the conditions of Proposition 11 are not fulfilled, the maximum likelihood approach of Chapter 4 provides anyway the optimal input states and the optimal POVMs in order to minimize the errors in the discrimination. We recall from Eq. (4.35) that the maximum of the probability of correct discrimination is given by $p(g|g) = L/|\mathbf{G}|$ where $L = \sum_{\mu \in \mathsf{S}} d_\mu k_\mu,\ k_\mu = \min\{d_\mu, m_\mu\}$. An interesting example of optimal discrimination is that of Ref.[42], where the permutations of N distinguishable particles becomes perfectly distinguishable asymptotically, under the condition that the single-particle Hilbert space has dimension $d \geq N/e$, where e is the Neper number. This is remarkably different from the classical case, where, in order to distinguish the permutations of N identical objects one needs to tag them with N different colors.

5.2 Achieving perfect discrimination with iterated uses of the unknown black box

Suppose that the unitaries in the representation $\{U_g\}$ are not perfectly distinguishable in a single run, i.e. that the condition of Proposition 11 is not fulfilled. Nevertheless, if we are allowed to perform the unknown transformation many times, a perfect discrimination is always possible, under the minimal requirement that no two of the unitaries are identical up to a phase. In the presence of group symmetry the minimal distinguishability requirement that $U_{g_1} = \lambda U_{g_2}$ implies $g_1 = g_2$ is equivalent to the condition

$$U_g = \lambda \mathbb{1} \quad \Longrightarrow \quad g = e \ . \tag{5.3}$$

A possible method is the one based on pairwise discrimination, which requires at least $N = |\mathbf{G}| - 1$ uses of the unknown black box. As already noticed, this method is clearly suboptimal, whence the question arises: which is the minimum number $N_{\min}$ of uses needed for perfect discrimination?

First of all, in order to perfectly discriminate the unitaries $\{U_g^{\otimes N} \mid g \in \mathbf{G}\}$, the Hilbert space $\mathcal{H}^{\otimes N}$ must contain at least $|\mathbf{G}|$ orthogonal vectors, whence we have the lower bound

$$N_{\min} \geq \log_d |\mathbf{G}| \; . \tag{5.4}$$

Notice, however, that in general this bound cannot be achieved. For example, in the case of the discrete phase shifts $\{U^k = e^{(2\pi i k \sigma_z)/|\mathbf{G}|} \mid k = 0, \dots, |\mathbf{G}|-1\}$ applied on a two-level system, it is easy to check that the actual value of $N_{\min}$ is $N_{\min} = |\mathbf{G}| - 1$, which is much greater than the lower bound $\log_2 |\mathbf{G}|$. In the following we will give first an abstract condition that characterize the number $N_{\min}$, and then a set of upper bounds that are easy to compute once the representation $\{U_g\}$ is given.

Applying Proposition 11 to the $N-$fold tensor representation $\{U_g^{\otimes N}\}$, it is immediate to obtain the following:

Proposition 12. *Consider the Clebsch-Gordan decomposition*

$$U_g^{\otimes N} = \bigoplus_{\mu \in \mathsf{S}^{(N)}} U_g^\mu \otimes \mathbb{1}_{m_\mu^{(N)}} \; . \tag{5.5}$$

The minimum number of uses $N_{\min}$ needed for perfect discrimination is the smallest integer N such that

1. $\mathsf{S}^{(N)} = \mathrm{Irr}(\mathbf{G}, \omega^N)$
2. $m_\mu^{(N)} \geq d_\mu \quad \forall \mu \in \mathrm{Irr}(\mathbf{G}, \omega^N)$.

This characterization is rather implicit, and, given the unitaries $\{U_g\}$ it is not immediate to establish how large $N_{\min}$ should be. Now we provide a set of upper bounds on $N_{\min}$ that can be simply calculated from the Clebsch-Gordan decomposition of $\{U_g\}$.

Bound 1. *Let $U_g = \bigoplus_{\mu \in \mathsf{S}} U_g^\mu \otimes \mathbb{1}_{m_\mu}$ a projective representation satisfying the minimal distinguishability requirement of Eq.* (5.3) *Then,*

$$N_{\min} \leq |\mathbf{G}| - L + 1 \tag{5.6}$$

where $L = \sum_{\mu \in \mathsf{S}} d_\mu k_\mu, \quad k_\mu = \min\{d_\mu, m_\mu\}$.

The proof of this bound exploits a simple lemma about linearly independent vectors obtained by Chefles [51]:

Lemma 4. *Let $\{|v_i\rangle \in \mathcal{H} \mid i = 1, \dots, n\}$ be a set of vectors, r of them linearly independent, with $r < n$. Let $\{|w_i\rangle \in \mathcal{H} \mid i, \dots, n\}$ be a set of vectors such that $|w_i\rangle \neq k|w_j\rangle$ for any $i \neq j$. Then, the set $\{|v_i\rangle|w_i\rangle \mid i = 1, \dots, n\}$ contains at least $r + 1$ linearly independent vectors.*

Proof of Bound 1. Take an optimal input state in the maximum likelihood approach, i.e. a state $|\psi\rangle$ of the form (4.30). According to Eq. (4.35), this state ensure a probability of successful discrimination equal to $p(g|g) = L/|\mathbf{G}|$, where $L = \sum_{\mu\in\mathsf{S}} d_\mu k_\mu$ is the number of linearly independent vectors in the orbit $\{U_g|\psi\rangle \mid g \in \mathbf{G}\}$. Thanks to the requirement (5.3), it is always possible to choose $|\psi\rangle$ in such a way that no unitary U_g, except the identity, has the eigenvector $|\psi\rangle$. Therefore, we are in condition to apply recursively Lemma 4, thus proving that the dimension of the orbit $\{U_g^{\otimes N}|\psi\rangle^{\otimes N}\}$ is at least $L + N - 1$. In particular, for $N = |\mathbf{G}| - L + 1$ the vectors $\{U_g^{\otimes N}|\psi\rangle^{\otimes N}\}$ are all linearly independent, namely the dimension of the orbit is $|\mathbf{G}|$. According to Theorem 11, this implies that the Hilbert space $\mathcal{H}^{\otimes N}$ contains another vector $|\Psi\rangle$ for which the probability of successful discrimination is $p_N(g|g) = |\mathbf{G}|/|\mathbf{G}| = 1$. ■

Bound 1 shows that a global discrimination is always more efficient than a pairwise one. In fact, the orbit of an input state $|\psi\rangle \in \mathcal{H}$ contains always at least two independent vectors, i.e. $L \geq 2$, therefore the number of uses $N = |\mathbf{G}| - L + 1$ is always smaller than $|\mathbf{G}| - 1$, which is the minimum number needed for pairwise discrimination.

In order to derive the second bound, we now consider the two conditions of Proposition 12 separately: we first find a number M_1 that guaranties that all the irreps are contained in the decomposition of $\{U_g^{\otimes M_1}\}$, and then a number M_2 such that the multiplicity of each irrep is enough large. To start with, we use the following

Lemma 5. *Let $U_g = \bigoplus_{\mu\in\mathsf{S}} U_g^\mu \otimes \mathbb{1}_{m_\mu}$ be a projective representation. Then, the number of linearly independent unitaries in the set $\{U_g\}$ is*

$$r = \dim \mathrm{Span}\{U_g\} = \sum_{\mu\in\mathsf{S}} d_\mu^2 \ . \tag{5.7}$$

Proof. Define the linear map $\mathcal{L} : \mathbb{C}^{|\mathbf{G}|} \to \mathcal{B}(\mathcal{H})$ which maps the vector $|\alpha\rangle = \sum \alpha(g)|g\rangle$ into the matrix $\mathcal{L}|\alpha\rangle = \sum_g \alpha(g)\, U_g$. By definition, $r = \dim \mathrm{Rng}(\mathcal{L}) = \dim \mathrm{Supp}(\mathcal{L}) = |\mathbf{G}| - \dim \mathrm{Ker}(\mathcal{L})$. Now we evaluate the dimension of the kernel. Suppose that $|\alpha\rangle \in \mathrm{Ker}(\mathcal{L})$, and decompose it on the basis of irreducible matrix elements, as in Theorem 6:

$$\alpha(g) = \sum_{\mu\in\mathrm{Irr}(\mathbf{G},\omega)} \sum_{i,j} \alpha_{ij}^\mu \ u_{ij}^{\mu *}(g) \ . \tag{5.8}$$

Using the orthogonality of irreducible matrix elements, it is immediate to see that $|\alpha\rangle$ is in the kernel if and only if the coefficients α^{μ}_{ij} are zero for any $\mu \in \mathsf{S}$ and for any $i, j = 1, \dots, d_\mu$. This implies $\dim \mathrm{Ker}(\mathcal{L}) = |\mathbf{G}| - \sum_{\mu\in\mathsf{S}} d_\mu^2$, whence $r = \sum_{\mu\in\mathsf{S}} d_\mu^2$. ■

Using Lemmas 4 and 5, we can prove the following

Proposition 13. *Let $U_g = \bigoplus_{\mu\in\mathsf{S}} U_g^\mu \otimes \mathbb{1}_{m_\mu}$ be a projective representation with cocycle ω, satisfying the minimal distinguishability requirement of Eq. (5.3). Define $M_1 = |\mathbf{G}| - r + 1$, where $r = \sum_{\mu\in\mathsf{S}} d_\mu^2$. Then, for $N \geq M_1$ the tensor representation $\{U_g^{\otimes N}\}$ contains all the irreps with cocycle ω^N.*

Proof. Since there are no unitaries in the set $\{U_g\}$ that are proportional to each other, we can apply recursively Lemma 4 to prove that the dimension of the set $\{U_g^{\otimes N}\}$ is at least $r + N - 1$. Then, for $N \geq M_1$, all the unitaries are linearly independent. On the other hand, Lemma 5 gives the equality $|\mathbf{G}| = \sum_{\mu\in\mathsf{S}^{(N)}} d_\mu^2$. Comparing this equality with Eq.(2.52), we finally obtain $S^{(N)} = \mathrm{Irr}(\mathbf{G}, \omega)$, namely the Clebsch-Gordan decomposition of $\{U_g^{\otimes N}\}$ contains all the irreps with cocycle ω^N. ■

In order to achieve a perfect discrimination, we also need that the irreps that show up in the Clebsch-Gordan decomposition of $\{U_g^{\otimes N}\}$ have a sufficient multiplicity $m_\mu^{(N)} \geq d_\mu$. The following Lemma is useful to quantify the total multiplicity in a tensor representation:

Lemma 6. *Let $\{U_g^{\otimes N}\} = \bigoplus_{\mu\in\mathsf{S}^{(N)}} U_\mu^{\otimes N} \otimes \mathbb{1}_{m_\mu^{(N)}}$ be the Clebsh-Gordan decomposition of the N-fold tensor representation of $\{U_g \in \mathcal{B}(\mathcal{H}) \mid g \in \mathbf{G}\}$. Then,*

$$\sum_{\mu\in\mathsf{S}^{(N)}} m_\mu^{(N)} \geq \frac{d^N}{\sqrt{|\mathbf{G}|}} \ , \tag{5.9}$$

where $d = \dim(\mathcal{H})$.

Proof. Decomposing the Hilbert space $\mathcal{H}^{\otimes N}$ as in Eq. (2.20) and computing the dimension on both sides, we obtain $\sum_{\mu\in\mathsf{S}^{(N)}} d_\mu m_\mu = d^N$. On the other hand, from Eq. (2.52) we have $d_\mu \leq \sqrt{|\mathbf{G}|}$ for any μ, whence Eq. (5.9) follows. ■

We are now able to prove our second bound:

Bound 2. *Let $U_g = \bigoplus_{\mu\in\mathsf{S}} U_g^\mu \otimes \mathbb{1}_{m_\mu}$ be a projective representation acting in the Hilbert space $\mathcal{H}$ and satisfying the minimal distinguishability requirement of Eq.* (5.3)*. Then the unitaries $\{U_g^{\otimes N}\}$ are perfectly distinguishable for any N such that*

$$N \geq |\mathbf{G}| - r + 1 + \log_d |\mathbf{G}| \ , \tag{5.10}$$

where $r = \sum_{\mu \in \mathsf{S}} d_\mu^2$ *and* $d = \dim \mathcal{H}$.

Proof. Define the numbers $M_1 \doteq |\mathbf{G}| - r + 1$ and $M_2 \doteq \lceil \log_d |\mathbf{G}| \rceil$, $\lceil x \rceil$ denoting the smallest integer which is larger than x. According to Proposition 13, the representation $\{U_g^{\otimes M_1}\}$ contains all irreps with cocycle ω^{M_1}, and, equivalently, the matrices $\{U_g^{\otimes M_1}\}$ are linearly independent (Lemma 5). Since the tensor product cannot decrease the number of linearly independent vectors, also the matrices $\{U_g^{\otimes M_1} \otimes U_g^\mu\}$ are linearly independent for any μ, i.e. also the representation $\{U_g^{\otimes M_1} \otimes U_g^\mu\}$ contains all the irreps of its factor system. Therefore, by writing the Clebsch-Gordan decomposition $U_g^{\otimes M_2} = \bigoplus_{\mu \in \mathsf{S}^{(M_2)}} U_g^\mu \otimes \mathbb{1}_{m_\mu^{(M_2)}}$, we have that the tensor representation $\{U_g^{\otimes M_1 + M_2}\}$ contains all irreps with cocycle $\omega^{M_1+M_2}$, each of them with multiplicity $m_\mu^{(M_1+M_2)} \geq \sum_{\mu \in \mathsf{S}(M_2)} m_\mu^{(M_2)} \geq d^{M_2}/\sqrt{|\mathbf{G}|}$, the last inequality coming from Lemma 6. Since $d^{M_2} \geq |\mathbf{G}|$ and $d_\mu \leq \sqrt{|\mathbf{G}|}$ for any μ, we have necessary $m_\mu^{(M_1+M_2)} \geq d_\mu \quad \forall \mu$. Hence, for any $N \geq M_1 + M_2$ the unitaries U_g^N can be perfectly discriminated, according to the condition of Prop. 12. ■

In many cases, the above bound allows one to dramatically cut down the number of uses of the unknown black box needed both in the pairwise discrimination method and in Bound 1.

Example. Consider the case of the "shift-and-multiply" group $\mathbf{G} = \mathbb{Z}_d \otimes \mathbb{Z}_d$, with the projective representation in the Hilbert space $\mathbb{C}^d$ given by

$$\mathsf{R}(\mathbf{G}) = \{U_{pq} = Z^p W^q \mid (p,q) \in \mathbb{Z}_d \otimes \mathbb{Z}_d\} \ , \tag{5.11}$$

where $Z = \sum_{k=0}^{d-1} |k \oplus 1\rangle\langle k|$ and $W = \sum_{k=0}^{d-1} e^{(2\pi i)/d} |k\rangle\langle k|$ for some orthonormal basis $\{|k\rangle \mid k = 0, \ldots, d-1\}$. The representation $\mathsf{R}(\mathbf{G})$ is irreducible, therefore the number of linearly independent unitaries is $r = d^2$. Since for the group $Z_d \otimes \mathbb{Z}_d$ has $|\mathbf{G}| = d^2$ elements. Bound 2 ensures that an unknown gate can be perfectly identified by applying it $N \geq 3$ times on an entangled input state. Remarkably, this is independent of the dimension d: no matters how large is the group, the unitaries $\{U_{pq}^{\otimes N}\}$ can be perfectly discriminated for any $N \geq 3$. Notice that the strategy of pairwise discrimination would require $N = d^2 - 1$ uses of the unknown black box, a number which grows quadratically with the dimension. Similarly, the method of Bound 1 would require $N \geq d^2 - d + 1$ uses.

In this example, Bound 2 works quite good. Combining it with the lower bound (5.4) we obtain $2 \leq N_{\min} \leq 3$. It is easily verified that the actual value is $N_{\min} = 2$, namely there is simple method for perfect discrimination that requires two uses of the unknown black box. Applying the unknown unitary U_{pq} to the state $\rho_0 = |0\rangle\langle 0|$, one obtains $U_{pq}|0\rangle\langle 0|U_{pq}^\dagger = |p\rangle\langle p|$, whence it is possible to determine p with zero error probability, just by measuring in the computational basis $\{|k\rangle \mid k = 0, \dots, d-1\}$. In the same way, applying U_{pq} to the state $\sigma_0 = |e_0\rangle\langle e_0|$ of the Fourier transformed basis $\{|e_k\rangle \mid k = 0, \dots, d-1\}$ defined by

$$|e_k\rangle = 1/\sqrt{d} \sum_{l=0}^{d-1} e^{(2\pi i l k)/d} \; |l\rangle \tag{5.12}$$

it is possible to determine q, with zero error probability. Hence, two copies are sufficient for perfectly identifying an unknown element of the group.

Now we will present our third bound. To this purpose, let us denote with $\chi(g) = \mathrm{Tr}[U_g]$ the characters of the representation $\{U_g\}$. A consequence of the minimal distinguishability requirement of Eq. (5.3) is that $|\chi(g)| < |\chi(e)|$ for any group element g different from the identity element e. In fact,

$$\begin{aligned}\chi(g) = \mathrm{Tr}[U_g] = \sum_{n=1}^{d} \langle n|U_g|n\rangle \leq \sum_{n=1}^{d} |\langle n|U_g|n\rangle| \\ \leq d = \mathrm{Tr}[\mathbb{1}] = \chi(e) \ .\end{aligned}$$

Here the equality is achieved if and only if U_g is a multiple of the identity, and, since the only multiple of the identity in the representation $\{U_g\}$ is U_e, the strict inequality $|\chi(g)| < |\chi(e)|$ follows for any $g \neq e$.

Let us define then the quantity

$$\alpha \doteq \max_{g \neq e} \left\{ \ \left| \frac{\chi(g)}{d} \right| \ \right\} \ . \tag{5.13}$$

Since $d = \chi(e)$, from the above discussion it is clear that $\alpha < 1$. The parameter α quantifies how much orthogonal are the unitaries with respect to the Hilbert-Schmidt product.

Now we will find the number of copies M which is sufficient to guarantee that the Clebsch-Gordan decomposition of $\{U_g^{\otimes M}\}$ contains all the irreps of the group.

Lemma 7. *If $\alpha = 0$, then for any number M the tensor product $\{U_g^{\otimes M}\}$ contains all the irreps with cocycle ω^M.*
If $\alpha > 0$, then $\{U_g^{\otimes M}\}$ contains all the irreps with cocycle ω^M for any M such that

$$M > \frac{1}{\log_d(\alpha^{-1})} \log_d |\mathsf{A}| \ , \tag{5.14}$$

where $|\mathsf{A}|$ is the cardinality the set

$$\mathsf{A} \doteq \{g \mid \chi(g) \neq 0 \ , \ g \neq e \ \} \ . \tag{5.15}$$

Proof. Let be $\mu \in \mathrm{Irr}(\mathbf{G})$ an irrep in the Clebsch-Gordan series of $U_g^{\otimes M}$. Then, according to Eq. (2.45), its multiplicity $m_\mu^{(M)}$ is given by the scalar product of characters

$$m_\mu^{(M)} = \frac{1}{|\mathbf{G}|} \sum_{g \in \mathbf{G}} \chi(g)^{\mu *} \; \chi(g)^M \ , \tag{5.16}$$

where $\chi(g)^N = \mathrm{Tr}[U_g^{\otimes N}]$. Now we have:

$$\begin{aligned} m_\mu^{(M)} &= \frac{d_\mu d^M}{|\mathbf{G}|} \left(1 + \sum_{g \in \mathsf{A}} \frac{\chi^{\mu *}(g)}{d_\mu} \; \frac{\chi(g)^M}{d^M} \right) && (5.17) \\ &\geq \frac{d_\mu d^M}{|\mathbf{G}|} \left(1 - \sum_{g \in \mathsf{A}} \left| \frac{\chi^\mu(g)}{d_\mu} \right| \; \left| \frac{\chi(g)}{d} \right|^M \right) && (5.18) \\ &\geq \frac{d_\mu d^M}{|\mathbf{G}|} \left(1 - \alpha^M \sum_{g \in \mathsf{A}} \left| \frac{\chi^\mu(g)}{d_\mu} \right| \right) && (5.19) \\ &\geq \frac{d_\mu d^M}{|\mathbf{G}|} \left(\, 1 - \alpha^M |\mathsf{A}| \, \right) \ . && (5.20) \end{aligned}$$

Clearly, if $\alpha = 0$, then all the multiplicities are nonzero, whence the representation $\{U_g^{\otimes M}\}$ contains all the irreps with cocycle ω^M. In the case $\alpha > 0$, if M satisfies (5.14), then the r.h.s. of Eq. (5.20) is positive, therefore all multiplicities are nonzero. ■

As stated by Proposition 12, to perfectly discriminate the unknown group elements we need a representation $\{U_g^{\otimes N}\}$ in which all irreps show up with multiplicity $m_\mu^{(N)} \geq d_\mu$. The following Proposition gives a condition on N under which this requirement is fulfilled.

Proposition 14. *If α=0, the unitaries $\{U_g^{\otimes N}\}$ are perfectly distinguishable for any $N \geq \log_d |\mathbf{G}|$.*
If $\alpha > 0$, then a perfect discrimination is achieved if the following condition is satisfied

$$\left\lceil \frac{d^N d_\mu}{|\mathbf{G}|} \ \left(1 - \alpha^N \ |\mathsf{A}| \ \right) \right\rceil \geq d_\mu \qquad \forall \mu \in \mathrm{Irr}(\mathbf{G}, \omega^N) \ , \tag{5.21}$$

$\lceil x \rceil$ denoting the smallest integer number greater than x.

Proof. Since the multiplicities are integer numbers, $m_\mu^{(N)}$ is greater than the upper integer part of the r.h.s. in Eq. (5.20). By requiring such a number to be greater than the dimension d_μ we obtain Eq. (5.14).■

Observation. The situation $\alpha = 0$ corresponds to the case where the unitaries $\{U_g\}$ form a *nice error basis* [52, 53]. In this case, the lower bound (5.4) can always be achieved, whence we have $N_{\min} = \lceil \log_d |\mathbf{G}| \rceil$. In other words, it is possible to perform perfect discrimination with the minimal dimension of the Hilbert space which is needed to encode the group transformations into orthogonal vectors. In particular, in the example of the "shift-and-multiply" group $\mathbf{G} = \mathbb{Z}^d \times \mathbb{Z}^d$, the above Proposition predicts correctly that the d^2 unitaries in the representation (5.11) can be prefectly discriminated with $N_{\min} = 2$ uses.

Corollary 2. *The unitaries $\{U_g^{\otimes N}\}$ can be perfectly discriminated if N satisfies*

$$d^N \left(1 - \alpha^N |\mathsf{A}|\right) \geq |\mathbf{G}| \ . \tag{5.22}$$

Proof. By definition, $\lceil x \rceil \geq x, \quad x \in \mathbb{R}$. Therefore, Eq. (5.22) implies Eq. (5.14). ■

The relation (5.22) allows to estimate the number of uses needed to discriminate with certainty among the possible choices just by considering elementary quantities such as the dimension of the Hilbert space, the traces of the unitaries, and the number of group elements. The minimum number N that satisfies Eq. (5.22) can be evaluated numerically, and essentially it has the dependence $N \approx k(\alpha) \log_d |\mathbf{G}|$, where $k > 1$ is a positive constant which is as greater as α is bigger ($k = 1$ for $\alpha = 0$).

Finally, we give now the third explicit bound on the number $N_{\min}$. To make more precise this point we derive now an explicit value $\tilde{N}$ that allows to discriminate perfectly. The bound is obtained by combining Lemma 7 with Lemma 6.

Bound 3. *Let $\{U_g\}$ be a projective representation satisfying the minimal distinguishability requirement of Eq. 5.3. Then, the unitaries $\{U_g^{\otimes N}\}$ are perfectly distinguishable for any N such that*

$$N \geq \left\lfloor \frac{\log_g |\mathbf{G}|}{\log_d (\alpha^{-1})} \right\rfloor + \lceil \log_d |\mathbf{G}| \rceil + 1 \ , \tag{5.23}$$

$\lfloor x \rfloor$ denoting the largest integer smaller than x.

Proof. Define the numbers $M_1 \doteq \left\lfloor \frac{\log_g |\mathbf{G}|}{\log_d (\alpha^{-1})} \right\rfloor + 1$ and $M_2 \doteq \lceil \log_d |\mathbf{G}| \rceil$. Due to Lemma 7, the representation $\{U_g^{\otimes M_1}\}$ contains all the irreps with cocycle ω^{M_1}. Equivalently, the unitaries $\{U_g^{\otimes M_1}\}$ are linearly independent. Since the tensor product cannot reduce the number of linearly independent vectors, also the unitaries $\{U_g^{\otimes M_1} \otimes U_g^{\mu}\}$ are linearly independent, therefore their decomposition contains all the irreps of their factor system. Therefore, the representation $\{U_g^{(M_1+M_2)}\}$ contains all the irreps with cocycle $\omega^{M_1+M_2}$, each of them with multiplicity $m_\mu^{(M_1+M_2)} \geq \sum_{\mu \in S^{(M_2)}} m_\mu^{(M_2)}$. Due to Lemma 6, one has $\sum_{\nu \in \mathsf{S}^{(M_2)}} m_\nu^{(M_2)} \geq d^{M_2}/\sqrt{|\mathbf{G}|} \geq \sqrt{|\mathbf{G}|}$. Since the relation $d_\mu \leq \sqrt{|\mathbf{G}|}$ holds for any μ, we finally obtain $m_\mu^{M_1+M_2} \geq d_\mu$, for any $\mu \in \mathrm{Irr}(\mathbf{G}, \omega^{M_1+M_2})$. Hence, for $N \geq M_1 + M_2$ the unitaries $U_g^{\otimes N}$ are perfectly distinguishable, according to Proposition 12. ■

An immediate consequence of Bound 3 is the following

Corollary 3. *The minimum number of copies $N_{\min}$ needed for perfect discrimination is forced to stay within the bounds*

$$\log_d |\mathbf{G}| \leq N_{\min} \leq k(\alpha)\ \log_d |\mathbf{G}|\ , \tag{5.24}$$

where $k(\alpha) \ = \ 2 + 1/\log_d(\alpha^{-1})$.

Proof. The upper bound is obtained by using $\lfloor x \rfloor \leq x$, $\lceil x \rceil \leq x + 1$ for any x, and recalling that $|\mathsf{A}| < |\mathbf{G}|$ (A is a proper subset of $\mathbf{G}$). The lower bound is the same of Eq. (5.4). ■

Equation (5.24) might suggest that the number $N_{\min}$ is always logarithmic in the number of group elements $\mathbf{G}$. Actually, this is the case when the maximum overlap α of the unitaries is independent of $\mathbf{G}$, as it happens in the example of nice error bases, where $\alpha = 0$. However, there are cases in which α is a function of $|\mathbf{G}|$, and, therefore the upper bound $N_{\min} \leq k(\alpha) \log_d |\mathbf{G}|$ is not of order $\mathcal{O}(\log_d |\mathbf{G}|)$. Notice that anyway, due to Bound 1, the number $N_{\min}$ is at most linear in $\mathbf{G}$. For example, this is the case of the discrete phase shifts $\{U^k = e^{(2\pi i k \sigma_z)/|\mathbf{G}|} \mid k = 0, \dots, |\mathbf{G}| - 1\}$, where the actual value of $N_{\min}$ is linear in the number of group elements, namely $N_{\min} = |\mathbf{G}| - 1$.

In this Chapter, we considered the application of the unknown gate $U_g^{\otimes N}$ to N identical systems, prepared in a suitable input state, which is possibly entangled. This parallel scheme might be compared with a sequential scheme where the unknown gate is applied N times to the same system, such as the one proposed in Ref. [48] for the discrimination of two unitaries. However, in this case the equivalence between parallel and sequential schemes cannot hold in general: For example, the four Pauli matrices cannot be perfectly distinguished when applied to a single two-level system, no matter how many times they are iterated. Another interesting question, is how to obtain a perfect discrimination of the unitaries $\{U_g\}$ by introducing an external reference system—described by the Hilbert space $\mathcal{H}_{\mathcal{R}}$—which is not sent through the unknown gate U_g. In this case, one can solve the problem by applying Proposition 11 to the representation $\{U_g^{\otimes N} \otimes \mathbb{1}_{\mathcal{R}}\}$. Notice that the effect of the reference system is to increase the multiplicities: if $m_\mu^{(N)}$ is the multiplicity of the irrep μ in the decomposition of $\{U_g^{\otimes N}\}$, then the multiplicity of μ in the decomposition of $\{U_g^{\otimes N} \otimes 1\!\!1_{d_\mathcal{R}}\}$ is $\tilde{m}_\mu^{(N)} = m_\mu^{(N)} d_\mathcal{R}$ where $d_\mathcal{R} = \dim(\mathcal{H}_\mathcal{R})$. Combining this observation with Lemma 7 we obtain the following

Bound 4. *Let $\{U_g\}$ be a projective representation satisfying the minimal distinguishability requirement and let α and A be defined as in Eqs.* (5.13) *and* (5.15)*, respectively.*
If $\alpha = 0$ the unitaries $\{U_g\}$ can be perfectly discriminated using a single probe and an external reference system of dimension $d_\mathcal{R} \geq \sqrt{|\mathbf{G}|}$. If $\alpha \neq 0$ a perfect discrimination is possible using N probes with

$$N \geq \frac{\log_d |\mathsf{A}|}{\log_d(\alpha^{-1})} \ , \tag{5.25}$$

and a reference system with dimension $d_\mathcal{R} \geq \sqrt{|\mathbf{G}|}$.

In the case of the "shift-and-multiply group" $\mathbf{G} = \mathbb{Z}_d \times \mathbb{Z}_d$ with the projective representation of Eq. (5.11), the above bound ensures that a perfect discrimination is possible with a single use of the unknown gate, exploiting entanglement with a $d-$dimensional reference system. We can recognize in this scheme the well-known dense coding for a $d-$dimensional system [54].

Chapter 6

Optimal estimation of symmetry transformations using entanglement

The strategy that minimizes the average value of a given cost function in the estimation of an unknown symmetry transformation can be explicitly derived under general assumptions. For a large class of cost functions, containing most of the physically meaningful examples, the optimal POVM is independent of the particular choice of the cost function, and coincides with the maximum likelihood POVM of Chapter 4. The optimal input states do depend instead on the particular cost function, nevertheless they can be searched without loss of generality among a class of states of a simple form, which makes evident the role of entanglement in the Clebsch-Gordan TPS induced by the group action. In this way, optimizing the input state is reduced to finding the minimum eigenvalue of a $|\mathsf{S}| \times |\mathsf{S}|$ matrix, S being the number of irreps in the decomposition of the unknown symmetry transformation. Exploiting these general results, one can simply solve problems such as the optimal communication of a spatial reference frame, the estimation of an unknown maximally entangled state, and, formally, the estimation of an unknown squeezing parameter.

6.1 General theory

6.1.1 External vs internal entanglement

Quantum entanglement is the origin of many of the most surprising advantages offered by the new technology of quantum information [3], such as computational speed-up [55, 56], quantum teleportation [57] and dense coding [54], secure protocols in cryptography [58], precision enhancement in quantum measurements [47, 59]. It is then natural to expect that an entangled input state can be useful to improve the estimation of an unknown channel. In this case, the idea is to entangle the system that undergoes the channel with another system, called the *reference system*, that undergoes the identity transformation. In the case of unitary channels, dense coding is the seminal example where this idea comes into play: the four unitaries $\{\sigma_x, \sigma_y, \sigma_z, \mathbb{1}\}$ becomes perfectly distinguishable when applied to a maximally entangled state, otherwise they can be distinguished only with error probability $p_e \geq 1/2$.

However, the role of the entanglement with a reference system in quantum estimation is far from being clear as it could appear at first sight. For example, in the discrimination between two unitaries [46, 47] there is no improvement coming from the use of entangled input state, and, moreover, a maximally entangled state is often worse than a separable one. In the estimation of an $\mathbb{SU}(d)$ transformation entanglement is very useful[60, 61], while in phase estimation—corresponding to the group $U(1)$—it is not useful at all[62]. The question then naturally arises: Is it possible to understand once for all the role of entanglement in a general fashion? Moreover, it is well known that entanglement is not an absolute concept, but it depends on the algebras that are chosen to represent the "local operations" on the subsystems of a compound system[63]. Different algebras define different subsystems, and hence different tensor product structures on the same Hilbert space. A second question is then: which kind of entanglement is useful for estimating an unknown channel?

According to the main target of this presentation, in this Chapter we will consider the case of unitary channels corresponding to symmetry transformations, i.e. elements of some projective representation $\mathsf{R}(\mathbf{G}) = \{U_g \mid g \in \mathbf{G}\}$. In this case, it is crucial to take in account that the Hilbert space supporting the representation $\{U_g\}$ already possesses a natural tensor product structure, namely the Clebsch-Gordan TPS of Eqs. (2.20) and (2.22). As we already noted in the previous chapters, the entanglement between representation and multiplicity spaces in the Clebsch-Gordan TPS is the key feature of the optimal input states in the maximum likelihood approach (Section 4.3), and is a necessary ingredient to obtain a perfect discrimination in the case of finite groups (Subsection 5.1.1). Here we prove that, for a large class of optimality criteria, this kind of entanglement is the only useful entanglement for the estimation of an unknown symmetry transformation $\{U_g\}$.

Using a simple group theoretical analysis, the role of the reference system in an estimation strategy becomes immediately clear: introducing it is equivalent to substituting the representation $\{U_g\}$, acting on $\mathcal{H}$, with the representation $\{U'_g = U_g \otimes \mathbb{1}_{\mathcal{R}}\}$, acting on $\mathcal{H} \otimes \mathcal{H}_{\mathcal{R}}$, where $\mathcal{H}_{\mathcal{R}}$ is the reference system's Hilbert space. If the representation $\{U_g\}$ has the decomposition $U_g = \bigoplus_{\mu \in \mathsf{S}} U_g^\mu \otimes \mathbb{1}_{m_\mu}$, then the representation $\{U'_g\}$ has decomposition

$$U'_g = \bigoplus_{\mu \in \mathsf{S}} U_g \otimes \mathbb{1}_{m_\mu d_{\mathcal{R}}} \ , \tag{6.1}$$

where $d_{\mathcal{R}} = \dim(\mathcal{H}_{\mathcal{R}})$. In other words, adding the reference system has only the effect of increasing the multiplicities, while the irreps contained in the Clebsch-Gordan decomposition remain exactly the same. Since any state $|\Psi\rangle \in \mathcal{H} \otimes \mathcal{H}_{\mathcal{R}}$ can be decomposed as $|\Psi\rangle = \bigoplus_{\mu \in \mathsf{S}} \ c_\mu |\Psi_\mu\rangle\!\rangle$ with $|\Psi_\mu\rangle\!\rangle$ being bipartite states in $\mathcal{H}_\mu \otimes \mathbb{C}^{m_\mu d_{\mathcal{R}}}$, is clear that it is not useful to have a multiplicity $m_\mu d_{\mathcal{R}}$ larger than the dimension d_μ of the representation space. In fact, the representation space $\mathcal{H}_\mu$ can entangle itself at most with a $d_\mu-$dimensional subspace of the multiplicity space $\mathbb{C}^{m_\mu d_{\mathcal{R}}}$. This simple observation has a very relevant consequence:

Proposition 15. *For the estimation of the representation $U_g = \bigoplus_{\mu \in \mathsf{S}} \ U_g^\mu \otimes \mathbb{1}_{m_\mu}$ it is not useful to have a reference system of dimension $d_{\mathcal{R}}$ larger than*

$$d_{\mathcal{R}}^{\min} = \max \left\{ \left\lceil \frac{d_\mu}{m_\mu} \right\rceil \ \mid \mu \in \mathsf{S} \right\} \ , \tag{6.2}$$

$\lceil x \rceil$ denoting the smallest integer such that $\lceil x \rceil \geq x$.

An immediate consequence of this Proposition if the following:

Corollary 4. *Let $\mathbf{G}$ be an Abelian group and $\{U_g \mid g \in \mathbf{G}\}$ a unitary representation. Then, an external reference system is useless.*

Proof. All the unitary irreps of an Abelian group are one-dimensional, i.e. $d_\mu = 1 \ \forall \mu \in \mathrm{Irr}(\mathbf{G}, 1)$, whence $d_{\mathcal{R}}^{\min} = 1$ in Prop. 15. Therefore, the reference system can be taken as one-dimensional, i.e. it can be neglected. ■

The above results do not refer to a particular approach to optimization, since they are just a simple consequence to the group structure of the problem. In particular, Corollary 4 implies that an external reference system is of no use for the problem of phase estimation, involving the Abelian group $U(1)$ ($U(1)^{\times d}$, for multiple phase estimation). In the research about quantum Cramér-Rao bound, this fact was reported as a new feature in Ref. [62], however the uselessness of a reference system for phase estimation was already a straightforward consequence of Ref. [64], also stressed in Refs. [15] and [16]. From the group theoretical point of view, this feature is indeed rather trivial.

The external reference system can be regarded as an additional resource, since having a reference system requires the ability of performing the identical transformation on it, i.e. the ability of preserving it from any possible source of decoherence. The above Proposition allows one to optimize the use of such a resource by minimizing the dimension of the system that needs to be protected against unwanted transformations. This is very useful in situations such as the absolute transmission of a spatial reference frame, where two distant parties (say Alice and Bob) want to establish a frame of Cartesian axes by sending directional qubits, i.e. spin 1/2 particles. In this case any qubit sent from Alice appears rotated to Bob, and estimating this rotation is equivalent for Bob to estimating the directions of Alice's axes. Of course, any directional qubit is affected by the same rotation, and therefore it cannot be used as a reference system. In this case, in order to have a reference system one should use different degrees of freedom that are rotationally invariant, or special encoding schemes that provide rotationally invariant logical qubits, as the one proposed by Bartlett, Rudolph, and Spekkens [4]. Nevertheless, the multiplicity spaces encapsulated into the Clebsch-Gordan TPS make possible to minimize (and asymptotically get rid of) this additional resource, by substituting external with internal entanglement. This allows Alice and Bob to optimally transmit a spatial frame only by sending directional qubits.

6.1.2 The generalized Holevo class of cost functions

In a pioneering work about phase estimation [65], Holevo introduced a class of cost functions $c(\hat{\phi}, \phi)$ having the following form:

$$c(\hat{\phi} - \phi) = \sum_{k \in \mathbb{Z}} a_k \, e^{-ik(\hat{\phi}-\phi)} \ , \qquad a_k \leq 0 \quad \forall k \neq 0 \ , \tag{6.3}$$

i.e. functions with negative coefficients in the Fourier series, except for the constant offset $a_0 = \int_{-\pi}^{\pi} \mathrm{d}\phi/(2\pi) \ c(\phi)$.

This class covers most of the physically meaningful optimality criteria, such as

- maximum likelihood, with $c_{ML}(\hat{\phi}-\phi) = -\delta(\hat{\phi}-\phi) = -\frac{1}{2\pi} \sum_{k\in\mathbb{Z}} e^{ik(\hat{\phi}-\phi)}$
- maximum fidelity, with $c_{MF}(\hat{\phi}-\phi) = 1 - \left|\langle\psi|U_\phi^\dagger U_{\hat{\psi}}|\psi\rangle\right|^2$, $|\psi\rangle$ being any state in $\mathcal{H}$
- minimum angular dispersion, with $c_{md}(\hat{\phi}-\phi) = 4\sin^2\frac{(\hat{\phi}-\phi)}{2}$.

Given an input state and a cost function in the Holevo class, the optimization of the POVM can be solved analytically in a simple way. Remarkably, for fixed input state, all cost functions in the class lead to the same optimal POVM[2].

In the following we will show how the results by Holevo can be generalized to treat not only the group $U(1)$ involved in phase estimation, but also any finite and compact Lie group $\mathbf{G}$. As the first step in this program, we give now a proper generalization of the Holevo class, by considering cost functions $c(\hat{g}, g)$ that satisfy the following two requirements:

First requirement. We require $c(\hat{g}, g)$ to be group invariant, namely

$$c(\hat{g}, g) = c(k\hat{g}, kg) \qquad \forall \hat{g}, g, k \in \mathbf{G} \tag{6.4}$$

(left-invariance), and

$$c(\hat{g}, g) = c(gk, \hat{g}k) \qquad \forall \hat{g}, g, k \in \mathbf{G} \tag{6.5}$$

(right-invariance). By using the Fourier analysis on the group (Sec. 2.10), one can prove that this assumption is equivalent to the expansion

$$c(\hat{g}, g) = \sum_{\sigma} a_\sigma \ \chi^{\sigma *}(\hat{g} g^{-1}) \ , \tag{6.6}$$

where $\chi^\sigma(g) \equiv \mathrm{Tr}[U^\sigma(g)]$ is the character of the irreducible representation σ, and the coefficients a_σ satisfy the identity $a^*_\sigma = a_{\sigma *} \quad \forall \sigma$, in order to have a real cost function (for the proof, see the Appendix of Ref. [18]).

Second requirement. We require all nonzero coefficients a_σ in the generalized Fourier expansion (6.6) to be negative, with the only exception of the coefficient a_{σ_0} corresponding to the trivial representation $U^{\sigma_0}(g) = 1 \quad \forall g$, which is allowed to be positive (the σ_0 term just adds a trivial offset to the cost function, since $\chi^{\sigma_0}(g) = 1 \quad \forall g$).

The above requirements define a direct generalization of the Holevo class, which is retrieved in the case of the group $\mathbf{G} = U(1)$. Our generalized Holevo class always contains physically meaningful cost functions, such as:

- maximum likelihood, with

  $$c_{ML}(\hat{g}, g) = -\delta(\hat{g} g^{-1}) = - \sum_{\mu \in \mathrm{Irr}(\mathbf{G}, \omega)} d_\mu \chi^*_\mu(\hat{g} g^{-1})$$

  (see Eq. (2.64))

- maximum entanglement fidelity $c_{ME}(\hat{g}, g) = 1 - \left| \langle\langle E | U^\dagger_g U_{\hat{g}} \otimes \mathbb{1} | E \rangle\rangle \right|^2$, where $|E\rangle\rangle \in \mathcal{H} \otimes \mathcal{H}$ is any maximally entangled state.

6.1.3 Optimal input states

Here optimality will be defined in the Bayesian approach with uniform prior distribution, corresponding to a complete ignorance about the "true" transformation. Accordingly, optimality will be defined as the minimization of the Bayes expected cost

$$\langle c \rangle = \int_{\mathbf{G}} \mathrm{d}\hat{g} \int_{\mathbf{G}} \mathrm{d}g \; c(\hat{g}, g) \; \mathrm{Tr}[M(\hat{g}) \; U_g \rho U_g^\dagger] \; , \tag{6.7}$$

$M(\mathrm{d}\hat{g})$ being the operator density of the estimating POVM.

Since the average cost (6.7) is a linear functional of the input state ρ, in the optimization problem we can restrict attention to *pure* input states $\rho = |\Psi\rangle\langle\Psi|$. Then the problem becomes equivalent to the optimal discrimination problem of states in the orbit

$$\mathcal{O} = \left\{ |\Psi_g\rangle\langle\Psi_g| \equiv U_g \; |\Psi\rangle\langle\Psi| U_g^\dagger \; | \; g \in \mathbf{G} \right\} \tag{6.8}$$

generated from the input state $|\Psi\rangle$ by the action of the representation $\{U_g\}$.

In the following we allow the use of an external reference system in the estimation strategy, and look for the optimal input states in $\mathcal{H} \otimes \mathcal{H}_{\mathcal{R}}$. In this situation we have the following:

Proposition 16. *Let $U_g = \bigoplus_{\mu \in \mathsf{S}} U_g^\mu \otimes \mathbb{1}_{m_\mu}$ be the Clebsch-Gordan decomposition of the unknown unitaries $\{U_g\}$. If the use of an external reference system is allowed, then there is no loss of generality in assuming the condition*

$$m_\mu = d_\mu \qquad \forall \mu \in \mathsf{S} \; . \tag{6.9}$$

Proof. Suppose $d_\mu > m_\mu$ for some representation μ. In this case, we can introduce a reference system $\mathcal{R}$ whose dimension is

$$d_{\mathcal{R}} \geq \max_{\mu \in \mathsf{S}} \left\{ \frac{d_\mu}{m_\mu} \right\} \; , \tag{6.10}$$

and replace U_g with its extension $U_g' = U_g \otimes \mathbb{1}_{\mathcal{R}}$, acting on the compound system. In this way, U_g' will satisfy the condition $m_\mu' \equiv m_\mu \times d_{\mathcal{R}} \geq d_\mu \quad \forall \mu$. On the other hand, any pure state $|\Psi\rangle$ can be decomposed in the form (4.7) with $|\Psi_\mu\rangle\rangle \in \mathcal{H}_\mu \otimes \mathbb{C}^{m_\mu}$ being bipartite states. Since the Schmidt number of each state $|\Psi_\mu\rangle\rangle$ cannot exceed $k_\mu = \min\{d_\mu, m_\mu\}$, we can switch our attention from the whole Hilbert space $\mathcal{H} \otimes \mathcal{H}_{\mathcal{R}} = \bigoplus_\mu \; \mathcal{H}_\mu \otimes \mathbb{C}^{m_\mu'}$ to the invariant subspace $\mathcal{H}' \equiv \bigoplus_\mu \; \mathcal{H}_\mu \otimes \mathbb{C}^{d_\mu}$, which contains the input state $|\Psi\rangle$ along with its orbit (6.8). In other words, without loss of generality we can always consider an input state in the Hilbert space

$$\mathcal{H}' = \bigoplus_{\mu} \mathcal{H}_\mu \otimes \mathbb{C}^{d_\mu} \ , \tag{6.11}$$

which is embedded in the larger space $\mathcal{H} \otimes \mathcal{H}_{\mathcal{R}}$. ■

Now we show that the best input state $|\Psi\rangle$ for estimating the group transformation of an unknown black box is a state of the form $|\Psi\rangle = \bigoplus_{\mu \in \mathsf{S}} c_\mu |\Psi_\mu\rangle\!\rangle$, with each $|\Psi_\mu\rangle\!\rangle$ maximally entangled, namely

$$|\Psi_\mu\rangle\!\rangle = \frac{1}{\sqrt{d_\mu}} \sum_{n=1}^{d_\mu} |\psi_n^\mu\rangle|\phi_n^\mu\rangle \ , \tag{6.12}$$

$\mathsf{B}_A^\mu = \{|\psi_n^\mu\rangle \mid n = 1, \ldots, d_\mu\}$ and $\mathsf{B}_B^\mu = \{|\phi_n^\mu\rangle \mid n = 1, \ldots, d_\mu\}$ being Schmidt bases for $\mathcal{H}_\mu$ and $\mathbb{C}^{d_\mu}$ respectively.

In order to deal with bipartite states, it is very useful to introduce a convenient notation [66]. Given two Hilbert spaces $\mathcal{H}_A$ and $\mathcal{H}_B$, and fixed two orthonormal bases $\mathcal{B}_A = \{|\phi_n\rangle \mid n = 1, \ldots, d_A\}$ and $\mathcal{B}_B = \{|\psi_n\rangle \mid n = 1, \ldots, d_B\}$ for $\mathcal{H}_A$ and $\mathcal{H}_B$ respectively, it is possible to associate in a one to one way any vector $|C\rangle\!\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$ with an operator $C : \mathcal{H}_B \longrightarrow H_A$ via the relation

$$|C\rangle\!\rangle = \sum_{m,n} \langle \phi_m | C | \psi_n \rangle \ |\phi_m\rangle|\psi_n\rangle \ . \tag{6.13}$$

With this notation, one has the simple relations

$$\langle\!\langle C|D\rangle\!\rangle = \mathrm{Tr}[C^\dagger D] \tag{6.14}$$

and

$$A \otimes B \ |C\rangle\!\rangle = |ACB^T\rangle\!\rangle \qquad \forall A \in \mathcal{B}(\mathcal{H}_A), \ \forall B \in \mathcal{B}(\mathcal{H}_B) \ , \tag{6.15}$$

where transposition T is defined with respect to the fixed bases.

Exploiting the notation (6.13)—with fixed bases B_A^μ and B_B^μ—the optimal input state $|\Psi\rangle$ must have the form

$$|\Psi\rangle = \bigoplus_{\mu \in \mathsf{S}} \frac{c_\mu}{\sqrt{d_\mu}} \ |W_\mu\rangle\!\rangle \ , \tag{6.16}$$

with $W_\mu \equiv \sum_n \ |\psi_n^\mu\rangle\langle\phi_n^\mu|$ unitary operators.

Theorem 13 (optimal input states). *With a suitable choice of the coefficients $\{c_\mu\}$, any input state of the form (6.16) achieves the minimum Bayes cost.*

Suppose that the minimum cost $\langle c\rangle^{Opt}$ is achieved by the input state $|\Phi\rangle = \bigoplus_\mu \; c_\mu \; |\Phi_\mu\rangle\!\rangle$ along with the estimation strategy described by the covariant density $M(g)$. The operator $K_h \equiv \bigoplus_\mu \; 1\!\!1_\mu \otimes \sqrt{d_\mu}\left(W_\mu^\dagger U_h^\mu \Phi_\mu\right)^T$ converts the orbit of an input state (6.16) into the orbit of the optimal input state $|\Phi\rangle$, since using identity (6.15), we have

$$K_h \; |\Psi_g\rangle = |\Phi_{gh}\rangle \; , \tag{6.17}$$

where $|\Psi_g\rangle = U_g|\Psi\rangle$ and $|\Phi_g\rangle = U_g|\Phi\rangle$. Consider now the POVM density $M'(g) \equiv \int \mathrm{d}h \; K_h^\dagger \; M(gh) \; K_h$. The POVM $M'(g)$ is normalized, since

$$\begin{aligned} \int \mathrm{d}g \; M'(g) \;\; &= \;\; \int \mathrm{d}g \int \mathrm{d}h \, K_h^\dagger \; M(gh) \; K_h \\ &= \;\; \int \mathrm{d}h \; K_h^\dagger K_h \\ &= \;\; 1\!\!1 \; , \end{aligned}$$

where we exchanged integrals over g and h, used invariance of the Haar measure $\mathrm{d}g$, and finally used Eq. (2.32) and the normalization of bipartite states $|\Phi_\mu\rangle\!\rangle$ in the form $\mathrm{Tr}[\Phi_\mu^\dagger \Phi_\mu] = 1$. A state $|\Psi\rangle$ of the form (6.16) along with the POVM $M'(g)$ achieves the minimum cost. In fact,

$$\begin{aligned} \langle c\rangle \;\; &= \;\; \int \mathrm{d}g \int \mathrm{d}\hat{g} \; c(\hat{g}, g) \; \langle \Psi_g| \; M'(\hat{g}) \; |\Psi_g\rangle \\ &= \;\; \int \mathrm{d}g \int \mathrm{d}\hat{g} \int \mathrm{d}h \; c(\hat{g}, g) \; \langle \Phi_{gh}| \; M(\hat{g}h) \; |\Phi_{gh}\rangle \\ &= \;\; \int \mathrm{d}g \int \mathrm{d}\hat{g} \int \mathrm{d}h \; c(\hat{g}h, gh) \langle \Phi_{gh}| \; M(\hat{g}h) \; |\Phi_{gh}\rangle \\ &= \;\; \int \mathrm{d}k \int \mathrm{d}\hat{k} \; c(\hat{k}, k) \; \langle \Phi_k| \; M(\hat{k}) \; |\Phi_k\rangle \\ &= \;\; \langle c\rangle^{Opt} \; , \end{aligned}$$

where we used right-invariance of both cost function and Haar measure. ■

6.1.4 Covariance properties of the estimating POVM

Since the whole orbit (6.8) is generated from the input state $|\Psi\rangle$ by the action $\mathsf{R}(\mathbf{G})$ of the group, there is no loss of generality in restricting attention to estimating POVM of the covariant form [2]

$$P(\mathrm{d}g) = M(g)\mathrm{d}g = U_g \; \Xi \; U_g^\dagger \; \mathrm{d}g \tag{6.18}$$

with Ξ a suitable positive operator satisfying the normalization condition (3.17). A covariant POVM yields a left-invariant probability distribution, namely $p(k\hat{g}|kg) = p(\hat{g}|g) \quad \forall k, \hat{g}, g \in \mathbf{G}$. Using both the left-invariance of the probability distribution and of the cost function, the average cost (6.7) can be written as

$$\langle c \rangle = \int \mathrm{d}g \ c(g,e) \ p(g|e) \tag{6.19}$$

where e is the identity element of the group $\mathbf{G}$.

For superpositions of maximally entangled states as in Eq. (6.16), the orbit $\mathcal{O}$ enjoys an additional symmetry that is reflected in an additional covariance property of the POVM. In fact, using the decomposition (2.22) and the identity (6.15), we can note that

$$\begin{aligned} |\Psi_g\rangle &= U_g \ |\Psi\rangle \\ &= \bigoplus_{\mu \in \mathsf{S}} \frac{c_\mu}{\sqrt{d_\mu}} \ (U_g^\mu \otimes \mathbb{1}_\mu) \ |W_\mu\rangle\!\rangle \\ &= \bigoplus_{\mu \in \mathsf{S}} \frac{c_\mu}{\sqrt{d_\mu}} \ [\mathbb{1}_\mu \otimes (W_\mu^\dagger U_g^\mu W_\mu)^T] \ |W_\mu\rangle\!\rangle \\ &= V_g^\dagger \ |\Psi\rangle \qquad \forall g \in \mathbf{G} \ , \end{aligned}$$

where

$$V_g \equiv \oplus_{\mu \in \mathsf{S}} \ (\mathbb{1}_\mu \otimes (W_\mu^\dagger U_g^\mu W_\mu)^*) \ , \tag{6.20}$$

is an element of a new projective representation $\mathsf{R}'(\mathbf{G})$ of the group $\mathbf{G}$. Notice that the two representations $\mathsf{R}(\mathbf{G})$ and $\mathsf{R}'(\mathbf{G})$ commute among themselves, and for $\mathsf{R}'(\mathbf{G})$ the role of representation and multiplicity spaces is exchanged with respect to $\mathsf{R}(\mathbf{G})$. Then, the following Lemma holds:

Lemma 8. *There is no loss of generality in assuming a covariant density* $M(g) = U_g \ \Xi \ U_g^\dagger$ *with*

$$[\Xi, U_g V_g] = 0 \qquad \forall g \in \mathbf{G} \ , \tag{6.21}$$

where U_g *and* V_g *are given in Eqs. (2.22) and (6.20), respectively.*

Proof. For any possible density $N(g)$ there is a covariant density with the above property and with the same average cost. In fact, the group average

$$M(g) = \int \mathrm{d}k \int \mathrm{d}h \ U_k^\dagger V_h^\dagger \ N(kgh^{-1}) \ U_k V_h \tag{6.22}$$

is covariant—namely $M(g) = U_g \ \Xi \ U_g^\dagger$ with $\Xi = M(e)$—and satisfies the required commutation relation (6.21). Both properties follow simply from the invariance of the Haar measure. To prove that the cost of the covariant density $M(g)$ is the same as the cost of $N(g)$ we use the property

$$U_k V_h \, |\Psi_g\rangle = |\Psi_{kgh^{-1}}\rangle \qquad \forall k, h, g \in \mathbf{G} \tag{6.23}$$

of the states generated from the input (6.16). In this way,

$$\begin{aligned}
\langle c\rangle_M &\equiv \int \mathrm{d}g \int \mathrm{d}\hat{g} \; c(\hat{g}, g) \; \langle\Psi_g| \; M(\hat{g}) \; |\Psi_g\rangle \\
&= \int \mathrm{d}g \int \mathrm{d}\hat{g} \int \mathrm{d}k \int \mathrm{d}h \; c(\hat{g}, g) \; \times \\
&\qquad\qquad \times \langle\Psi_{kgh^{-1}}| \; N(k\hat{g}h^{-1}) \; |\Psi_{kgh^{-1}}\rangle \\
&= \int \mathrm{d}g \int \mathrm{d}\hat{g} \int \mathrm{d}k \int \mathrm{d}h \; c(k\hat{g}h^{-1}, kgh^{-1}) \; \times \\
&\qquad\qquad \times \langle\Psi_{kgh^{-1}}| \; N(k\hat{g}h^{-1}) \; |\Psi_{kgh^{-1}}\rangle \\
&= \int \mathrm{d}r \int \mathrm{d}\hat{r} \; c(\hat{r}, r) \; \langle\Psi_r| \; N(\hat{r}) \; |\Psi_r\rangle \\
&\equiv \langle c\rangle_N \; ,
\end{aligned}$$

where we used the left- and right-invariance of the cost function $c(\hat{g}, g)$. ■

Let us diagonalize the operator Ξ and express its (non-normalized) eigenvectors in the decomposition (2.20):

$$\begin{aligned}
\Xi &= \sum_{i=1}^{r} |\eta^i\rangle\langle\eta^i| \\
&= \sum_i \bigoplus_{\mu,\nu} \sqrt{d_\mu d_\nu} \; |\eta^i_\mu\rangle\!\rangle\langle\!\langle\eta^i_\nu| \; ,
\end{aligned} \tag{6.24}$$

where the factor $\sqrt{d_\mu}$ has been inserted just for later convenience.

Lemma 9. *Any covariant density $M(g) = U_g \; \Xi \; U_g^\dagger$ with the commutation property (6.21) must satisfy the two relations:*

$$\sum_i \eta^{i\dagger}_\mu \eta^i_\mu = \mathbb{1}_\mu \qquad \forall \mu \in \mathsf{S} \; , \tag{6.25}$$

and

$$\sum_i \eta^i_\mu \eta^{i\dagger}_\mu = \mathbb{1}_\mu \qquad \forall \mu \in \mathsf{S} \; . \tag{6.26}$$

Proof. Consider the normalization constraint given by Eq. (3.17), i.e. $\mathrm{Tr}_{\mathcal{H}_\mu}[\Xi] = d_\mu \ 1\!\!1_\mu \quad \forall \mu \in \mathsf{S}$. By explicit computation,

$$\begin{aligned}
\mathrm{Tr}_{\mathcal{H}_\mu}[\Xi] &= d_\mu \sum_i \ \mathrm{Tr}_{\mathcal{H}_\mu}\left[|\eta^i_\mu\rangle\!\rangle\langle\!\langle\eta^i_\mu| \ \right] \\
&= d_\mu \sum_i \ \eta^{iT}_\mu \ \mathrm{Tr}_{\mathcal{H}_\mu}[\ |1\!\!1_\mu\rangle\!\rangle\langle\!\langle 1\!\!1_\mu| \ ] \ \eta^{i*}_\mu \\
&= d_\mu \ \sum_i \ \eta^{iT}_\mu \eta^{i*}_\mu \ .
\end{aligned}$$

Substituting this expression in the normalization constraint $\mathrm{Tr}_{\mathcal{H}_\mu}[\Xi] = d_\mu 1\!\!1_{m_\mu}$ and taking the complex conjugate we get (6.25). Moreover, using the commutation relation (6.21), we can transform the group average with respect to $\{U_g\}$ in a group average with respect to $\{V_g\}$, namely

$$\begin{aligned}
1\!\!1 &= \int \mathrm{d}g \ U_g \Xi U^\dagger_g \\
&= \int \mathrm{d}g \ U_g \ (U^\dagger_g V^\dagger_g \ \Xi \ U_g V_g) \ U^\dagger_g \\
&= \int \mathrm{d}g \ V^\dagger_g \ \Xi \ V_g \ .
\end{aligned}$$

Then, using Eq. (2.32) for the group average we obtain the relation $\mathrm{Tr}_{\mathbb{C}^{m_\mu}}[\Xi] = d_\mu 1\!\!1_{\mathcal{H}_\mu}$. Then, Eq. (6.26) can be proved along in the same way as Eq. (6.25). ■

6.1.5 Optimal POVMs

We are now able to find the optimal covariant POVM for the estimation of a group transformation applied to a coherent superpositions of maximally entangled states.

Theorem 14 (optimal POVM). *In the estimation of the states in the orbit $\mathcal{O}$ generated from the input state*

$$|\Psi\rangle = \bigoplus_{\mu \in \mathsf{S}} c_\mu \ \frac{|W_\mu\rangle\!\rangle}{\sqrt{d_\mu}} \ , \tag{6.27}$$

where W_μ are unitary operators, the covariant POVM given by $M(\mathrm{d}g) = U_g|\eta\rangle\langle\eta|U^\dagger_g \ \mathrm{d}g$ with

$$|\eta\rangle = \bigoplus_{\mu \in \mathsf{S}} \ \sqrt{d_\mu} \ e^{i\,\arg(c_\mu)} \ |W_\mu\rangle\!\rangle \tag{6.28}$$

is optimal for any cost function $c(\hat{g},g)$ of the form

$$c(\hat{g},g)=\sum_{\sigma}\ a_{\sigma}\ \chi^{\sigma *}(\hat{g}g^{-1})\ , \tag{6.29}$$

with $a_{\sigma}\leq 0 \qquad \forall \sigma\neq\sigma_0$.
The average cost corresponding to the optimal estimation strategy is

$$\langle c\rangle^{Opt}=\sum_{\mu,\nu}\ |c_{\mu}|\ C_{\mu\nu}\ |c_{\nu}|\ , \tag{6.30}$$

where the cost matrix $C_{\mu\nu}$ is defined by

$$C_{\mu\nu}\doteq a_{\sigma_0}\ \delta_{\mu\nu}+\sum_{\sigma\neq\sigma_0}\ a_{\sigma}\ m_{\sigma}^{(\mu\nu)}\ , \tag{6.31}$$

$m_{\sigma}^{(\mu\nu)}$ being the multiplicity of the irreducible representation σ in the Clebsch-Gordan series of the tensor product $U_g^{\mu}\otimes U_g^{\nu *}$.

Proof. We will show that Eq. (6.30) gives a lower bound for the average cost, and that the POVM $\Xi=|\eta\rangle\langle\eta|$ with $|\eta\rangle$ given by Eq. (6.28) achieves this bound. By using identities (6.14) and (6.15), and the form (6.24) for the operator Ξ, Eq. (6.19) becomes

$$\begin{aligned}\langle c\rangle &= \int \mathrm{d}g\ c(g,e)\times \\ &\quad \times\sum_{i}\ \sum_{\mu,\nu}\ c_{\mu}^{*}c_{\nu}\ \mathrm{Tr}\left[W_{\mu}^{\dagger}\ U_g^{\mu}\ \eta_{\mu}^{i}\otimes W_{\nu}^{T}\ U_g^{\nu *}\ \eta_{\nu}^{i*}\right]\ .\end{aligned}$$

Let's expand $c(g,e)$ as in (6.29). Subtracting from the average cost $\langle c\rangle$ the constant term a_{σ_0}, which is not relevant for the optimization, we get

$$\begin{aligned}\langle c\rangle - a_{\sigma_0} &= \sum_{i}\sum_{\mu,\nu}\ c_{\mu}^{*}c_{\nu}\ \times \\ &\quad\times\sum_{\sigma\neq\sigma_0}\ \frac{a_{\sigma}}{d_{\sigma}}\ \mathrm{Tr}\left[\Pi_{\sigma}^{(\mu\nu)}\ (\eta_{\mu}^{i}W_{\mu}^{\dagger}\otimes\eta_{\nu}^{i*}\ W_{\nu}^{T})\right]\ ,\end{aligned}$$

where we defined

$$\Pi_{\sigma}^{(\mu\nu)}\equiv d_{\sigma}\ \int \mathrm{d}g\ \chi^{\sigma *}(g)\ U_g^{\mu}\otimes U_g^{\nu *}\ . \tag{6.32}$$

According to Eq. (2.41), $\Pi_{\sigma}^{(\mu\nu)}$ is the projection onto the direct sum of all the subspaces of $\mathcal{H}_{\mu}\otimes\mathcal{H}_{\nu}$ that carry the irreducible representation σ in the tensor product $U_g^{\mu} \otimes U_g^{\nu *}$. Clearly $\Pi_\sigma^{(\mu\nu)}$ is nonzero if and only if the Clebsch-Gordan series of $U_g^{\mu} \otimes U_g^{\nu^*}$ contains σ with nonzero multiplicity $m_\sigma^{(\mu\nu)}$. Notice also that $\mathrm{Tr}[\Pi_\sigma^{(\mu\nu)}] = d_\sigma m_\sigma^{(\mu\nu)}$, by definition of $\Pi_\sigma^{(\mu\nu)}$.

Denoting by $\sum'_{\mu,\nu,\sigma}$ the sum over μ, ν and all σ except σ_0, the average cost can be bounded as follows

$$\begin{aligned}
\langle c \rangle - a_{\sigma_0} \geq & \sum_{\mu,\nu,\sigma}{}' \frac{a_\sigma}{d_\sigma} \left| c_\mu c_\nu \sum_i \mathrm{Tr}\left[\Pi_\sigma^{(\mu\nu)} (\eta_\mu^i \; W_\mu^\dagger \otimes \eta_\nu^{i*} \; W_\nu^T)\right] \right| \\
\geq & \sum_{\mu,\nu,\sigma}{}' \frac{a_\sigma}{d_\sigma} |c_\mu c_\nu| \sqrt{\left(\sum_i \mathrm{Tr}\left[\Pi_\sigma^{(\mu\nu)} (\eta_\mu^i \eta_\mu^{i\dagger} \otimes \mathbb{1}_\nu)\right] \right)} \\
& \times \sqrt{\left(\sum_j \mathrm{Tr}\left[\Pi_\sigma^{(\mu\nu)} (\mathbb{1}_\mu \otimes W_\mu^* \eta_\nu^{jT} \eta_\nu^{j*} W_\mu^T)\right] \right)} \; ,
\end{aligned}$$

since all a_σ are nonpositive. The second inequality follows from Cauchy-Schwartz inequality with respect to the scalar product $\langle \mathbf{A}, \mathbf{B} \rangle \equiv \sum_i \mathrm{Tr}\left[A_i^\dagger B_i\right]$, where we take $A_i^\dagger = \Pi_\sigma^{(\mu\nu)} (\eta_\mu^i W_\mu^\dagger \otimes \mathbb{1}_\nu)$ and $B_i = (\mathbb{1}_\mu \otimes \eta_\nu^{i*} W_\mu^T) \; \Pi_\sigma^{(\mu\nu)}$. Exploiting the relations (6.25) and (6.26), and using that $\mathrm{Tr}\left[\Pi_\sigma^{(\mu\nu)}\right] = d_\sigma m_\sigma^{(\mu\nu)}$, we obtain the bound

$$\begin{aligned}
\langle c \rangle \;\; \geq & \;\; a_{\sigma_0} + \sum_{\mu,\nu,\sigma}{}' a_\sigma \; m_\sigma^{(\mu\nu)} \; |c_\mu c_\nu| \\
\equiv & \;\; \langle c \rangle^{Opt} \; .
\end{aligned} \tag{6.33}$$

It is straightforward to see that the choice of a covariant POVM with $\Xi = |\eta\rangle\langle\eta|$ with $|\eta\rangle$ given by (6.28) achieves this lower bound. ■

6.1.6 Consequences

The general result of Theorem 14 has some remarkable consequences:

- Up to the constant term a_{σ_0}, the minimum cost (6.30) is simply given by the expectation value of the *cost matrix* (6.31) over the normalized vector $\mathbf{v} \equiv (\; |c_\mu| \;)$. Therefore the optimal input state is obtained just by finding the eigenvector corresponding to the minimum eigenvalue of the cost matrix. In other words, the optimal state for the estimation of an unknown parameter is always a superposition of maximally entangled states, with the coefficients in the superposition modulated by the particular choice of the cost function. Notice the simplification of the optimization problem provided by Theorem 14: instead of optimizing a state in the Hilbert space $\mathcal{H} = \bigoplus_{\mu \in \mathsf{S}} \mathcal{H}_\mu \otimes \mathbb{C}^{m_\mu}$ we need only to optimize a vector in $\mathbb{R}^{|\mathsf{S}|}$, where $|\mathsf{S}|$ is the number of irreducible representations contained in the Clebsch-Gordan decomposition of the unitaries $\{U_g\}$.

- The optimal POVM of Theorem 14 is the same optimal POVM arising from the maximum likelihood criterion in Theorem 10. In fact, this criterion corresponds to the particular choice of the delta cost function

$$\begin{aligned} c(\hat{g}, g) &= -\delta(\hat{g}, g) \\ &= -\sum_\sigma d_\sigma \, \chi^\sigma(\hat{g} g^{-1}) \ , \end{aligned}$$

  which is of the form (6.29). In other words, in the case of superpositions of maximally entangled states, the result of Theorem 14 can be viewed as the extension of the maximum likelihood approach of Chapter 4 to arbitrary cost functions.

- In the optimization of covariant POVM's it is often assumed that the operator Ξ corresponding to an optimal estimation can be taken with unit rank. However, this is true for the maximization of the mutual information [31], but not necessarily for the minimization of the Bayes cost. Actually, for mixed states some counterexamples are known [67, 68], and for pure states there is no general proof that the POVM minimizing the average Bayes cost can be chosen with rank one. Therefore, it is important to emphasize that here the rank-one property of the optimal POVM of Theorem 14 is a result of the derivation, not an assumption.

6.2 Applications

6.2.1 Absolute alignment of spatial frames using quantum systems

Suppose that two distant parties (Alice and Bob) have misaligned Cartesian axes, and that they want to communicate in order to align them. If they share a common spatial reference frame, then they can align their axes with arbitrary precision by communicating a sufficiently long string of abstract bits, which describes the orientation of their axes relatively to the common reference frame. However, there are cases in which the common spatial frame is not available, or hard to be accessed, and the only way to communicate is to send real physical systems carrying directional information, such as gyroscopes. We refer to this situation, where no common reference frame is assumed, as the *absolute alignment* of Cartesian frames.

In a quantum communication scenario, Alice can transmit to Bob the directions of her Cartesian axes by sending a set of N spin 1/2 particles prepared in a given state $|A\rangle \in (\mathbb{C}^2)^{\otimes N}$. Since Bob's axes are rotated with respect to Alice's ones, in his reference frame each spin appears rotated by the same unknown rotation, and hence the state $|A\rangle$ appears in Bob's reference frame as $U_g^{\otimes N}|A\rangle$, where U_g is the $\mathbb{SU}(2)$ matrix representing the unknown spatial rotation $g \in \mathbb{SO}(3)$[1]. Estimating the rotation g enables Bob to align his axes with Alice's ones, thus achieving the goal of the absolute alignment problem. Of course, since a finite dimensional Hilbert space does not allow the perfect discrimination of a continuous set of signals, the issue here is to optimize the accuracy of the estimation strategy for a given finite N, by properly choosing Bob's measurement and Alice's input state $|A\rangle$. Moreover, in this problem it is suitable to optimize the estimation strategy without the use of an external reference system, since this allows to establish which is the best precision that can be achieved by transmitting only directional qubits (i.e. spin 1/2 particles), without the need of additional degrees of freedom. Remarkably, as we will show in the following, optimal precision achievable with a reference system asymptotically coincides with the optimal precision achievable without it, i.e. at any rate the reference system is asymptotically useless.

In order to find the optimal strategy for the estimation of the unknown rotation connecting Bob's axes with Alice's ones, we first need to fix our optimality criterion. Here we choose the Bayesian approach with uniform prior $\mathrm{d}g$ (the invariant Haar measure over the group $\mathbb{SU}(2)$) which expresses Bob's complete lack of knowledge about Alice's reference frame. The optimization consists in minimizing the average error

$$\overline{e} = \int \mathrm{d}g_* \int \mathrm{d}g \; p(g|g_*) e(g, g_*) \ , \tag{6.34}$$

[1] A rotation $g \in \mathbb{SO}(3)$ in the three dimensional space is unitarily represented in the Hilbert space of a spin 1/2 particle by a matrix in $U_g \in \mathbb{SU}(2)$. This homomorphism is two-to-one, namely the same rotation $g \in \mathbb{SO}(3)$ is associated to the two matrices $\pm U_g \in \mathbb{SU}(2)$. Accordingly, $\mathbb{SO}(3) \cong \mathbb{SU}(2)/\mathbb{Z}_2$. Of course, the two matrices $\pm U_g$ induce the same transformation on physical states, i.e. $\mathcal{A}_g(\rho) = U_g \rho U_g^\dagger \quad \forall \rho \in \mathcal{S}(\mathcal{H})$. Therefore we can regard the alignment problem as an estimation of $\mathbb{SO}(3)$, or, equivalently as an estimation of $\mathbb{SU}(2)$ with nontrivial stability group $\mathbb{Z}_2$.

where the transmission error $e(\hat{g}, g)$ is the cost function defined as

$$e(\hat{g}, g) \doteq \sum_{\alpha=x,y,z} |g n_\alpha^B - \hat{g} n_\alpha^B|^2 \ , \tag{6.35}$$

n_α^B being the unit vector pointing in the direction of Bob's $\alpha-$axis, for $\alpha = x, y, z$. Such a cost function, which quantifies the deviation between the estimated axes and the true ones, is an element of the generalized Holevo class defined in Par. 6.1.2. In fact, it has the form

$$e(\hat{g}, g) = 6 - 2\chi_1(\hat{g} g^{-1}) \ , \tag{6.36}$$

where $\chi_1(g) \equiv \mathrm{Tr}[U_g^1]$ is the character of $\{U_g^1\}$, the irrep of the rotation group labeled by the quantum number $j = 1$. Notice that $\chi_1(g) = \chi_1(g)^*$ since for $j = 1$ the rotation matrices are real.

In order to apply our general results to this problem we need the Clebsch-Gordan TPS induced by the representation $\{U_g^{\otimes N}\}$ on the Hilbert space $\mathcal{H}^{\otimes N}$. This is given by the decomposition

$$\mathcal{H}^{\otimes N} = \bigoplus_{j=j_0}^{N/2} \mathcal{H}_j \otimes \mathbb{C}^{m_j} \ , \tag{6.37}$$

where j is the quantum number of the total angular momentum, which ranges from $j_0 = 0(1/2)$ to $N/2$ for N even (odd). The dimensions of the representation spaces $\mathcal{H}_j$ are

$$d_j = 2j + 1 \ , \tag{6.38}$$

while the multiplicities are given by

$$m_j^{(N)} = \frac{2j+1}{\frac{N}{2} + j + 1} \binom{N}{\frac{N}{2} + j} \ . \tag{6.39}$$

Notice that the condition $m_j^{(N)} \geq d_j$ is satisfied for any $j < \frac{N}{2}$, while for $j = N/2$ one has $m_{N/2}^{(N)} = 1 < d_{N/2} = N + 1$. Therefore, if the use of an external reference system is not allowed, our general results cannot be directly applied to the alignment problem. However, we can consider the subspaces $\mathcal{K}_N \subset \mathcal{H}^{\otimes N}$ defined as

$$\mathcal{K}_N = \bigoplus_{j=j_0}^{N/2-1} \mathcal{H}_j \otimes \mathbb{C}^{m_j^{(N)}} \ , \tag{6.40}$$

Notice that the space $\mathcal{H}^{\otimes N}$ is isomorphic to a subspace of $\mathcal{K}_{N+2}$, therefore we have the chain of inclusions $\mathcal{K}_{N+2} \supseteq \mathcal{H}^{\otimes N} \supseteq \mathcal{K}_N$. Since the optimization in a larger Hilbert space gives a smaller cost, we have the relation

$$\epsilon_{N+2}^{\min} \leq \overline{e}_N^{\min} \leq \epsilon_N^{\min} \ , \tag{6.41}$$

where $\epsilon_N^{\min}(\overline{e}_N^{\min})$ is the minimum costs achievable with input states in $\mathcal{K}_N(\mathcal{H}^{\otimes N})$. Since in the Hilbert spaces $\mathcal{K}_N$ the relation $m_j^{(N)} \geq d_j$ *is satisfied* for any j, we are now in condition to apply our results and to obtain the minimum cost $\epsilon_N^{\min}$. The asymptotic scaling of the actual error $\overline{e}_N^{\min}$ is then derived by Eq. (6.41). Using Theorem 10 we can immediately calculate the minimum cost $\epsilon_N^{\min}$ by finding the minimum eigenvalue of the $(N/2-1) \times (N/2-1)$ cost matrix C_{ij}, which for odd N is given by

$$C_{ij} = -2(\delta_{i,j+1} + \delta_{i,j-1} - 2\delta_{i,j}) \tag{6.42}$$

whose expression directly follows from Eq. (6.31). The above matrix can be diagonalized in terms of Chebyshev polynomials, and its smallest eigenvalue is $\gamma_N = 4\left[1 - \cos(\frac{2\pi}{N+1})\right] \approx \frac{8\pi^2}{N^2}$. Using Eq. (6.30), we then have the minimum cost $\epsilon_N^{\min} = 6 + \gamma_N = \frac{8\pi^2}{N^2}$, and the same result holds for even N. Finally, using Eq. (6.41) we obtain the asymptotic scaling

$$\overline{e}_N^{\min} = \frac{8\pi^2}{N^2} \ , \tag{6.43}$$

which represents the ultimate precision allowed by quantum mechanics in the absolute alignment of spatial reference frames[2].

[2] An alternative and remarkably simple way to obtain the asymptotic scaling is to consider the cost matrix C_{ij}, acting in the Hilbert space $\mathbb{C}^{N/2-1}$, as the discretized version of the differential operator $C = -2\frac{\mathrm{d}^2}{\mathrm{d}x^2}$, acting in the Hilbert space $L_0^2([0, N/2])$ of square-summable functions with zero boundary conditions. This trick was exploited in Ref. [69] for the problem of optimal phase estimation, which resorts in the same minimization problem arising for $\mathbb{SU}(2)$ [70]. The eigenvectors of the differential operator $C = -2\frac{\mathrm{d}^2}{\mathrm{d}x^2}$ are the functions

$$c^{(k)}(x) = \mathcal{N}_k \ \sin\left(\frac{2\pi k x}{N}\right) \qquad k = \pm 1, \pm 2, \ldots \tag{6.44}$$

where $\mathcal{N}_k$ is a normalization constant, and the eigenvalues are

$$\gamma_N^{(k)} = \frac{8\pi^2 k^2}{N^2} \ . \tag{6.45}$$

Of course, the minimum cost is achieved with the choice $k = \pm 1$.

Notice that, if $\hat{g}$ and g are close, then the transmission error $e(\hat{g}, g)$, as defined in Eq. (6.35) is approximately the sum of the squares of the angles between the estimated and the true directions, i.e.

$$e(\hat{g}, g) \approx \sum_{\alpha=x,y,z} \theta_\alpha^2 \ , \tag{6.46}$$

where θ_α is the angle between the directions $\hat{g}n_\alpha^B$ and gn_α^B. Therefore, since for large N the probability distribution of Bob's outcomes is concentrated in a narrow neighborhood of the true rotation g, we have that the average transmission error asymptotically coincides with the sum of the variances, i.e. $\overline{e}_N^{\min} \approx \sum_{\alpha=x,y,z} \sigma_\alpha^2$. Due to the covariance of the estimating POVM, the variances in the three directions are equal, and we have asymptotically

$$\sigma_\alpha = \sqrt{\frac{8}{3}}\ \frac{\pi}{N} \quad \alpha = x, y, z \ . \tag{6.47}$$

Remarkably, the use of an external reference system does not allow to obtain a better scheme than the one presented before. In fact, the reference system only allows one to satisfy the condition $m_j^{(N)} \geq d_j$ for any $j = j_0, \ldots, N/2$, which is the same condition satisfied in the subspace $\mathcal{K}_{N+2}$ defined in Eq. (6.40). The minimum error in a scheme assisted by the reference system would be then equal to $\epsilon_{N+2}^{\min} \approx 8\pi^2/N^2$, which is exactly the same scaling of Eq. (6.43). Therefore, asymptotically there is no need of this additional resource.

It is also interesting to compare the optimal scaling with the best scaling that can be obtained if the equivalent representations of the rotation group are neglected, i.e. if one restrict himself to considering states of the form

$$|A\rangle = \bigoplus_{j=j_0}^{N/2} a_j \ |\psi_j\rangle|\phi_j\rangle \ , \tag{6.48}$$

where there is no entanglement between the representation and multiplicity spaces. In this case, the best achievable scaling—that for some time was mistakenly considered to be the optimal one [71, 72]—is only $\overline{e}_N^{\min} = \frac{8}{N}$, i.e. in terms of variances

$$\sigma_\alpha = \sqrt{\frac{8}{N}} \quad \alpha = x, y, z \ . \tag{6.49}$$

The difference in the two performances is due to the use of quantum entanglement, whose signature is the typical gain of a factor $\sqrt{N}$ in the variances.

Recently it has been proved by Giovannetti, Lloyd and Maccone that an improvement of order $\sqrt{N}$ is the maximum gain achievable with the use of entanglement in the estimation of the action of a one-parameter group[49, 50]. However, for non-Abelian groups this gain is just an upper bound, which is not proved to be achievable. In the case of $\mathbb{SU}(2)$ rotation, we see here an example where the bound is achieved. The generalization of the results presented in this Subsection to the case of the group $\mathbb{SU}(d)$ with $d \geq 2$ has been recently given by Kahn [73]. Also in this case, using the Theorems 13 and 14 along with a judicious choice of the coefficients c_μ in Eq. (6.16), it is possible to obtain a Bayes expected error of the order $\mathcal{O}(1/N^2)$, corresponding to a variance of order $\mathcal{O}(1/N)$ in the estimated $\mathbb{SU}(d)$ parameters.

The absolute alignment of spatial frames considered in this Subsection can be also achieved *secretly*, i.e. Alice can communicate to Bob the orientation of her Cartesian axes in such a way that no eavesdropper Eve can gain information about it. This can be done by randomizing the choice of the input state according to a random sequence of bits, secretly shared between Alice and Bob. With a proper choice of the encoding, Bob can infer the direction of Alice's axes with the optimal precision of Eq. (6.43), while Eve cannot access any directional information [74]. Remarkably, the number of secret bits needed in this protocol is asymptotically $3 \log N$, which is exactly the secret classical capacity associated to the presence of a privately shared reference frame [5].

6.2.2 Estimation of an unknown maximally entangled state

Maximally entangled states are a fundamental resource for quantum teleportation [57] and for quantum cryptography [58]. To achieve ideal teleportation, Alice and Bob must know with precision which maximally entangled state they are sharing, otherwise the fidelity of the state received by Bob with the original state from Alice can be significantly lowered. Similar arguments apply to the cryptographic schemes where the correlations arising from entanglement are exploited to generate a secret key.

Here we consider the problem of estimating in the best way a completely unknown maximally entangled state, provided that N identical copies are available. This is done as an application of Theorem 14. Let us consider a state $|\psi\rangle\rangle \in \mathcal{H} \otimes \mathcal{H}$, with $\dim(\mathcal{H}) = d$. In terms of the notation (6.13), this state is maximally entangled if and only if $\psi = \frac{1}{\sqrt{d}}\, U$, where U is some unitary operator. Using property (6.15), any maximally entangled state can be written as

$$|\psi_g\rangle\!\rangle = \frac{1}{\sqrt{d}} \left(U_g \otimes 1\!\!1\right) |1\!\!1\rangle\!\rangle \ , \tag{6.50}$$

where U_g is an element of the group $\mathbb{SU}(d)$.

If N identical copies of the unknown state $|\psi_g\rangle\!\rangle$ are given, then the problem becomes to find the best estimate for parameter g encoded into the states of the form $|\Psi_g\rangle\!\rangle = |\psi_g\rangle\!\rangle^{\otimes N}$. Optimality is defined here as the maximization of the Uhlmann fidelity between the true state and the estimated one:

$$f(\hat{g}, g) = |\langle\!\langle \psi_g | \psi_{\hat{g}} \rangle\!\rangle|^2 \ . \tag{6.51}$$

Using the definition (6.50) and the property (6.14), we obtain

$$f(\hat{g}, g) = \frac{1}{d^2} \ |\chi(\hat{g} g^{-1})|^2 \ . \tag{6.52}$$

where $\chi(g) = \mathrm{Tr}[U_g]$. The maximization of the fidelity corresponds to the minimization of the cost function

$$c(\hat{g}, g) = 1 - f(\hat{g}, g) \ , \tag{6.53}$$

which is of the form (6.29). In particular, for $d = 2$, $|\chi(g)|^2 = 1 + \chi^1(g)$, where $\chi^1(g) = \mathrm{Tr}[U_g^1]$ is the character of the irreducible representation of $\mathbb{SU}(2)$ with angular momentum $j = 1$,whence we have

$$c(\hat{g}, g) = \frac{1}{4} \ \left(3 - \chi^1(g^{-1}\hat{g})\right) \ . \tag{6.54}$$

All the states of the form $|\Psi_g\rangle = |\psi_g\rangle\!\rangle^{\otimes N}$ are generated from the input state

$$|\Psi\rangle = \frac{1}{\sqrt{d^N}} \ |1\!\!1\rangle\!\rangle^{\otimes N} \tag{6.55}$$

by the action of the representation $\{(U_g \otimes 1\!\!1)^{\otimes N} \mid U_g \in \mathbb{SU}(d)\}$.

The following Lemma, whose proof can be found in the Appendix of Ref. [18], provides the decomposition of the input state in the Clebsch-Gordan TPS

Lemma 10. *Using suitable bases for the multiplicity spaces in decomposition (2.20), the input state (6.55) can be written as*

$$|\Psi\rangle = \bigoplus_{\mu \in \mathsf{S}} \ \frac{c_\mu}{\sqrt{d_\mu}} \ |1\!\!1_\mu\rangle\!\rangle \ , \tag{6.56}$$

where the sum runs over the irreducible representations of $\mathbb{SU}(d)$ occurring in the Clebsch-Gordan series of $\{U_g^{\otimes N}\}$, and

$$c_\mu = \sqrt{\frac{d_\mu m_\mu}{d^N}} \ , \tag{6.57}$$

d_μ and m_μ being respectively the dimension and the multiplicity of the representation $\mu \in \mathsf{S}$.

Thank to this Lemma we can exploit directly the result of Theorem 14 to get the optimal POVM and to calculate the average fidelity. We carry on the calculation of the optimal fidelity in the simplest case $d = 2$. As usual, the irreps of $\mathbb{SU}(2)$ contained in the representation $\{U_g^{\otimes N}\}$ are labeled by the quantum number j, ranging from $j_0 = 0(\frac{1}{2})$ to $\frac{N}{2}$ for N being even (odd), respectively. The minimum cost can be evaluated using Theorem 14 as

$$\langle c\rangle^{Opt} = \frac{3}{4} + \sum_{i,j=j_0}^{\frac{N}{2}} |c_i| \ C_{ij} \ |c_j| \tag{6.58}$$

Using Eq. (6.57) with the values of dimensions and multiplicities given by Eqs. (6.38) and (6.39), the coefficients of the state become

$$c_i = g(i) \sqrt{\frac{1}{2^N} \binom{N}{\frac{N}{2}+i}} \ , \tag{6.59}$$

where

$$g(i) = \frac{2i+1}{\sqrt{\frac{N}{2}+i+1}} \ . \tag{6.60}$$

On the other hand, the matrix C_{ij} is calculated according to the definition (6.31), namely by evaluating the multiplicity of the representation with angular momentum $k = 1$ in the Clebsch-Gordan series of the tensor product $U_g^i \otimes U_g^{j*}$. In this way we get for odd N

$$C_{ij} \ = \ \frac{1}{4}(2\delta_{i,j} - \delta_{i,j+1} - \delta_{i,j-1}) \ . \tag{6.61}$$

The case of even N differs only for the border term with $j = 0$, but the asymptotics is exactly the same as for N odd. Since $\sum_i |c_i|^2 = 1$, we have

$$\langle c\rangle = \frac{1}{2}\left(1 - \sum_{j=j_0}^{\frac{N}{2}-1} c_j c_{j+1}\right) \ . \tag{6.62}$$

To obtain the asymptotic behavior of the optimal fidelity, we can approximate the binomial distribution in (6.59) with a Gaussian $G_\sigma(x)$ with mean $\bar{x}=0$ and variance $\sigma^2=\frac{N}{4}$. Since the sum in (6.62) runs over a large interval with respect to σ, we can also approximate it with an integral over $[0,+\infty]$. All these approximations hold up to order higher than $\frac{1}{N}$. Thus the evaluation of the optimal fidelity is reduced to the evaluation of the integral

$$I=\int_0^\infty \mathrm{d}x\ g(x)g(x+1)\ G_\sigma(x)\ , \tag{6.63}$$

whose leading order can be obtained from Taylor expansion. In this way, we derive the asymptotic cost

$$\langle c\rangle^{Opt}=\frac{3}{4N}\ , \tag{6.64}$$

corresponding to the optimal fidelity

$$\langle f\rangle^{Opt}=1-\frac{3}{4N}\ . \tag{6.65}$$

Remarkably, the Bayes cost with uniform a priori distribution has the same asymptotic behavior of the cost of the optimal locally unbiased estimator obtained in [61, 75], for any possible value g of the true parameter. This means that in the present unbiased case the covariant measurement of Theorem 14 is optimal not only on average but also pointwise.

6.2.3 Estimation of a squeezing parameter

In the following, we will present the optimal estimation of an unknown squeezing transformation in a given direction, acting on an arbitrary state of the radiation field. This problem arises in the experimental situation where a degenerate parametric amplifier is pumped by a strong coherent field with a fixed phase relation with the state to be amplified, and one is interested in optimally characterizing the amplifier gain.

Consider a single-mode radiation field with bosonic operators a and $a^\dagger$, satisfying the canonical commutation relations $[a,a^\dagger]=1$. The squeezing transformation is defined as follows

$$S(r)=e^{-irK}\ ,\qquad K=i\,\frac{a^{\dagger 2}-a^2}{2} \tag{6.66}$$

where r is a real parameter. In the quadrature representation $\psi(x)=\langle x|\psi\rangle$, where $|x\rangle$ denotes the Dirac-normalized eigenvector of the quadrature operator $X=(a+a^\dagger)/2$, the effect of squeezing on the wavefunction is given by

$$\psi(x)\longrightarrow e^{-r/2}\psi(e^{-r}x)\ . \tag{6.67}$$

The estimation of an unknown squeezing is actually an example of estimation of an element of the additive group $\mathbf{G} = \mathbb{R}$, which has been studied in many variants in the literature [76] (see also p. 188 of Holevo's book [2]). Notice that, since the group $\mathbf{G} = \mathbb{R}$ is noncompact and its irreps form a continuous set, this example is not rigorously covered by the results presented in this Chapter. However, in the following we will see that it is possible to formally extend the validity of the results, thus obtaining the best possible estimation of a squeezing parameter.

Optimality is defined in terms of a *cost function* $c(\hat{r} - r)$, that quantifies the cost of estimating $\hat{r}$ when the true value is r. Since the group of squeezing transformations is noncompact, we cannot adopt the Bayesian approach with uniform prior, namely the uniform measure on the real line cannot be a probability density, being non-normalizable. Therefore, we choose the minimax approach, and define the optimal POVM as the one that minimizes the worst-case cost

$$c_{\max} = \sup_{r \in \mathbb{R}} \left\{ \int_{-\infty}^{+\infty} d\hat{r} \; p(\hat{r}|r) c(\hat{r} - r) \right\} \ , \tag{6.68}$$

where $p(\hat{r}|r)$ is the conditional probability density of $\hat{r}$ given the true value r. Then, formally extending the definition of generalized Holevo class, we consider cost functions with the Fourier expansion[3]

$$c(r) = \int_{-\infty}^{+\infty} d\mu \; a_\mu e^{i\mu r} \ , \qquad a_\mu \leq 0 \quad \forall \mu \neq 0 \ , \tag{6.69}$$

where $e^{i\mu r}$ are the characters of the unidimensional representations of the additive group $\mathbf{G} = \mathbb{R}$. Again, this class contains a large number of optimality criteria, such as the maximum likelihood

$$c_{ML}(r) = -\delta(r) = -\frac{1}{2\pi} \int_{-\infty}^{+\infty} \mathrm{d}\mu \; e^{i\mu r} \ , \tag{6.70}$$

and the maximum fidelity $c_{MF}(r) = 1 - |\langle \psi | S(r) | \psi \rangle|^2$.

In order to obtain the optimal measurement we need to find the decomposition of the Hilbert space induced by the Clebsch-Gordan decomposition of the unitaries $\{S(r) = e^{-irK}\}$, a rather simple task since for Abelian groups

[3] Since the irreps form a continuous set, one could object that the definition of the Fourier coefficients in a zero measure set (the point $\mu = 0$) is irrelevant. This definition has to be interpreted in the distributional sense, namely we allow singular Fourier coefficients such as the delta distribution $a_\mu = \delta(\mu)$, corresponding to the constant offset $f(r) = \int_{-\infty}^{+\infty} \mathrm{d}\mu \; k\delta(\mu) \; e^{i\mu r} \equiv 1$.

the irreducible subspaces are the eigenspaces of the generator. The spectrum of the generator K is the whole real line, and the eigenvalue equation is

$$K|\mu, s\rangle = \mu|\mu, s\rangle \ , \tag{6.71}$$

where $\mu \in \mathbb{R}$ is the eigenvalue, and s is a degeneracy index with two possible values ± 1. The explicit expression of the generalized eigenvectors of K in the quadrature representation is given by [77]

$$\langle x|\mu, s\rangle = \frac{1}{\sqrt{2\pi}} |x|^{i\mu - \frac{1}{2}} \ \theta(sx) \ , \tag{6.72}$$

where $\theta(x)$ is the Heaviside step-function [$\theta(x) = 1$ for $x > 0$, $\theta(x) = 0$ for $x < 0$]. The vectors $|\mu, s\rangle$ are orthogonal in the Dirac sense, namely $\langle \mu, r|\nu, s\rangle = \delta_{rs} \ \delta(\mu - \nu)$, and provide the resolution of the identity

$$\int_{-\infty}^{+\infty} d\mu \ \Pi_\mu = 1\!\!1 \ , \tag{6.73}$$

where $\Pi_\mu = \sum_{s=\pm 1} \ |\mu, s\rangle\langle \mu, s|$ is the projector onto the Dirac-eigenspace of K corresponding to the eigenvalue μ. Accordingly, the Hilbert space $\mathcal{H}$ can be decomposed as a *direct integral*

$$\mathcal{H} = \int_{-\infty}^{+\infty} d\mu \ \mathcal{M}_\mu \ , \tag{6.74}$$

which is the analogue of the Clebsch-Gordan TPS of Eq (2.20) in the case of a continuous set of irreps, with one-dimensional representation spaces $\mathcal{H}_\mu$ and two-dimensional multiplicity spaces $\mathcal{M}_\mu$. More precisely, here the Hilbert space $\mathcal{M}_\mu$ is the two-dimensional vector space spanned by the vectors $|\mu, \pm 1\rangle$, and equipped with the scalar product

$$\langle v_\mu|w_\mu\rangle = \sum_{s=\pm 1} v_s^{\mu *} w_s^\mu \tag{6.75}$$

for $|v_\mu\rangle = \sum_{s=\pm 1} v_s^\mu |\mu, s\rangle$, $|w_\mu\rangle = \sum_{s=\pm 1} w_s^\mu |\mu, s\rangle$. Accordingly, the norm in $\mathcal{M}_\mu$ is $||v_\mu|| = \sqrt{\langle v_\mu|v_\mu\rangle}$.

Using the completeness relation (6.73), we can write any pure state $|\psi\rangle \in \mathcal{H}$ as

$$|\psi\rangle = \int_{-\infty}^{+\infty} d\mu \ c_\mu |\psi_\mu\rangle \ , \tag{6.76}$$

where $c_\mu = ||\Pi_\mu|\psi\rangle||$, and

$$|\psi_\mu\rangle = \frac{\Pi_\mu|\psi\rangle}{||\Pi_\mu|\psi\rangle||} \tag{6.77}$$

is the normalized projection of $|\psi\rangle$ onto $\mathcal{M}_\mu$. This is the continuous version of the decomposition of Eq. (6.27), where the maximally entangled states $|W_\mu\rangle\!\rangle/\sqrt{d_\mu}$ are replaced by the vectors $|\psi_\mu\rangle$. Since the irreps of the additive group are one-dimensional, here there is no entanglement between representation and multiplicity spaces. Moreover, as stated in Corollary 4, a reference system is of no use in this problem.

The last ingredient needed to write down the optimal POVM of Theorem 14 in the present case is the formal analogue of the dimension d_μ in the group average of Eq. (2.32). This is given by the following

Proposition 17. *Let O be an operator on $\mathcal{H}$. Then, its group average $\overline{O} = \int_{-\infty}^{+\infty} \mathrm{d}r \ S(r) O S^\dagger(r)$ is given by*

$$\overline{O} = \int_{-\infty}^{+\infty} \mathrm{d}\mu \ \frac{\Pi_\mu O \Pi_\mu}{2\pi} \ . \tag{6.78}$$

Comparing the above relation with Eq. (2.32), we then obtain the value of the formal dimension $d_\mu = 1/(2\pi)$.

We are now in condition to exploit Theorem 14, thus getting the optimal covariant POVM in the form

$$P(\mathrm{d}r) = S(r)|\eta\rangle\langle\eta|S^\dagger(r) \ \mathrm{d}r \ , \tag{6.79}$$

where

$$|\eta\rangle = \int_{-\infty}^{+\infty} \frac{d\mu}{\sqrt{2\pi}} \, |\psi_\mu\rangle \ . \tag{6.80}$$

Such a POVM is optimal for any cost function of the form (6.69), and in particular, for the maximum likelihood criterion. Notice the correspondence of the vectors

$$|\eta(r)\rangle \doteq S(r)|\eta\rangle \tag{6.81}$$

with the Susskind-Glogower vectors[78]

$$|e(\varphi)\rangle \doteq \sum_{n=0}^{\infty} \frac{e^{in\varphi}}{\sqrt{2\pi}} |n\rangle \tag{6.82}$$

that arise in the context of optimal phase estimation (here $|n\rangle$ are the non-degenerate eigenvectors of the photon number operator $a^\dagger a$). The vectors $|\eta(r)\rangle$ are orthogonal in the Dirac sense, namely the optimal POVM is a von Neumann measurement. The projection $|\psi_\mu\rangle$ of Eq. (6.77), contained in the expression (6.80) makes the optimal measurement depend on the input state

$|\psi\rangle$. Accordingly, one obtains different non-commuting observables for the measurement of the parameter r, corresponding to different input states.

The optimal measurement (6.79) can be compared with that given in Ref. [79], which is described in our notation by the POVM

$$\widetilde{P}(\mathrm{d}r) = \sum_{s=\pm 1} |\eta_s(r)\rangle\langle\eta_s(r)|\ \mathrm{d}r\ , \tag{6.83}$$

where

$$|\eta_s(r)\rangle = \int_{-\infty}^{+\infty} \frac{d\mu}{\sqrt{2\pi}}\ e^{-ir\mu}|\mu, s\rangle\ . \tag{6.84}$$

Using Eq. (6.72), it is easy to see that $|\eta_\pm(r)\rangle$ are eigenvectors of the quadrature X corresponding to the eigenvalues $\pm e^r$, and hence the POVM (6.83) corresponds to measuring the observable $\ln|X|$, *independently of the input state.* It is simple to realize that this estimation scheme is not optimal, as there are input states for which the observable $\ln|X|$ gives a very inaccurate estimation compared with the optimal one.

In the rest of this Subsection, we will examine the performances of the optimal estimation strategy of Eq. (6.79) for particular classes of input states, such as coherent and displaced squeezed states. Using Eqs. (6.76) and (6.77) the optimal probability distribution for an input state $|\psi\rangle$ is given by

$$\begin{aligned} p(\hat{r}|r)\mathrm{d}\hat{r} &= \langle\psi|S(r)^\dagger\ P(\mathrm{d}\hat{r})\ S(r)|\psi\rangle \\ &= |\langle\eta|\ S(r-\hat{r})\ |\psi\rangle|^2\,\mathrm{d}\hat{r} \\ &= \frac{1}{2\pi}\left|\int_{-\infty}^{+\infty} d\mu\, e^{i(\hat{r}-r)\mu}\sqrt{\langle\psi|\Pi_\mu|\psi\rangle}\right|^2 \mathrm{d}\hat{r}\ . \end{aligned} \tag{6.85}$$

Moreover, since the probability distribution depends only on the difference $\hat{r}-r$, from now on we will write $p(\hat{r}-r)$ instead of $p(\hat{r}|r)$. Representing the projection Π_μ as $\Pi_\mu = \int_{-\infty}^{+\infty}\frac{d\lambda}{2\pi}\, e^{i\lambda(\mu-K)}$, the probability distribution of Eq. (6.85) can be rewritten as

$$p(r) = \left|\int_{-\infty}^{+\infty}\frac{d\mu}{2\pi}\, e^{ir\mu}\sqrt{\int_{-\infty}^{+\infty} d\lambda\, e^{i\lambda\mu}\,\langle\psi|S(\lambda)|\psi\rangle}\right|^2\ . \tag{6.86}$$

- **Coherent states.** In the case of a coherent input state $|\alpha\rangle$, the probability distribution (6.86) can be specified as follows

$$p(r) = e^{-|\alpha|^2}\left|\int_{-\infty}^{+\infty}\frac{d\mu}{2\pi}\, e^{-i\mu r}\ \times \right. \tag{6.87}$$

$$\left.\sqrt{\int_{-\infty}^{+\infty}\frac{d\lambda}{\sqrt{\cosh\lambda}}\, e^{i\lambda\mu}\, e^{\frac{1}{2}\tanh\lambda(\alpha^{*2}-\alpha^2)}\, e^{\frac{|\alpha|^2}{\cosh\lambda}}}\ \right|^2\ .$$

In particular, in the case of the vacuum state ($\alpha = 0$) the optimal estimation can be compared with the scheme of Ref. [79]. The two probability distributions are plotted in Fig. (6.1). One can notice that the probability distribution of Eq. (6.87) is more concentrated in a neighborhood the true value, and there is no bias, i.e. the mean value coincides with the true one. Notice also that the true value is also the most likely one, as a consequence of the fact that the estimation strategy is optimal also for the maximum likelihood criterion[4].

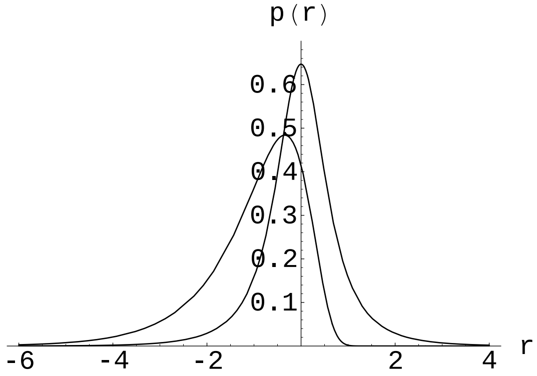

Figure 6.1: Probability distributions for the estimation of squeezing on a vacuum input state. The asymmetric distribution comes from the suboptimal measurement of Ref. [79] in Eq. (6.83). The symmetric distribution corresponds to the optimal measurement of Eq. (6.79).

The precision of the estimation can be enhanced by increasing the average photon number of the input state $|\alpha\rangle$. Numerically, the probability distribution (6.87) has been plotted for increasing real values of α in Fig. 6.2, where one can easily observe the corresponding improvement in the estimation.

For large values of $|\alpha|$, from Eq. (6.87) one obtains asymptotically the Gaussian distribution

$$p(r) = \sqrt{\frac{2|\alpha|^2}{\pi}}\, e^{-2|\alpha|^2 r^2} \ , \tag{6.88}$$

[4] The maximum likelihood approach selects the measurement that makes the probability density at the true value as high as possible. Conversely, in the case of unimodular groups, it is possible to prove (see Proposition 19 of Chapter 7) that the probability distribution resulting from the maximum likelihood criterion achieves its maximum value in correspondence with the true value.

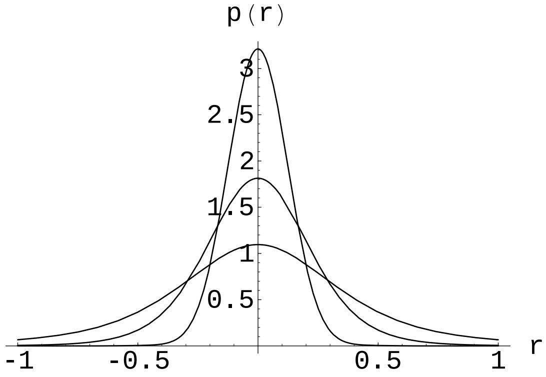

Figure 6.2: Optimal probability distribution of squeezing for input coherent states. The distribution becomes sharper for increasing values of the coherent-state amplitude ($\alpha = 1$, 2, 4.)

that provides a r.m.s error on the estimation of r as

$$\Delta r = \frac{1}{2\sqrt{\bar{n}}} \ , \tag{6.89}$$

where $\bar{n} = |\alpha|^2$ is the mean photon number. Notice that, for real α the optimal estimation is obtained asymptotically by measuring the quadrature X and estimating $r = \ln|x/\alpha|$ in correspondence with the outcome x. In fact, the probability distribution of the quadrature X in a coherent state with real α is

$$\begin{aligned} |\langle x|\alpha\rangle|^2 \mathrm{d}x &= \sqrt{\frac{2}{\pi}}\ e^{-2(x-\alpha)^2}\ \mathrm{d}x & (6.90)\\ &= \sqrt{\frac{2}{\pi}}\ e^{-2|\alpha|^2(1-\frac{x}{\alpha})^2}\ \mathrm{d}x & (6.91)\\ &= \sqrt{\frac{2|\alpha|^2}{\pi}}\ e^{-2|\alpha|^2(1-e^r)^2}\ e^r \mathrm{d}r & (6.92)\\ &\approx \sqrt{\frac{2|\alpha|^2}{\pi}}\ e^{-2|\alpha|^2 r^2}\ \mathrm{d}r \ , & (6.93)\end{aligned}$$

whence $r = \ln|x/\alpha|$ is distributed according to the optimal probability distribution. On the other hand, if α is purely imaginary, one achieves asymptotically the optimal estimation by measuring the quadrature Y—instead of X—and estimating $r = \ln|y/\alpha|$. As we already stressed, the optimal measurements do depend strongly on the input state. Contrarily, the scheme of Ref. [79], which consists in measuring the observable $\ln|X|$ independently of the input state, gives a r.m.s. error which for purely imaginary α takes the constant value $\Delta r = 1,1107$ independently of $|\alpha|$.

- **Displaced squeezed states.** The asymptotic performance of the optimal estimation using coherent input states, given by Eq. (6.89), can be improved to
$$\Delta r = \frac{1}{2\bar{n}} \tag{6.94}$$
by using displaced squeezed states $|\alpha, z\rangle = D(\alpha)S(z)|0\rangle$, with $\alpha, z \in \mathbb{R}$. In fact, from the relation
$$D(\alpha)S(z) = S(z)D(\alpha e^{-z}) \ , \tag{6.95}$$
the probability distribution $p(r)$ is given by Eq. (6.87) just by replacing α with αe^{-z}. In particular, for large αe^{-z}, one has the Gaussian
$$p(r) = \sqrt{\frac{2(\alpha e^{-z})^2}{\pi}} \, e^{-2(\alpha e^{-z})^2 \ r^2} \ . \tag{6.96}$$
Accordingly, the uncertainty in the estimation of a squeezing on the input state $|\alpha, z\rangle$ is
$$\Delta r = \frac{1}{2|\alpha|e^{-z}} \ . \tag{6.97}$$
In the asymptotic limit of large number of photons $\bar{n} = |\alpha|^2 + \sinh^2 z$, the minimization of the r.m.s. gives the optimal scaling $\Delta r = 1/(2\bar{n})$, for $\alpha = \sqrt{\bar{n}/2}$ and $z = -1/2 \ln(2\bar{n})$. This corresponds to approximate the eigenvectors of the quadrature operator X. The optimal performance in the asymptotic regime can be achieved by measuring the quadrature X and estimating $r = \ln|x/\alpha|$, in correspondence to the outcome x, i.e. by measuring the observable $\ln|X/\alpha|$ as in Ref. [79].

  The same results can be obtained also by using displaced squeezed states with purely imaginary displacement, i.e. states of the form $|i\alpha, z\rangle = D(i\alpha)S(z)|0\rangle$ with $\alpha, z \in \mathbb{R}$. In this case, the relation
$$D(i\alpha)S(z) = S(z)D(i\alpha e^{z}) \ , \tag{6.98}$$
allows one to obtain the optimal probability distribution from Eq. (6.87) with the substitution $\alpha \to i\alpha e^z$. Making the same substitution in Eq. (6.88) one gets the asymptotic expression
$$p(r) = \sqrt{\frac{2(\alpha e^{z})^2}{\pi}} \, e^{-2(\alpha e^{z})^2 \ r^2} \ . \tag{6.99}$$

Accordingly, the uncertainty in the estimation is

$$\Delta r = \frac{1}{2|\alpha|e^z} \ . \tag{6.100}$$

Then, taking $\alpha = \sqrt{\bar{n}/2}$ and $z = 1/2 \ln(2\bar{n})$ one achieves the optimal scaling $\Delta r = 1/(2\bar{n})$. The corresponding input states are approximate eigenvectors of the conjugate quadrature Y, and asymptotically the optimal estimation scheme is to measure Y and to estimate $r = -\ln|y/\alpha|$ in correspondence with the outcome y, i.e. to measure the observable $\ln|Y/\alpha|$. Again, it is interesting to note that the different input states $|\alpha, z\rangle$ and $|i\alpha, z\rangle$ require different non-commuting observables for the optimal estimation, even though we want to estimate the same parameter.

Chapter 7

Quantum estimation with nonunimodular groups

The estimation of a real squeezing and the estimation of a real displacement in the radiation field are two rather simple problems involving a one-parameter group isomorphic to the real line. However, if we want to estimate both parameters jointly the problem becomes highly nontrivial, being an example of joint measurement of two noncommuting observables, similar to the case of position and momentum. Jointly estimating a squeezing and a displacement is equivalent to estimating an element of the affine group "$ax+b$" of translations and dilations on the real line, which is the typical example of a nonunimodular group, i.e. of a group where the left-invariant Haar measure is different from the right-invariant one. In this Chapter we derive the maximum likelihood measurements for the estimation of nonunimodular group transformations, finding some unexpected features that never appeared in the estimation problems considered so far. In the concrete example of squeezing and displacement, the presented results allow one to find a remarkable relation between the uncertainties in the joint measurement and the uncertainties in the two optimal separate measurements.

7.1 Basic notions about nonunimodular groups

7.1.1 Generalities

In the following we consider the general class of locally compact groups. It is well known that any such group $\mathbf{G}$ admits a left-invariant and a right-invariant Haar measure, which are both unique up to a constant[12]. Here

$\mathrm{d}_L(g)$ denotes the left-invariant measure:

$$\mathrm{d}_L(hg) = \mathrm{d}_L(g) \qquad \forall h, g \in \mathbf{G} \ , \tag{7.1}$$

and denotes $\mathrm{d}_R(g)$ the right-invariant measure:

$$\mathrm{d}_R(gh) = \mathrm{d}_R(g) \qquad \forall g, h \in \mathbf{G} \ . \tag{7.2}$$

In general, the two invariant measures may differ, and the relation

$$\mathrm{d}_L(g) = \Delta(g) \ \mathrm{d}_R(g) \tag{7.3}$$

defines a positive function $\Delta(g) > 0$, called the *modular function*. Since the left- and right-invariant measures are unique up to a constant, one can always choose the constant in order to have

$$\Delta(e) = 1 \ . \tag{7.4}$$

With this normalization, the modular function enjoys the property

$$\Delta(gh) = \Delta(g)\Delta(h) \qquad \forall g, h \in \mathbf{G} \ , \tag{7.5}$$

and one also has

$$\mathrm{d}_L(g^{-1}) = \mathrm{d}_R(g) \ . \tag{7.6}$$

If the modular function is constant, i.e. $\Delta(g) \equiv 1 \quad \forall g$, the group $\mathbf{G}$ is called unimodular. All the examples considered so far felt in this class, but now we will proceed further.

7.1.2 Orthogonality relations

A clear presentation about the orthogonality relations for nonunimodular groups is given in Ref. [37]. The definition of square-summable irrep, given in Sec. 2.6 for unimodular groups, can be extended to nonunimodular groups in a straightforward way:

Definition 15. *Let* $\mathbf{G}$ *be a locally compact Lie group. A projective irrep* $\{U_g\}$ *acting in the Hilbert space* $\mathcal{H}$ *is called* square-summable *if there is a non-zero vector* $|\psi\rangle \in \mathcal{H}$ *such that*

$$\int_{\mathbf{G}} \mathrm{d}_L g \ |\langle\psi|U_g|\psi\rangle|^2 < \infty \ . \tag{7.7}$$

A vector $|\psi\rangle$ satisfying the property (7.7) is called *admissible.*

Given an operator $O \in \mathcal{B}(\mathcal{H})$ and an irrep $\{U_g\}$, one can consider the group average either with respect to the left-invariant measure

$$\overline{O} = \int_{\mathbf{G}} \mathrm{d}_L g \; U_g O U_g^\dagger \ , \tag{7.8}$$

or with respect to the right-invariant one

$$\overline{O}_R = \int_{\mathbf{G}} \mathrm{d}_R g \; U_g O U_g^\dagger \ . \tag{7.9}$$

If the integral of Eq. (7.8) converges, the left-average $\overline{O}$ commutes with the irreducible representation, and hence it must be proportional to the identity. Contrarily, the right-average $\overline{O}_R$ does not commute with the unitaries $\{U_g\}$, namely is not proportional to the identity. This fact is reflected in the following Theorem by Duflo, Moore, and Carey:

Theorem 15 ([80, 81]). *Let* $\mathbf{G}$ *a locally compact group and* $\{U_g\}$ *a square-summable irrep. Then, the group average* $\overline{O}$ *of a given operator is*

$$\overline{O} = \mathrm{Tr}[O D^{-1}] \ \mathbb{1} \tag{7.10}$$

where D *is the* formal degree operator*, i.e. the unique positive self-adjoint operator defined by*

$$D^{-1} = \int_{\mathbf{G}} \mathrm{d}_R g \; U_g |\phi\rangle\langle\phi| U_g^\dagger \ , \tag{7.11}$$

$|\phi\rangle$ *being any normalized state in* $\mathcal{H}$*.*

In the following we will refer to the formal degree operator D as to the *Duflo-Moore-Carey operator (DMC).* The definition of the DMC operator in the above Theorem makes evident its analogy with the formal dimension in the case of unimodular groups. In fact, for unimodular groups the group average formula (7.10) holds with $D^{-1} = \mathbb{1}/d$, where the number d is the formal dimension defined in Eq. (2.35).

Using Theorem 15, one can evaluate the integral in Eq. (7.7):

$$\int_{\mathbf{G}} \mathrm{d}_L g \ |\langle\psi|U_g|\psi\rangle|^2 = \langle\psi|D^{-1}|\psi\rangle \ , \tag{7.12}$$

thus observing that a vector $|\psi\rangle \in \mathcal{H}$ is admissible if and only if it is in the domain of D^{-1}. Analogously, we can call *admissible* the operators satisfying $\mathrm{Tr}[OD^{-1}] < \infty$, i.e. the operators for which the group average $\overline{O}$ converges.

To conclude this Section, we give the immediate generalization of Theorem 15 to the case of a reducible representations $\{U_g\}$ that are a direct sum of a countable number of square-summable irreps. This is the main formula that will be useful for applications.

Proposition 18. *Let* $\mathbf{G}$ *a locally compact group and* $\{U_g \mid g \in \mathbf{G}\}$ *a projective representation acting in the Hilbert space* $\mathcal{H}$*, with Clebsch-Gordan decomposition* $U_g = \bigoplus_{\mu \in \mathsf{S}} U_g^\mu \otimes \mathbb{1}_{\mathcal{M}_\mu}$*. Let*

$$\mathcal{H} = \bigoplus_{\mu \in \mathsf{S}} \mathcal{H}_\mu \otimes \mathcal{M}_\mu \tag{7.13}$$

be the Clebsch-Gordan TPS. Then, the group average $\overline{O}$ *of the operator* $O \in \mathcal{B}(\mathcal{H})$ *is given by*

$$\overline{O} = \bigoplus_{\mu \in \mathsf{S}} \mathbb{1}_{\mathcal{H}_\mu} \otimes \mathrm{Tr}_{\mathcal{H}_\mu}[O\ D_\mu^{-1} \otimes \mathbb{1}_{\mathcal{M}_\mu}]\ , \tag{7.14}$$

where $\mathrm{Tr}_{\mathcal{H}_\mu}$ *denotes the trace over the first factor in the tensor product* $\mathcal{H}_\mu \otimes \mathcal{M}_\mu$*, and* D_μ *is the formal degree operator defined as*

$$D_\mu^{-1} = \int_{\mathbf{G}} \mathrm{d}_R g\ U_g |\psi_\mu\rangle\langle\psi_\mu| U_g^\dagger\ , \tag{7.15}$$

$|\psi_\mu\rangle$ *being any normalized vector in* $\mathcal{H}_\mu$*.*

7.2 Maximum likelihood strategy

Here we derive the best estimate of the signal $g \in \mathbf{G}$ encoded in the states $\{U_g|\Psi\rangle\}$, where $|\Psi\rangle \in \mathcal{H}$ is the input state and $\{U_g\}$ is a projective representation with discrete Clebsch-Gordan series. Optimality is defined in the minimax sense, according to the maximum likelihood criterion. The representation $\{U_g\}$ will be a projective representation with discrete Clebsch-Gordan decomposition $U_g = \bigoplus_{\mu \in \mathsf{S}} U_g^\mu \otimes \mathbb{1}_{\mathcal{M}_\mu}$.

Since optimality is defined in the minimax sense, with an invariant cost function, we can consider without loss of generality a covariant POVM

$$P(\mathrm{d}_L g) = U_g \Xi U_g^\dagger\ \mathrm{d}_L g\ . \tag{7.16}$$

The possibility of introducing a density for POVMs that are covariant under the action of a generally nonunimodilar group has been studied in Ref. [82]. The normalization condition

$$\int_{\mathbf{G}} P(\mathrm{d}_L g) = \mathbb{1} \tag{7.17}$$

can be rewritten as

$$\overline{\Xi} = \mathbb{1} \ . \tag{7.18}$$

Then, using Eq. (7.14) for evaluating the group average $\overline{\Xi}$, this condition becomes

$$\mathrm{Tr}_{\mathcal{H}_\mu}[D_\mu \otimes \mathbb{1}_{\mathcal{M}_\mu} \ \Xi] = \mathbb{1}_{\mathcal{M}_\mu} \ . \tag{7.19}$$

According to the maximum likelihood approach, we need to find the covariant measurement that maximizes the probability density that the estimated transformation coincides with the true one, i.e.

$$p(g|g) \ d_L g = \mathrm{Tr}[P(\mathrm{d}_L g) U_g \rho U_g^\dagger] = \mathrm{Tr}[\Xi \rho] \ \mathrm{d}_L g \ . \tag{7.20}$$

For a pure input state $\rho = |\Psi\rangle\langle\Psi|$, the maximization of the likelihood over all possible operators $\Xi \geq 0$ satisfying the constraints (7.19) follows in a simple way by a repeated use of Schwartz inequality. In fact, the input state $|\Psi\rangle$ can be written in the decomposition (7.13) as

$$|\Psi\rangle = \bigoplus_{\mu \in \mathsf{S}} \ c_\mu \ |\Psi_\mu\rangle\!\rangle \ , \tag{7.21}$$

where each $|\Psi_\mu\rangle\!\rangle \in \mathcal{H}_\mu \otimes \mathcal{M}_\mu$ is a bipartite state. From Schwartz inequality, we have

$$p(g|g) \leq \sum_{\mu,\nu} \ \left| c_\mu^* c_\nu \ \langle\!\langle \Psi_\mu | \Xi | \Psi_\nu \rangle\!\rangle \right| \leq \left(\sum_\mu |c_\mu| \sqrt{\langle\!\langle \Psi_\mu | \Xi | \Psi_\mu \rangle\!\rangle} \right)^2 \tag{7.22}$$

At this point, we assume that each bipartite state $|\Psi_\mu\rangle\!\rangle$ is in the domain of the operator $D_\mu^{-1/2} \otimes \mathbb{1}_{\mathcal{M}_\mu}$. This assumption is not restrictive, since the domain of a self-adjoint operator is dense in the Hilbert space. In this way, it is possible to write $|\Psi_\mu\rangle\!\rangle = D_\mu^{1/2} D_\mu^{-1/2} \otimes \mathbb{1}_{\mathcal{M}_\mu} |\Psi_\mu\rangle\!\rangle$ and to exploit the Schmidt decomposition of the (non-normalized) vector $D_\mu^{-1/2} \otimes \mathbb{1}_{\mathcal{M}_\mu} \ |\Psi_\mu\rangle\!\rangle$:

$$|\Psi_\mu\rangle\!\rangle = \sum_{m=1}^{r_\mu} \ \sqrt{\lambda_m^\mu} \ \sqrt{D_\mu} \otimes \mathbb{1}_{\mathcal{M}_\mu} |\tilde{\psi}_m^\mu\rangle |\tilde{\phi}_m^\mu\rangle \ , \tag{7.23}$$

where r_μ is the Schmidt rank, $\lambda_m^\mu \geq 0$ are Schmidt coefficients such that $\sum_{m=1}^{r_\mu} \lambda_m^\mu = \langle\!\langle \Psi_\mu | D_\mu^{-1} \otimes \mathbb{1}_{\mathcal{M}_\mu} | \Psi_\mu \rangle\!\rangle$, and $|\tilde{\psi}_m^\mu\rangle, |\tilde{\psi}_m^\mu\rangle$ are the elements of two orthonormal bases for $\mathcal{H}_\mu$ and $\mathcal{M}_\mu$, respectively. The form (7.23) is very convenient for optimization, in fact we can use again Schwartz inequality and obtain

$$\langle\!\langle \Psi_\mu | \Xi | \Psi_\mu \rangle\!\rangle \leq \left(\sum_m \sqrt{\lambda_m^\mu \ \langle \tilde{\psi}_m^\mu | \langle \tilde{\phi}_m^\mu | \sqrt{D_\mu} \otimes \mathbb{1}_{\mathcal{M}_\mu} \Xi \sqrt{D_\mu} \otimes \mathbb{1}_{\mathcal{M}_\mu} | \tilde{\psi}_m^\mu \rangle | \tilde{\phi}_m^\mu \rangle} \right)^2 \tag{7.24}$$

Finally, we have

$$\langle\tilde{\psi}^\mu_m|\langle\tilde{\phi}^\mu_m|\sqrt{D_\mu}\otimes \mathbb{1}_{\mathcal{M}_\mu}\Xi\sqrt{D_\mu}\otimes \mathbb{1}_{\mathcal{M}_\mu}|\tilde{\psi}^\mu_m\rangle|\tilde{\phi}^\mu_m\rangle \leq \langle\tilde{\phi}^\mu_m|\ \mathrm{Tr}_{\mathcal{H}_\mu}[D_\mu\otimes \mathbb{1}_{\mathcal{M}_\mu}\ \Xi]\ |\tilde{\phi}^\mu_m\rangle \\ = 1\ ,$$

the last equality following from the normalization constraint (7.19). Therefore, the previous chain of inequalities proves the upper bound

$$p(g|g) \leq \left(\sum_{\mu\in\mathsf{S}} |c_\mu|\ \sum_{m=1}^{r_\mu} \sqrt{\lambda^\mu_m}\right)^2 = L^{opt}\ , \tag{7.25}$$

that holds for any covariant POVM. On the other hand, it is immediate to check that the bound is achieved by the covariant POVM given by $\Xi = |\eta\rangle\langle\eta|$ with

$$|\eta\rangle = \bigoplus_{\mu\in\mathsf{S}} e^{i\arg(c_\mu)}\ \sum_{m=1}^{r_\mu} D_\mu^{-1/2}\otimes \mathbb{1}_{\mathcal{M}_\mu}\ |\tilde{\psi}^\mu_m\rangle|\tilde{\phi}^\mu_m\rangle\ , \tag{7.26}$$

$\arg(z)$ denoting the argument of a complex number, i.e. $z = |z|e^{i\arg(z)}$. The normalization of such a POVM follows from Eq. (7.19), and one has

$$\int \mathrm{d}_L g U_g \Xi U_g^\dagger = \bigoplus_{\mu} \mathbb{1}_{\mathcal{H}_\mu}\otimes \sum_{m=1}^{r_\mu} |\tilde{\phi}^\mu_m\rangle\langle\tilde{\phi}^\mu_m|\ , \tag{7.27}$$

namely, the POVM is complete in the subspace $\mathcal{H}_\Psi$ spanned by the orbit of $|\Psi\rangle$, and can be trivially completed to the whole Hilbert space without affecting the probability distribution. Notice that, if the group $\mathbf{G}$ is unimodular, namely $D_\mu = \mathbb{1}_{\mathcal{H}_\mu}/d_\mu$ for some positive constant d_μ, then we correctly retrieve the result of Theorem 10 in Chapter 4.

The case of $\{U_g\}$ being a direct sum of inequivalent irreps. The expression for the optimal covariant POVM can be further simplified in the case when all the multiplicity spaces $\mathcal{M}_\mu$ are one-dimensional, i.e. when the representation $\{U_g\}$ is a direct sum of inequivalent irreps. In this case, we can decompose an input state $|\psi\rangle \in \mathcal{H}$ as $|\psi\rangle = \bigoplus_\mu\ c_\mu|\psi_\mu\rangle$ (as in Eq. (7.21), but without the need of introducing bipartite states), and now the decomposition (7.23) becomes trivial, namely

$$|\psi_\mu\rangle = \sqrt{\lambda_\mu}\ \sqrt{D_\mu}|\tilde{\psi}_\mu\rangle\ , \tag{7.28}$$

where

$$|\tilde{\psi}_\mu\rangle = \frac{D_\mu^{-1/2}|\psi_\mu\rangle}{||D_\mu^{-1/2}|\psi_\mu\rangle||}\ , \tag{7.29}$$

and

$$\lambda_\mu = ||D_\mu^{-1/2}|\psi_\mu\rangle|| = \langle\psi_\mu|D_\mu^{-1}|\psi_\mu\rangle \ . \tag{7.30}$$

Therefore, Eq. (7.26) for the optimal POVM becomes

$$|\eta\rangle = \bigoplus_{\mu\in\mathsf{S}} e^{i\arg(c_\mu)} \frac{D_\mu^{-1}|\psi_\mu\rangle}{\sqrt{\langle\psi_\mu|D_\mu^{-1}|\psi_\mu\rangle}} = \bigoplus_{\mu\in\mathsf{S}} \frac{D_\mu^{-1}|\psi\rangle}{\sqrt{\langle\psi|D_\mu^{-1}|\psi\rangle}} \ , \tag{7.31}$$

and the corresponding optimal likelihood is given by

$$L^{opt} = |\langle\eta|\psi\rangle|^2 = \left(\sum_{\mu\in\mathsf{S}} \sqrt{\langle\psi|D_\mu^{-1}|\psi\rangle}\right)^2 \ . \tag{7.32}$$

7.2.1 Two unexpected features

- *Square-root measurements do not maximize the likelihood.* A possible strategy to estimate an unknown quantum state, randomly drawn from a given family, is given by the so-called square-root measurements (SRM), firstly introduced by Hausladen and Wootters [41]. As it was showed in Section 4.2, for unimodular groups the square-root measurements and the maximum likelihood measurement coincide. This fact was originally observed for the discrete group of phase shifts in Refs. [83, 84, 85], and then, in the general case of unimodular groups in Ref. [16]. However, as we will see in the following, the case of nonunimodular groups represents an exception to the fact that SRM are optimal for the maximum likelihood criterion in the presence of a physical symmetry. In fact, the SRM for the estimation of a group transformation acting on a fixed state ρ is given by the POVM $M_{sq}(g) = F^{-1/2}\ U_g\rho U_g^\dagger\ F^{-1/2}$, where
$$F = \int_{\mathbf{G}} \mathrm{d}_L g\ U_g\rho U_g^\dagger \tag{7.33}$$
(the POVM $M_{sq}(g)$ is obviously normalized with respect to the left-invariant measure $\mathrm{d}_L g$). The comparison with the maximum-likelihood measurements of the previous section is particularly simple in the case of group representations $\{U_g\}$ that are direct sum of inequivalent irreps. In fact, for a pure state $\rho = |\psi\rangle\langle\psi|$ with $|\psi\rangle = \bigoplus_\mu\ c_\mu|\psi_\mu\rangle$, the integral (7.33) is easily calculated by using Eq. (7.14), namely $F = \bigoplus_\mu |c_\mu|^2\langle\psi_\mu|D_\mu|\psi_\mu\rangle\ \mathbb{1}_{\mathcal{H}_\mu}$. Notice here that the square-root measurement can be defined only if $|\psi_\mu\rangle$ is in the domain of $D_\mu^{1/2}$. Therefore, the square-root measurement is given by the covariant POVM

$M_{sq}(g) = U_g|\eta_{sq}\rangle\langle\eta_{sq}|U_g^\dagger$, where

$$|\eta_{sq}\rangle = \bigoplus_\mu e^{i\arg(c_\mu)}\ \frac{|\psi_\mu\rangle}{\sqrt{\langle\psi_\mu|D_\mu|\psi_\mu\rangle}}\ . \tag{7.34}$$

This covariant POVM is different from the optimal one given in (7.31), and does not achieve the optimal value (7.32) for the likelihood. When $|\psi_\mu\rangle$ is in the domain of both $D_\mu^{1/2}$ and $D_\mu^{-1/2}$, we can compare the values of the likelihood as follows. One has

$$L^{sq} = \left(\sum_\mu \frac{|c_\mu|}{\sqrt{\langle\psi_\mu|D_\mu|\psi_\mu\rangle}}\right)^2 \leq \left(\sum_\mu |c_\mu|\sqrt{\langle\psi_\mu|D_\mu^{-1}|\psi_\mu\rangle}\right)^2 = L^{opt}\ , \tag{7.35}$$

where we used the inequality $\langle\psi_\mu|D_\mu^{-1}|\psi_\mu\rangle\ \langle\psi_\mu|D_\mu|\psi_\mu\rangle \geq 1$, from Schwartz inequality applied to the vectors $D_\mu^{1/2}|\psi_\mu\rangle$ and $D_\mu^{-1/2}|\psi_\mu\rangle$.

- *The true value is not the most likely one.* In the maximum likelihood approach, one optimizes the choice of the POVM in order to maximize the probability density that the estimated value of the parameters coincides with the true one. Intuitively, one could expect that the probability density $p(\hat{g}|g) = \mathrm{Tr}[M(\hat{g})U_g\rho U_g^\dagger]$ for the optimal POVM achieves its maximum at the value $\hat{g} = g$. This is true for unimodular groups, but fails to hold for nonunimodular groups.

Proposition 19. *Let the group* $\mathbf{G}$ *be unimodular. If the covariant density* $M(\hat{g})$ *maximizes the likelihood for a given input state* ρ*, then the probability distribution* $p(\hat{g}|g)$ *of the estimate* $\hat{g}$ *on the state* $U_g\rho U_g^\dagger$ *achieves its maximum for* $\hat{g} = g$.

Proof. Suppose that the most likely value does not coincide with the true one. Then we can rigidly shift the whole probability distribution with a post-processing operation that brings the most likely value to the true one. In fact, if the maximum of $p(\hat{g}|g)$ occurs at $\hat{g} = gh$, we can always replace $M(\hat{g})$ with a new covariant density $M'(\hat{g}) = U_{\hat{g}}\ \Xi'\ U_{\hat{g}}^\dagger$, where

$$\Xi' = U_h \Xi U_h^\dagger\ . \tag{7.36}$$

The normalization of the new POVM follows from the fact that for unimodular groups the DMC operators are trivially proportional to the identity, and therefore the operator Ξ' satisfy the normalization constraints (7.19) as well. Moreover, the probability distribution $p'(\hat{g}|g)$

associated with $M'(\hat{g})$ enjoys the property $p'(\hat{g}|g) = p(\hat{g}h|g)$, whence it achieves the maximum in $\hat{g} = g$. In this way, the likelihood of $M'(\hat{g})$ would be higher than the likelihood of the density $M(\hat{g})$. But this cannot happen since $M(\hat{g})$ is the optimal maximum-likelihood density. Therefore $p(\hat{g}|g)$ must be maximum in $\hat{g} = g$. $\blacksquare$

For nonunimodular groups the previous argument does not apply, since the POVM given by (7.36) is no longer normalized. In fact, the operator Ξ' does not satisfy the normalization constraints (7.19), since the DMC operators do not commute with the unitaries U_h. In other words, one is not allowed to rigidly shift the probability distribution in order to match the most likely value with the true one. As we will see in the explicit example of the estimation of real squeezing and displacement, this situation can indeed happen. In order to reduce the discrepancy between the true value and the most likely one, a suitable choice of the input states is needed. For example, in the simple case of $\{U_g\}$ being a direct sum of inequivalent irreps, if the projection of the input state onto the irreducible subspaces are eigenvectors of the DMC operators, then the most likely value coincides with the true one. In fact, for any input state $|\psi\rangle = \bigoplus_\mu c_\mu|\psi_\mu\rangle$, using Schwarz inequality, we have

$$p^{opt}(\hat{g}|g) = \left|\sum_\mu |c_\mu| \frac{\langle\psi_\mu|U_{g^{-1}\hat{g}} D_\mu^{-1}|\psi_\mu\rangle}{\sqrt{\langle\psi_\mu|D_\mu^{-1}|\psi_\mu\rangle}}\right|^2 \leq \left(\sum_\mu |c_\mu| \sqrt{\frac{\langle\psi_\mu|D_\mu^{-2}|\psi_\mu\rangle}{\langle\psi_\mu|D_\mu^{-1}|\psi_\mu\rangle}}\right)^2 , \tag{7.37}$$

and if each $|\psi_\mu\rangle$ is eigenvector of D_μ, then the last expression is equal to $p(g|g)$, then the true value is the most likely one.

7.3 Joint estimation of real squeezing and displacement

7.3.1 Translation and dilation

In the following we will apply the general framework of Section 7.2 to the case of joint estimation of real squeezing and displacement of a single-mode radiation field with bosonic operators a and $a^\dagger$ satisfying the CCR $[a, a^\dagger] = 1$. Given the wavefunction of a pure state $|\psi\rangle$ in the X-representation $\psi(x) = \langle x|\psi\rangle$, where $|x\rangle$ denotes the Dirac-normalized eigenstate of the quadrature operator $X = (a + a^\dagger)/2$, the affine transformation on the real line given by

$x \to e^r x + x'$ is represented by the unitary transformation

$$\psi(x) \to e^{-r/2}\ \psi(e^{-r}(x - x'))\ , \qquad x', r \in \mathbb{R}\ . \tag{7.38}$$

This transformation corresponds to the action of the unitary operator $U(x', r) = D(x')S(r)$ on the ket $|\psi\rangle$, where

$$\begin{aligned} D(x) &= \exp\left[x(a^\dagger - a)\right]\ , \\ S(r) &= \exp\left[\frac{r}{2}(a^{\dagger 2} - a^2)\right]\ , \end{aligned} \tag{7.39}$$

represent the displacement and the squeezing operator with real argument, respectively. In other words, the operators

$$\{U_{x,r} = D(x)S(r) \mid x, r \in \mathbb{R}\}\ , \tag{7.40}$$

provide a unitary representation of the affine group in the Hilbert space of wavefunctions. The affine group is nonunimodular, and in the above parametrization the left- and right-invariant measures are given by $\mathrm{d}_L g = e^{-r}\mathrm{d}r\mathrm{d}x$, and $\mathrm{d}_R g = \mathrm{d}r\mathrm{d}x$, respectively.

7.3.2 The maximum likelihood POVM

In order to exploit the results of Sec. 7.2, we need to know the Clebsch-Gordan decomposition of the representation $\{U_{x,r}\}$, the irreducible subspaces, and the DMC operators. All these informations are given in the following, while their proof can be found in the Appendix of Ref. [21].

The Clebsch-Gordan series of the representation $\{U_{x,r}\}$ consists on two irreps, that we indicate with the symbols + and −. Accordingly, the Hilbert space splits into two irreducible subspaces, i.e.

$$\mathcal{H} = \mathcal{H}_+ \oplus \mathcal{H}_-\ . \tag{7.41}$$

Comparing this decomposition with the general case (7.13), we see that the subspaces $\mathcal{H}_+$ and $\mathcal{H}_-$ are the representation spaces, while the multiplicity spaces $\mathcal{M}_+$ and $\mathcal{M}_-$ are trivially one-dimensional. The representation spaces $\mathcal{H}_+$ and $\mathcal{H}_-$ can be easily characterized in terms of the quadrature $Y = \frac{a-a^\dagger}{2i}$. In fact, writing the wavefunctions in the Y-representation as $\psi(y) = \langle y|\psi\rangle$, where $|y\rangle$ are the Dirac-normalized eigenvectors of Y, we have $\mathcal{H}_+ = \{|\psi\rangle \mid \psi(y) = 0 \quad \forall y < 0\}$ and $\mathcal{H}_- = \{|\psi\rangle \mid \psi(y) = 0 \quad \forall y > 0\}$. Therefore, the projection operators onto $\mathcal{H}_+$ and $\mathcal{H}_-$ can be written respectively as $\mathbb{1}_+ = \theta(Y)$ and $\mathbb{1}_- = \theta(-Y)$, where $\theta(x)$ is the customary step-function $[\theta(x) = 1$ for $x \geq 0$, $\theta(x) = 0$ for $x < 0]$. Moreover, the DMC operators are

$$D_\pm = \pi\ \frac{\theta(\pm Y)}{|Y|}\ . \tag{7.42}$$

With these tools we are now able to provide the optimal covariant measurement for the joint estimation of real squeezing and displacement on a given state of the radiation field. Let us denote by $|\psi\rangle$ the input state that undergoes to unknown squeezing and displacement transformations. Decomposing the input state on the subspaces $\mathcal{H}_+$ and $\mathcal{H}_-$ as $|\psi\rangle = c_+|\psi_+\rangle + c_-|\psi_-\rangle$, we can exploit Eq. (7.31) and write explicitly the optimal density as $M(x,r) = U_{x,r}|\eta\rangle\langle\eta|U^\dagger_{x,r}$ where

$$|\eta\rangle = \frac{|Y|\theta(Y)|\psi\rangle}{\sqrt{\pi\ \langle\psi||Y|\theta(Y)|\psi\rangle}} + \frac{|Y|\theta(-Y)|\psi\rangle}{\sqrt{\pi\ \langle\psi||Y|\theta(-Y)|\psi\rangle}}\ . \tag{7.43}$$

The optimal likelihood is then

$$L^{opt} = \frac{1}{\pi}\ \left(\sqrt{\langle\psi|\ |Y|\theta(Y)\ |\psi\rangle} + \sqrt{\langle\psi|\ |Y|\theta(-Y)\ |\psi\rangle}\right)^2\ , \tag{7.44}$$

according to the general expression of Eq. (7.32). As already mentioned, the expression of the likelihood provides some insight about the states that are most sensitive in detecting an unknown combination of real squeezing and displacement. Essentially, one can improve the likelihood by increasing the expectation value of $|Y|$, the modulus of the quadrature Y. In addition, the use of wavefunctions that in the Y-representation are non-zero both in the positive half-line and in the negative half-line allows to exploit the interference of the components $|\psi_+\rangle$ and $|\psi_-\rangle$ to enhance the value of the likelihood.

In the following we analyze the performances of the optimal estimation for particular classes of input states, in particular coherent and displaced squeezed states.

- **Coherent states.** Using Eq. (7.43) for the optimal POVM, we can obtain the probability distribution of the estimated squeezing and displacement parameter for a given input state. In particular, for a coherent input state $|\alpha\rangle$ the sensitivity of the measurement can be significantly improved by increasing the imaginary part of α, this corresponding to taking coherent states with a high expectation value of $|Y|$. The probability distribution for the joint estimation of squeezing and displacement on the vacuum state and on a coherent state with $\alpha = 10i$ has been reported in Figs. 1 and 2, respectively. Comparing the two figures, a remarkable improvement in the precision of the measurement can be observed in an enhancement of the likelihood, along with a narrowing of the probability distribution. Moreover, we can observe that for the vacuum state the maximum of the probability density is not achieved by the true value (which is given by to $x = r = 0$). The discrepancy between the most likely value and the true one, due to the fact that the affine group is nonunimodular, essentially disappears by increasing the expectation value of $|Y|$. As we can see from Fig. 2, for $\alpha = 10i$ the probability distribution is approximately a Gaussian centered around the true value $x = r = 0$.

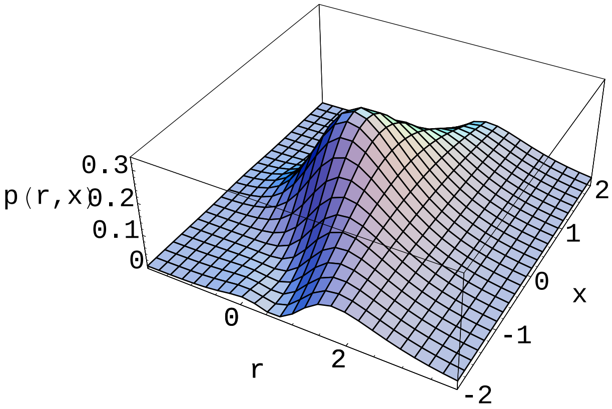

Figure 7.1: Optimal probability distribution for the joint estimation of the squeezing parameter and real displacement for the vacuum state. The maximum of the probability density does not coincide with the true value $x = 0, r = 0$.

Now it is interesting to focus on the asymptotic behavior of the probability distribution for a coherent state $|i\alpha\rangle$ when the amplitude α goes to infinity. In the asymptotic regime, the probability distribution

$$p_\alpha(x,r)\mathrm{d}x\mathrm{d}re^{-r} = |\langle\eta|\; U_{x,r}\; |i\alpha\rangle|^2\, \mathrm{d}x\mathrm{d}re^{-r}\ , \tag{7.45}$$

given by optimal vector $|\eta\rangle$ in Eq. (7.43), can be further simplified. In fact, the wavefunction of the coherent state $|i\alpha\rangle$ is

$$\langle y|i\alpha\rangle = \left(\frac{2}{\pi}\right)^{1/4}\; e^{-(y-\alpha)^2}\ , \tag{7.46}$$

which, for large α, lies almost completely in the positive half-line. In other words, in the expression (7.43) we can asymptotically neglect the component in the subspace $\mathcal{H}_-$, and drop the modulus from $|Y|$. In

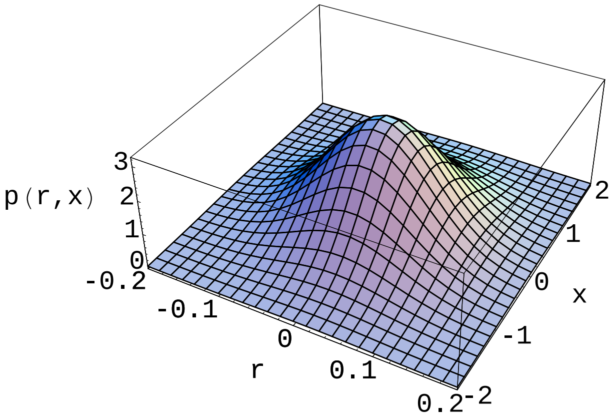

Figure 7.2: Optimal probability distribution for the joint estimation of the squeezing parameter and real displacement for a coherent state with complex amplitude $\alpha = 10i$.

this way, the probability distribution (7.45) can be approximated as

$$p_\alpha(x,r)\mathrm{d}x\mathrm{d}re^{-r} \approx \frac{1}{\pi}\,\frac{|\langle i\alpha|\; U_{x,r}Y\; |i\alpha\rangle|^2}{\langle i\alpha|\; Y\; |i\alpha\rangle}\;\mathrm{d}x\mathrm{d}re^{-r}\ . \tag{7.47}$$

Neglecting the higher order terms, we thus obtain the Gaussian distribution

$$p_\alpha(x,r)\mathrm{d}x\mathrm{d}re^{-r} = \frac{\alpha}{\pi}\;e^{-\alpha^2 r^2}\;e^{-x^2}\;\mathrm{d}x\mathrm{d}r\ . \tag{7.48}$$

Notice that, asymptotically the most likely values of the unknown parameters x and r are the true ones $x = r = 0$, and, in addition, also the mean values of x and r coincide with the true ones, namely the estimation is unbiased. Comparing the expression (7.48) with the numeric plot of Fig. (2) we can notice that the asymptotic regime is obtained already for a small value of the coherent amplitude $\alpha = 10$.

From the asymptotic expression (7.48) can see that the uncertainty in the estimation of the squeezing parameter r goes to zero with the number of photons $\bar{n} = |\alpha|^2$, namely the r.m.s error is

$$\Delta r = \frac{1}{\sqrt{2\bar{n}}}\ , \tag{7.49}$$

while the uncertainty in the estimation of the displacement x remains fixed, with the value

$$\Delta x = \frac{1}{\sqrt{2}} \ . \tag{7.50}$$

- **Displaced squeezed states.** In the case of highly excited coherent states, while the error in the estimation of r goes to zero with the number of photons, the error in estimating x remains fixed. However, it is possible to choose the input state in such a way that both variances vanish in the asymptotic limit. To this purpose, we consider now displaced squeezed states of the form

$$|i\alpha, z\rangle = D(i\alpha)S(z)|0\rangle \qquad \alpha, z \in \mathbb{R} \ . \tag{7.51}$$

Such states have the wavefunction

$$\langle y|i\alpha, z\rangle = \left(\frac{2e^{2z}}{\pi}\right)^{1/4} \ e^{-(y-\alpha)^2 e^{2z}} \ , \tag{7.52}$$

namely a Gaussian centered around the mean value α, with standard deviation $\sigma = 1/(\sqrt{2}e^z)$. Clearly, if the conditions $\alpha \gg 1$ and $\alpha \gg \sigma$ are simultaneously satisfied, such a Gaussian lies almost completely in the positive half-line. Therefore, in the asymptotic limit $\alpha \to +\infty$, $\alpha \gg e^{-z}$, the optimal probability distribution can be approximated as

$$p_{\alpha,z}(x,r) \approx \frac{1}{\pi} \frac{|\langle i\alpha, z| \ U_{x,r}Y \ |i\alpha, z\rangle|^2}{\langle i\alpha, z| \ Y \ |i\alpha, z\rangle} \ . \tag{7.53}$$

By calculating the expectation values and keeping the leading order terms, we then obtain the asymptotic distribution

$$p_{\alpha,z}(x,r)\mathrm{d}x\mathrm{d}re^{-r} = \frac{\alpha}{\pi} e^{-(\alpha e^z r)^2} e^{-(xe^{-z})^2} \mathrm{d}x\mathrm{d}r \ . \tag{7.54}$$

Again, in the asymptotic limit the most likely values in the probability distribution coincide with the true ones, and, moreover, the estimation is unbiased.

The r.m.s. error in the estimation of squeezing and displacement are now given by

$$\Delta r = \frac{1}{\sqrt{2}\alpha e^z} \tag{7.55}$$

and

$$\Delta x = \frac{1}{\sqrt{2}e^{-z}} \ , \tag{7.56}$$

respectively. In order to have both errors vanishing, one needs simultaneously $\alpha e^{z} \gg 1$ and $e^{-z} \gg 1$. For example, we can have an isotropic distribution $\Delta r = \Delta x$ with the choice $\alpha = e^{-2z}$. In the isotropic case, only a small fraction of order $\sqrt{\bar{n}}$ of the total number of photons $\bar{n} = |\alpha|^2 + \sinh^2(z)$ must be used to generate the input state from the vacuum with the squeezing $S(z)$, while almost all the photons in the input state are created by the displacement $D(i\alpha)$. Since one has $\Delta_x = \Delta_r = 1/(\sqrt{2}e^{-z}) \approx 1/(\sqrt{2}\bar{n}^{1/4})$, the convergence to the asymptotic regime is quite slow: the uncertainty goes to zero only with order $1/\bar{n}^{1/4}$ in the average number of photons. As an example in the asymptotic regime, we report in Fig. 3 a numeric plot of the exact probability distribution in the case $\alpha = e^{-2z} = 4000$.

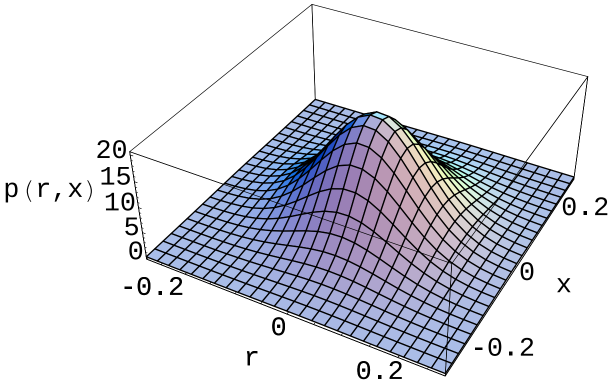

Figure 7.3: Optimal probability distribution for the joint estimation on a displaced squeezed states with $\alpha = e^{-2z} = 4000$.

Remark: *Displaced squeezed states with real squeezing and displacement.*
Since a displacement is estimated efficiently for input states that are well localized in position, and the squeezing is estimated efficiently for input states that are concentrated in a region far from the origin of the phase space, one might think that the displaced squeezed states of the form

$$|\alpha, z\rangle = D(\alpha)S(z)|0\rangle \qquad \alpha, z \in \mathbb{R} \tag{7.57}$$

provide a good estimation for $\alpha >> e^{z}$. However, due to covariance, these state are as good as the vacuum, i.e. they provide a rather poor estimation.

7.3.3 Asymptotic uncertainty relations

It is interesting to compare the precision achieved by the joint measurement with the precision that could be achieved if the parameters x and r were measured separately. In particular, in the cases of coherent and displaced squeezed states, the product of asymptotic uncertainties in the optimal joint measurement is exactly twice the Heisenberg limit, as it happens also in the joint measurement of position and momentum. To see this, we first need the optimal estimation strategies for the separate measurements of real displacement and squeezing. In the case of displacement, the optimal estimation for coherent and displaced squeezed states is obtained by measuring the observable $X = (a + a^\dagger)/2$ and, in case, subtracting a constant offset[1]. In the case of squeezing, instead, the optimal estimation is given in Subsection 6.2.3 of the previous Chapter.

- **Coherent states.** In the case of coherent states $|i\alpha\rangle$ with $\alpha \in \mathbb{R}$, the uncertainty in the measurement of the observable X is

$$\Delta x^{opt} = 1/2 \ . \tag{7.61}$$

On the other hand, for $|\alpha| >> 1$ the uncertainty in the optimal estimation of squeezing is $\Delta r^{opt} = 1/(2\sqrt{\bar{n}})$ with $\bar{n} = |\alpha|^2$, according to Eq. (6.89). This level of precision is asymptotically achieved by measuring

[1] In general, given the input state

$$|\psi\rangle = \int_{-\infty}^{+\infty} \mathrm{d}y \ \psi(y)|y\rangle \ , \tag{7.58}$$

and a cost function $c(\hat{x} - x)$ in the generalized Holevo class

$$c(\hat{x} - x) = \int_{-\infty}^{+\infty} \mathrm{d}y \ c(y) e^{2iy(\hat{x}-x)} \qquad c(y) \leq 0 \quad \forall y \neq 0 \ , \tag{7.59}$$

it is possible to prove in the minimax approach that the optimal POVM for the estimation of the displacement $D(x) = e^{-2ixY}$ is the covariant POVM $M(\mathrm{d}\hat{x}) = D(x)|\eta\rangle\langle\eta|D^\dagger(x) \ \mathrm{d}\hat{x}$ given by

$$|\eta\rangle = \int_{-\infty}^{+\infty} \mathrm{d}y \ e^{i \arg(\psi(y))}|y\rangle \ . \tag{7.60}$$

Accordingly, the optimal POVM for the coherent state $|\alpha\rangle$ is the observable $X - \mathrm{Re}(\alpha)$, while for the displaced squeezed state $|i\alpha, z\rangle$ it is just the observable X.

the observable $\ln|Y/\alpha|$. Notice that the commutator between the two observables X and $\ln|Y/\alpha|$ is

$$[X, \ln|Y/\alpha|] = \frac{i}{2Y} \ , \tag{7.62}$$

whence we have the Heisenberg-Robertson inequality

$$\Delta X \Delta \ln|Y/\alpha| \geq \left| \left\langle \frac{1}{4Y} \right\rangle \right| \tag{7.63}$$

($\langle A \rangle$ and ΔA denote the expectation value of the operator A on a given state $|\psi\rangle$, and the variance $\Delta A = \langle A^2 \rangle - \langle A \rangle$, respectively). In the asymptotic regime $\alpha >> 1$, since $|\langle i\alpha|1/(4Y)|i\alpha\rangle| \approx 1/(4|\alpha|)$, the coherent states attain the Heisenberg limit, i.e.

$$\Delta x^{opt} \Delta r^{opt} = \left| \left\langle \frac{1}{4Y} \right\rangle \right| \ . \tag{7.64}$$

In other words, the coherent states $|i\alpha\rangle$ are characterized asymptotically as minimum uncertainty states.

Comparing the uncertainties Δx^{opt} and Δr^{opt} in the optimal separate measurements with the values of the uncertainties of Eqs. (7.49) and (7.50) we obtain the remarkable relation

$$\left\{ \begin{array}{lcl} \Delta x & = & \sqrt{2}\ \Delta x^{opt} \\ \Delta r & = & \sqrt{2}\ \Delta r^{opt} \ . \end{array} \right. \tag{7.65}$$

As a consequence, the product of the uncertainties in the joint estimation is exactly twice the Heisenberg limit:

$$\Delta x \Delta r = 2 \Delta x^{opt}\ \Delta r^{opt} = \left| \left\langle \frac{1}{2Y} \right\rangle \right| \ . \tag{7.66}$$

Surprisingly, this is the same relation that occurs in the optimal joint measurement of two conjugated quadratures X and Y, achieved by the heterodyne measurement[20].

- **Displaces squeezed states.** In the case of displaced squeezed states of the form $|i\alpha, z\rangle = D(i\alpha)S(r)$ with $\alpha, z \in \mathbb{R}$, the uncertainty in the measurement of the observable X is

$$\Delta x^{opt} = 1/(2e^{-z}) \ . \tag{7.67}$$

On the other hand, in the asymptotic regime $\alpha >> 1, \alpha >> e^{-z}$ the optimal estimation of squeezing is obtained by measuring the observable $\ln(|Y/\alpha|)$, thus giving the uncertainty $\Delta r^{opt} = 1/(2\alpha e^{z})$, as in Eq. (7.55). Since asymptotically $|\langle i\alpha, z|1/(4Y)|i\alpha, z\rangle \approx 1/(4|\alpha|)$, the Heisenberg limit is achieved also in this case, i.e.

$$\Delta x^{opt}\Delta r^{opt} = \left|\left\langle \frac{1}{4Y} \right\rangle\right| \qquad \alpha >> 1, \ \alpha >> e^{-z} \ . \tag{7.68}$$

Hence, the displaced squeezed $|i\alpha, z\rangle$ are characterized asymptotically as minimum uncertainty states.

Moreover, comparing the uncertainties Δx^{opt} and Δr^{opt} with the uncertainties Δx and Δr (Eqs. (7.56) and (7.55), respectively) we still obtain the remarkable relation

$$\left\{ \begin{array}{lcl} \Delta x & = & \sqrt{2}\ \Delta x^{opt} \\ \Delta r & = & \sqrt{2}\ \Delta r^{opt}\ , \end{array} \right. \tag{7.69}$$

and, hence the relation with the Heisenberg limit

$$\Delta x \Delta r = 2\Delta x^{opt}\ \Delta r^{opt} = \left|\left\langle \frac{1}{2Y} \right\rangle\right| \ . \tag{7.70}$$

Chapter 8

Extremal quantum measurements

The previous chapters treated the optimal estimation of signal states that are generated by the action of a group on an input state. Optimality was defined as the minimization of an average cost, either in the Bayes approach or in the minimax. The generalization of such results to more complex setups—e.g. signal states forming the union of different orbits—and to different optimization criteria—e.g. the maximization of the mutual information—is a rather hard topic, which in general do not admit an analytic solution. In those cases, an intermediate step for the identification of optimal strategies is the characterization of the extremal POVMs, which gives a hint when the optimization problem has a convexity property. In this Chapter we provide a characterization of the extremal POVMs for finite dimensional Hilbert spaces, with particular attention to the case of covariant POVMs.

8.1 The convex set of POVMs

In the quantum mechanical description, the possible experiments devised to estimate a certain parameter $\omega \in \Omega$ are represented by operator-valued probability measures (POVMs) on the outcome space Ω. Knowing the POVM P of the measuring apparatus and the state ρ of the observed system, it is indeed possible to predict the probability of any experimental event via the Born rule

$$p_\rho(B) = \mathrm{Tr}[P(B)\rho] \ , \tag{8.1}$$

where $P(B) \geq 0$ is the operator associated with the event B. The set of all such experiments is then represented by the set $\mathsf{M}(\Omega)$ of all possible POVMs having the outcome space Ω.

It is immediate to see that the set of POVMs with outcome space Ω is convex, namely if P_1 and P_2 are elements of $\mathsf{M}(\Omega)$ (i.e. they satisfy the requirements of Eqs. (1.3), (1.4), and (1.5)), then also the convex combination $pP_1 + (1-p)P_2$ is an element of $M(\Omega)$ for any value of the probability p, $0 \leq p \leq 1$. From an experimental point of view, the convex combination $pP_1 + (1-p)P_2$ can be viewed as the randomization between two measuring apparatuses with different statistics, and it can be obtained by actively switching from an apparatus to the other depending on the outcome "1" or "2" of some classical random process, e.g. the toss up of a coin. More realistically, randomizations can occur if some of the parameters of the measuring apparatus are fluctuating during the experiment, or if the experimenter does not know the exact values of all parameters and is forced to assume a probability distribution over them. In a broad sense, convex combinations can be always regarded as a form of classical noise.

Not all measurements can be obtained by randomizing different apparatuses: there are indeed measurements that cannot be decomposed as a convex combination. Such measurements, called *extremal*, that are free from classical noise, are expected to represent efficient strategies for the estimation of a parameter. In particular, here we focus the attention on the statistics of the measurements, and therefore we are interested in the search of *extremal POVMs*:

Definition 16. *A POVM $P \in \mathsf{M}(\omega)$ is called* extremal *if for any couple of POVMs $P_1, P_2 \in \mathsf{M}(\Omega)$ and for any probability $0 < p < 1$ the following implication holds*

$$pP_1 + (1-p)P_2 = P \quad \Longrightarrow \quad P_1 = P_2 = P \ . \tag{8.2}$$

The set of extremal points of the convex set $\mathsf{M}(\Omega)$ will be denoted as $\partial\mathsf{M}(\Omega)$.

A classic result of convex analysis[86] states that for compact convex sets any point of the set can be decomposed in a convex combination of extremal points. Therefore, if the set $\mathsf{M}(\Omega)$ is compact we can write any POVM $P \in \mathsf{M}(\Omega)$ as

$$P = \sum p_i P_i \qquad P_i \in \partial\mathsf{M}(\Omega),\ p_i \geq 0,\ \sum_i p_i = 1 \ . \tag{8.3}$$

In this case, the characterization of the extremal POVMs is very useful for the solution of optimization problems. In fact, when optimality is defined as the minimization of the Bayes cost $\langle c \rangle$, one has to minimize a linear functional of the POVM. The convex decomposition (8.3) shows that if a

POVM P minimizes the cost, then all POVMs $\{P_i\}$ minimize the cost as well. Therefore, there is no loss of generality in restricting the search of the optimal measurement in the set $\partial \mathsf{M}(\Omega)$ of extremal POVMs. Analogously, when optimality is defined as the maximization of the mutual information, since the mutual information $I[P]$ is a convex functional of the POVM, one has $I[P] \leq \sum p_i I[P_i]$, and, again, if the POVM P is optimal, then any extremal POVM P_i in the decomposition is optimal as well.

Notice that, if the convex set $\mathsf{M}(\Omega)$ is not compact, a convex decomposition in extremal POVMs as in Eq. (8.3) is not possible in general. Nevertheless, one might try to prove the existence of a continuous decomposition

$$P = \int_{\partial \mathsf{M}(\Omega)} \mu(\mathrm{d}x) P_x \ , \tag{8.4}$$

where $\mu(\mathrm{d}x)$ is a probability measure on the set of extremal POVMs. To the best of our knowledge, the existence of such a decomposition is still an open problem. However, the characterization of extremal POVMs remains an interesting step in order to understand the role of classical noise in the statistics of a quantum measurement.

8.2 Extremal POVMs in finite dimensional systems

The study of extremal POVMs with finite outcome space $\Omega = \{\omega_1, \dots, \omega_N\}$ was addressed by Parthasarathy in the framework of C^*-algebras[87] and by Størmer in Ref. [88]. More recently, an equivalent characterization was given in Ref.[89] in the case of POVMs in finite dimensional Hilbert spaces. Here we first present an alternative derivation of this characterization, and subsequently generalize it to the case of countable or continuous outcome spaces. In this case we will show that, even though the outcome space is not finite, the extremal POVMs have a finite number of effective outcomes, i.e. any extremal POVM P can be written as

$$P(B) = \sum_{i=\mathcal{I}} \chi_B(\omega_i) \ P_i \ , \tag{8.5}$$

where $\mathcal{I}$ is a finite index set, $\{\omega_i \in \Omega \mid i \in \mathcal{I}\}$ is the support of P, $\{P_i \mid i \in \mathcal{I}\}$ is a POVM, and $\chi_B(\omega)$ is the characteristic function of the set B ($\chi_B(\omega) = 1$ for $\omega \in B$ and $\chi_B(\omega) = 0$ otherwise). This result suggests that the information encoded by a finite-dimensional quantum system might be optimally read out with a measurement with only a finite number of outcomes, i.e.

that one might get rid of measurements with a continuous space of outcomes. However, if the set Ω is not finite, the convex set of POVMs is not compact, and in order to prove that optimization can be restricted without loss of generality to the extremal POVMs, one should prove the existence of a decomposition of the form (8.4).

In order to characterize the extremal POVMs, we will use the standard method of perturbations, which is based on the following

Definition 17. *Let P be an element of $\mathsf{M}(\Omega)$. A Hermitian operator-valued function $Z : B \longmapsto Z(B) \in \mathcal{B}(\mathcal{H})$ is a* perturbation *of P if there exists an $\epsilon > 0$ such that $P + tZ \in \mathsf{M}(\Omega)$ for any $t \in [-\epsilon, \epsilon]$, i.e. if*

$$Z(\Omega) = 0 \tag{8.6}$$

and

$$P(B) + tZ(B) \geq 0 \qquad \forall t \in [-\epsilon, \epsilon] \ . \tag{8.7}$$

By this definition, it is immediate to realize that a POVM is extremal if and only if is admits only the trivial perturbation $Z(B) \equiv 0 \quad \forall B \in \sigma(\Omega)$.

8.2.1 Extremal POVMs with finite outcome space

A POVM with finite outcome space $\Omega = \mathcal{I}$ can be represented either as a vector of operators $\{P_i \mid i \in \mathcal{I}\}$, or, introducing the auxiliary Hilbert space

$$\mathcal{K} = \bigoplus_{i \in \mathcal{I}} \mathcal{W}_i, \qquad \mathcal{W}_i \cong \mathcal{H} \quad \forall i \in \mathcal{I} \ , \tag{8.8}$$

as a single block operator

$$P = \bigoplus_{i \in \mathcal{I}} P_i \ . \tag{8.9}$$

The relation between a POVM and its block operator defines a one-to-one linear map from the convex set $\mathsf{M}(\Omega)$ and a convex subset C of the space of block operators, here denoted as $\bigoplus_{i \in \mathcal{I}} \mathcal{B}(\mathcal{W}_i)$. The convex set $\mathsf{C} \subset \bigoplus_{i \in \mathcal{I}} \mathcal{B}(\mathcal{W}_i)$ is defined by the two constraints

$$P \geq 0 \ , \tag{8.10}$$

and

$$\mathcal{L}(P) = \mathbb{1} \ , \tag{8.11}$$

where $\mathcal{L} : \bigoplus_{i \in \mathcal{I}} \mathcal{B}(\mathcal{W}_i) \to \mathcal{B}(\mathcal{H})$ is the linear map defined by

$$\mathcal{L}(P) = \sum_{i \in \mathcal{I}} P_i \ . \tag{8.12}$$

Consider the trace norm $||\cdot||_1$ in the space of block operators, defined by $||O||_1 \doteq \sum_{i\in\mathcal{I}} \mathrm{Tr}|O_i|$, for $O = \bigoplus_{i\in\mathcal{I}} O_i$. Then, we have

Proposition 20. *The convex set* C *is compact in the block operator norm* $||\cdot||$.

Proof. Since C is a subset of the finite dimensional space $\bigoplus_{i\in\mathcal{I}} \mathcal{B}(\mathcal{W}_i) \quad \mathcal{W}_i \cong \mathcal{H}$, it is enough to show that C is bounded and closed. First, for any element $P \in \mathsf{C}$ one has $||P||_1 = d$, i.e. C is bounded. Moreover, any Cauchy sequence P_n converges to a block operator $P \in \bigoplus_{i\in\mathcal{I}} \mathcal{B}(\mathcal{W}_i)$. Such an operator is nonnegative and satisfies the relation $\mathcal{L}(P) = \mathbb{1}$. In fact, since $\mathcal{L}$ is continuous, one has $||\mathcal{L}(P) - \mathbb{1}||_1 = ||\mathcal{L}(P - P_n)||_1 \to 0$. Therefore, C is closed. ■

As a consequence, any POVM $\{P_i\}$ can be written as a convex combination of a finite number of extremal POVMs.

A perturbation of the block operator $P \in \mathsf{C}$ it given by a Hermitian block operator $Z \in \bigoplus_{i\in\mathcal{I}} \mathcal{B}(\mathcal{W}_i)$ with the properties

$$\exists\epsilon > 0 : \qquad P + tZ \geq 0, \quad \forall t \in [-\epsilon, \epsilon] \tag{8.13}$$

and

$$\mathcal{L}(Z) = 0 \ . \tag{8.14}$$

The first condition can be simplified by the following

Lemma 11. *Let* $\mathcal{H}$ *be a Hilbert space and* $A \in \mathcal{B}(\mathcal{H})$ *a nonnegative operator. Then, for any Hermitian operator* $B \in \mathcal{B}(\mathcal{H})$ *the condition*

$$\exists\epsilon > 0 : \qquad A \pm tB \geq 0 \quad \forall t \in [-\epsilon, \epsilon] \tag{8.15}$$

implies the inclusion of the supports

$$\mathrm{Supp}(B) \subseteq \mathrm{Supp}(A) \ . \tag{8.16}$$

Moreover, if $\mathcal{H}$ *is finite dimensional the two conditions are equivalent.*

Proof. Suppose that the condition (8.15) holds, then for any $|\phi\rangle \in \mathrm{Ker}(A)$ it must be $\langle\phi|B|\phi\rangle = 0$. Moreover, for any vector $|\psi\rangle \in \mathcal{H}$ one has:

$$|\langle\psi|B|\phi\rangle| = \frac{1}{t}|\langle\psi|(A+tB)|\phi\rangle| \leq \frac{1}{t}\sqrt{\langle\psi|(A+tB)|\psi\rangle \ \langle\phi|(A+tB)|\phi\rangle} = 0 \ .$$

Hence $\mathrm{Ker}\,(A) \subseteq \mathrm{Ker}\,(B)$, implying $\mathrm{Supp}\,(B) \subseteq \mathrm{Supp}\,(A)$. In the case of a finite dimensional Hilbert space, suppose that (8.16) holds. Let us denote by λ the smallest nonzero eigenvalue of A and by $||B||$ the operator norm of B, then condition (8.15) holds with $\epsilon = \frac{\lambda}{||B||}$. ■

Using the previous Lemma, we immediately a first characterization of extremal POVMs:

Theorem 16 (minimal support condition). *Let P be an element of C. Then, P is extremal if and only if for any element $Q \in \mathsf{C}$ the following implication holds*

$$\mathrm{Supp}(Q) \subseteq \mathrm{Supp}(P) \Longrightarrow Q = P \ . \tag{8.17}$$

Proof. Suppose that P is extremal. The operator $Z \doteq P - Q$ is a perturbation of P. In fact, since $\mathcal{H}$ is finite dimensional and $\mathcal{I}$ is finite, Z acts on a finite dimensional Hilbert space, and therefore the fact that $\mathrm{Supp}(Z) \subseteq \mathrm{Supp}(P)$ along with Lemma 11 implies that Z satisfies the condition Eq. (8.13). Moreover, $\mathcal{L}(Z) = \mathcal{L}(P-Q) = 0$, whence also the condition (8.14) is satisfied. Since P is extremal, $Z = P-Q$ must be zero. Conversely, suppose that the implication (8.52) holds and the Z is a perturbation for P. Then, the condition (8.13) implies $\mathrm{Supp}(Z) \subseteq \mathrm{Supp}(P)$, due to Lemma 11. Therefore, one has $\mathrm{Supp}(P + tZ) \subseteq \mathrm{Supp}(P)$, whence $Z = 0$, i.e. P is extremal. $\blacksquare$

As we will see in Section 8.4, this simple result has some important consequences in the optimization of an estimation strategy. A straightforward consequence of the minimal support condition is the following:

Corollary 5. *Any orthogonal POVM is extremal.*

Proof. Let $\{P_i\}$ be an orthogonal POVM and $\{Q_i\}$ a POVM with the property $Q_i \subseteq P_i \quad \forall i$. Then, from the equality $\sum_i Q_i = \sum_i P_i$ one obtains $Q_i = P_i \quad \forall i$. This obviously implies the equality of the block operators P and Q, whence $\{P_i\}$ is extremal $\blacksquare$

A deeper characterization of the extremal POVMs can be obtained by writing $P = \bigoplus_{i \in \mathcal{I}} P_i$ and introducing the projector Π_i on the support $\mathrm{Supp}(P_i)$.

Theorem 17 (spanning set condition). *Let $P = \bigoplus_{i \in \mathcal{I}} P_i$ be an element of C, and let $\{F_{mn} \mid m,n = 1,\dots,d\}$ be the operators defined by*

$$F_{mn} \doteq \bigoplus_{i \in \mathcal{I}} \Pi_i |m\rangle\langle n| \Pi_i \ , \tag{8.18}$$

where $\{|m\rangle \mid m = 1,\dots,d\}$ is an orthonormal basis for $\mathcal{W}_i \cong \mathcal{H}$. Then, P is extremal if and only if

$$\mathrm{Span}\{F_{mn} \mid m,n = 1,\dots,d\} = \bigoplus_{i \in \mathcal{I}} \mathcal{B}(\mathrm{Supp}(P_i)) \ . \tag{8.19}$$

Proof. Suppose that the condition (8.19) holds. Then, let $Z = \bigoplus_{i\in\mathcal{I}} Z_i$ be a perturbation for P. Due to Lemma 11, one has $\mathrm{Supp}(Z_i) \subseteq \mathrm{Supp}(P_i)$, i.e. $Z \in \bigoplus_{i\in\mathcal{I}} \mathcal{B}(\mathrm{Supp}(P_i))$. Moreover, we can write Z as $Z = \bigoplus_{i\in\mathcal{I}} \Pi_i Z_i \Pi_i$: in this way, condition (8.14) implies

$$\langle n| \left(\sum_{i\in\mathcal{I}} \Pi_i Z_i \Pi_i \right) |m\rangle = 0 \qquad \forall m,n \ = 1,\dots,d \ , \tag{8.20}$$

which is equivalent to $\mathrm{Tr}[ZF_{mn}] = 0 \ \forall m,n = 1,\dots d$, i.e. Z is orthogonal to all operators F_{mn} in the Hilbert-Schmidt product. Since the operators $\{F_{mn}\}$ span the whole algebra of block operators, the only operator which is orthogonal to all of them is the null operator. Hence, $Z = 0$, i.e. P is extremal. Conversely, suppose that P is extremal and suppose that condition (8.19) does not hold. This means that there exists a block operator $0 \neq Z \in \bigoplus_{i\in\mathcal{I}} \mathcal{B}(\mathrm{Supp}(P_i))$ which is not in the span of the set $\{F_{mn}\}$. In particular, we would have $\mathrm{Tr}[ZF_{mn}] = 0 \quad \forall m,n$. The operator Z can be embedded in the larger algebra $\bigoplus_{i\in\mathcal{I}} \mathcal{B}(\mathcal{W}_i) \quad \mathcal{W}_i \cong \mathcal{H}$. Then, Z satisfies the condition (8.11). Moreover, since $\mathrm{Supp}(Z_i) \subseteq \mathrm{Supp}(P_i)$, Lemma 11 guarantees that also the condition 8.13 is fulfilled. Hence, $Z \neq 0$ is a perturbation, in contradiction with the fact that P is extremal. ■

The above Theorem gives a remarkable bound on the ranks of the POVM operators P_i in an extremal POVM:

Corollary 6. *Let $P = \bigoplus_{i\in\mathcal{I}} P_i$ be an extremal point of S. Then, the ranks $r_i = \mathrm{rank}(P_i)$ must satisfy the relation*

$$\sum_{i\in\mathcal{I}} r_i^2 \leq d^2 \ . \tag{8.21}$$

Proof. The dimension of the algebra $\bigoplus_{i\in\mathcal{I}} \mathcal{B}(\mathrm{Supp}(P_i))$ is $\sum_{i\in\mathcal{I}} r_i^2$. Due to condition (8.19), such a dimension cannot exceed the cardinality of the spanning set $\{F_{mn} \mid m,n = 1,\dots,d\}$, which is d^2. ■

Another immediate consequence regards the number of nonzero operators in an extremal POVM:

Corollary 7. *Let $\{P_i\}$ be an extremal POVM with outcome space $\mathcal{I}$. Then, the number of nonzero operators $P_i \neq 0$ cannot exceed d^2.*

If we define as *effective outcome* an outcome that corresponds to a nonzero operator, i.e. an outcome that has nonzero probability for at least one state, then we have that for an extremal POVM in the Hilbert space $\mathcal{H} \cong \mathbb{C}^d$ the number of effective outcomes cannot exceed d^2. Any POVM with a larger number of effective outcomes is necessarily obtained by introducing a classical randomness in the measuring apparatus. Therefore, as long as finite outcome spaces are concerned, the optimal extraction of classical information can be achieved by a POVM with less than d^2 effective outcomes.

8.2.2 Extremal POVMs with infinite outcome space

Here we consider POVMs for the finite dimensional Hilbert space $\mathcal{H} \cong \mathbb{C}^d$ with infinite outcome space Ω, that can be either discrete or continuous.

The case of discrete outcome space is a simple generalization of the characterization of the previous Subsection. In this case we have the following

Theorem 18. *Let P_i be an extremal POVM in $\mathcal{H} \cong \mathbb{C}^d$ with discrete outcome space $\mathcal{I}$, and $\mathsf{E} \subset \mathcal{I}$ the subset defined by $\mathsf{E} = \{i \in \mathsf{I} \mid P_i \neq 0\}$. Then, one has $|\mathsf{E}| \leq d^2$.*

Proof. Suppose that there are d^2+1 nonzero operators, corresponding to the subset $\mathsf{E} \supset \{i_1, \dots, i_{d^2+1}\}$. Define $\lambda^{i_k}_{\min}$ and Π_{i_k} as the minimum nonzero eigenvalue and the projector on the support of P_{i_k}, respectively. Then, define the vectors $|f_{mn}\rangle \in \mathbb{C}^{d^2+1}$ with components $f^k_{mn} = \lambda^{i_k}_{\min} \langle m|\Pi_{i_k}|n\rangle$. Since the vectors $|f_{mn}\rangle$ are d^2, there should be a nonzero vector $|g\rangle$ in $\mathbb{C}^{d^2+1}$ such that $\langle g|f_{mn}\rangle = 0 \quad \forall m, n = 1, \dots d$. Moreover, since $f^{k*}_{mn} = f^k_{nm}$, the components g^k can be chosen to be real. Then, we claim that the Hermitian operators defined as

$$Z_i = \begin{cases} \lambda^{i_k}_{\min}\, g^k\, \Pi_{i_k} & i = i_k \in G \\ 0 & i \notin G \end{cases} \tag{8.22}$$

yield a perturbation for P. In fact, the condition (8.6) is satisfied

$$\langle m| \sum_{i \in \mathcal{I}} Z_i |n\rangle = \sum_{k=1}^{d^2+1} g^k f^k_{nm} = \langle g|f_{nm}\rangle = 0 \ . \tag{8.23}$$

On the other hand, the positivity condition (8.7) holds with $\epsilon = 1/C$, where $C = \max_{i=1,\dots,d^2+1} |g(\omega_i)|^2$. Since we constructed a nonzero perturbation, P_i is not extremal, i.e. we have a contradiction. ■

Thanks to this observation, the problem of characterizing the extremal POVMs is reduced to the case of a finite outcome space:

Theorem 19. *Let $\{P_i\}$ be a POVM for $\mathcal{H} \cong \mathbb{C}^d$ with discrete outcome space $\mathcal{I}$ and $\mathsf{E} \subset \mathcal{I}$ be the subset $\mathsf{E} = \{i \in \mathcal{I} \mid P_i \neq 0\}$. Then, $\{P_i\}$ is extremal if and only if*

1. $|\mathsf{E}| \leq d^2$

2. $\{P_i \mid i \in \mathsf{E}\}$ *is an extremal POVM with finite outcome space* E*.*

Also in the case of discrete infinite outcome space the number of effective outcomes of an extremal POVM must be smaller than d^2.

The problem of characterizing the extremal POVMs with continuous outcome spaces is much harder than the ones showed before, and until now no solution has been presented. Here we provide a characterization in the case of the outcome space Ω being a subset of $\mathbb{R}^n$ for some n, equipped with the $\sigma-$algebra of Borel subsets $\sigma(\Omega)$.

First of all, we prove that any element $P \in \mathsf{M}(\Omega)$ admits a density with respect to a suitable scalar measure on Ω.

Lemma 12. *Let P be an element of $\mathsf{M}(\Omega)$. Then, there exist a finite scalar measure $\mu(\mathrm{d}\omega)$ and an operator-valued function $M : \omega \mapsto M(\omega) \in \mathcal{B}(\mathcal{H})$, uniquely defined μ-almost everywhere such that*

$$P(B) = \int_B \mu(\mathrm{d}\omega)\ M(\omega) \qquad \forall B \in \sigma(\Omega)\ . \tag{8.24}$$

Moreover, the operators $M(\omega)$ are essentially nonnegative and bounded, namely

$$\begin{aligned} M(\omega) &\geq 0 \qquad \mu - a.e. & (8.25)\\ \mathrm{Tr}[M(\omega)] &= 1 \qquad \mu - a.e. & (8.26)\end{aligned}$$

Proof. Define the measure $\mu(\mathrm{d}\omega)$ via the relation

$$\mu(B) = \mathrm{Tr}[P(B)]\ . \tag{8.27}$$

With this definition μ dominates P, i.e. $P(B) \leq \mu(B)\mathbb{1} \quad \forall B$. This implies that P admits an operator density with respect to μ, which is defined $\mu-$almost everywhere and necessarily positive therein. Moreover, one has

$$\int_B \mu(\mathrm{d}\omega)\ \mathrm{Tr}[M(\omega)] = \mathrm{Tr}[P(B)] = \mu(B) = \int_B \mu(\mathrm{d}\omega) \qquad \forall B\ , \tag{8.28}$$

whence $\mathrm{Tr}[M(\omega)] = 1$ except for zero measure sets. ■

Moreover, any perturbation of a given POVM $P \in \mathsf{M}(\Omega)$ admits an operator density with respect to the same scalar measure:

Lemma 13. *Let P be a POVM in $\mathsf{M}(\Omega)$ and Z a perturbation of P. Then, there exists an operator valued function $Y : \omega \mapsto Y(\omega)$, uniquely deined $\mu-$almost everywhere, such that*

$$Z(B) = \int_B \mu(\mathrm{d}\omega)\ Y(\omega) \qquad \forall B \in \sigma(\Omega)\ , \tag{8.29}$$

where $\mu(\mathrm{d}\omega)$ is the finite scalar measure defined by $\mu(B) = \mathrm{Tr}[P(B)] \quad \forall B$.

Proof. Suppose that Z is a perturbation of P. For any couple of vectors $|\phi\rangle$ and $|\psi\rangle$ in the Hilbert space, consider the matrix element $\langle\phi|Z(B)|\psi\rangle$: we have

$$\begin{aligned} |\langle\phi|Z(B)|\psi\rangle| &\leq |\langle\phi|Z_+(B)|\psi\rangle| + |\langle\phi|Z_-(B)|\psi\rangle| \\ &\leq \sqrt{\langle\phi|Z_+(B)|\phi\rangle\ \langle\psi|Z_+(B)|\psi\rangle} + \sqrt{\langle\phi|Z_-(B)|\phi\rangle\ \langle\psi|Z_-(B)|\psi\rangle} \\ &\leq \mathrm{Tr}[Z_+(B)] + \mathrm{Tr}[Z_-(B)] \\ &\leq 1/\epsilon\ \mathrm{Tr}[M(B)] = 1/\epsilon\ \mu(B)\ . \end{aligned}$$

The fist inequality comes from triangular inequality applied to $Z(B) = Z_+(B) - Z_-(B)$, where $Z_+(B)$ ($Z_-(B)$) is the positive (negative) part of $Z(B)$. The third inequality comes from the fact that, due to condition (8.13), one has $\mathrm{Tr}[Z_\pm(B)] \leq 1/\epsilon\ \mathrm{Tr}[M(B)]$. Since the moduli of the matrix elements are dominated by the scalar measure $\mu(\mathrm{d}\omega)$, $Z(B)$ admits an operator density with respect to $\mu(\mathrm{d}\omega)$. ■

In conclusion, Lemma 12 states that any POVM $P \in \mathsf{M}(\Omega)$ lies in the convex subset of POVMs $\mathsf{M}(\Omega,\mu)$ that have a density with respect to the measure μ, while Lemma 13 shows that *all* POVMs contained in a convex decomposition of P must lie in $\mathsf{M}(\Omega,\mu)$.

The perturbations of a given POVM can be easily characterized in terms of their operator density:

Proposition 21. *Let P be an element of $\mathsf{M}(\Omega)$, with operator density $M(\omega)$ with respect to the measure $\mu(B) = \mathrm{Tr}[P(B)]$. A Hermitian operator-valued function $Z : B \longmapsto Z(B)$ is a perturbation of P is and only if*

1. *Z admits a density $Y(\omega)$ with respect to the measure $\mu(\mathrm{d}\omega)$*

2. *the following relations hold*

$$\int_\Omega \mu(\mathrm{d}\omega)\ Y(\omega) = 0\ , \tag{8.30}$$

and

$$\exists\epsilon > 0 : \qquad M(\omega) + tY(\omega) \geq 0 \quad \forall t \in [-\epsilon, \epsilon] \tag{8.31}$$

$\mu-$almost everywhere.

Proof. The existence of a density is ensured by Lemma 13, while equations (8.30) and (8.31) are just the conditions (8.6) and (8.7) expressed in terms of the operator density. ■

Define the *support* E of the measure μ as the complement in Ω of the largest subset of zero measure $B_0 = \bigcup_{\mu(B)=0} B$. Note that, by definition, the support is a measurable subset of Ω. In terms of the support E, we then have the following necessary condition:

Theorem 20. *Let $P : B \longmapsto P(B) \in \mathcal{B}(\mathcal{H})$ be an extremal POVM for Ω, with operator density $M(\omega)$ with respect to the measure $\mu(\mathrm{d}\omega) = \mathrm{Tr}[P(\mathrm{d}\omega)]$. Then the support of μ must be finite, with $|\mathsf{E}| \leq d^2$.*

Proof. First, if E is continuous there exists a nonzero perturbation for P. Consider a compact subset $K \subseteq \mathsf{E}$ and define the functions

$$f_{mn}(\omega) = \chi_K(\omega) \ \ \langle m|M(\omega)|n\rangle \tag{8.32}$$

where χ_K is the characteristic function of the set K and $|m\rangle, |n\rangle$ are elements of an orthonormal basis for $\mathcal{H} \cong \mathbb{C}^d$. Notice that the functions $f_{mn}(\omega)$ are essentially bounded, namely

$$|f_{mn}(\omega)| \leq \sqrt{\langle m|M(\omega)|m\rangle\langle n|M(\omega)|n\rangle} \leq \mathrm{Tr}[M(\omega)] = 1 \quad \mu - a.e. \tag{8.33}$$

the last equality coming from Lemma 12. Thus $f_{mn} \in L^\infty(K,\mu)$. Since K is compact, this also implies that f_{mn} are elements of the Hilbert space $L^2(K,\mu)$. Then, since the space $L^2(K,\mu) \cap L^\infty(K,\mu)$ is infinite dimensional, and the set $\{f_{mn}\}$ is finite, there exists a function $0 \neq g(\omega) \in L^2(K,\mu) \cap L^\infty(K,\mu)$ which is orthogonal to any element f_{mn}, i.e.

$$\int_K \mu(\mathrm{d}\omega) \ g(\omega)^* f_{mn}(\omega) = 0 \quad \forall m,n \ . \tag{8.34}$$

Moreover, since $f_{nn}(\omega) = f^*_{mn}(\omega)$, the function $g(\omega)$ can be chosen to be real. Then, we claim that the operator function

$$Y(\omega) = \chi_K(\omega) \ g(\omega) \ M(\omega) \tag{8.35}$$

defines a perturbation for P. The condition (8.30) is satisfied, namely

$$\langle m| \left(\int_\Omega \mu(\mathrm{d}\omega) \ Y(\omega)\right) \ |n\rangle = \int_K \mu(\mathrm{d}\omega) \ g(\omega) f_{mn}(\omega) = 0 \qquad \forall m,n \ .$$

Moreover, since $g(\omega)$ is in $L^\infty(K,\mu)$, there exists a constant $C > 0$ such that $|g(\omega)| \leq C$ $\mu-$almost everywhere in K. Then, condition (8.31) is then satisfied with $\epsilon = 1/C$. Therefore $Y(\omega)$ represents a nonzero perturbation, i.e. P is not extremal. In this way we proved that E cannot be continuous. In the case of discrete E with $|\mathsf{E}| > d^2$ it is possible to construct a perturbation in the same way as in Theorem 18. Hence, E must be discrete, with $|\mathsf{E}| \leq d^2$. ■

The conclusion of the above Theorem is that any extremal POVM must be necessary of the form

$$P(B) = \sum_{i \in \mathsf{E}} \chi_B(\omega_i)\ P_i\ , \qquad |\mathsf{E}| \leq d^2 \tag{8.36}$$

where $\{P_i\}$ is a POVM for the finite outcome space E. Again, the number of effective outcomes must be finite. This leads directly to the following characterization of the extremal POVMs

Theorem 21. *Let P be a POVM for the Hilbert space $\mathcal{H} \cong \mathbb{C}^d$ with outcome space $\Omega \subseteq\ Reals^n$. Then, P is extremal if and only if*

$$P(B) = \sum_{i=1}^{|\mathsf{E}|} \chi_B(\omega_i)\ P_i\ , \tag{8.37}$$

where E is a finite set with $|\mathsf{E}| \leq d^2$ and $\{P_i\}$ is an extremal POVM with finite outcome space E.

8.3 Extremal covariant POVMs in finite dimensional systems

In this Section we will study the extremal point of the convex set of covariant POVMs in a finite dimensional Hilbert space with given outcome space Ω. This convex set is a convex subset of the set $\mathsf{M}(\Omega)$ of arbitrary POVMs with outcome space Ω, and therefore an extremal covariant POVM may be not an extremal point of $\mathsf{M}(\Omega)$. Typically, for continuous groups, or for finite groups a number of elements larger than d^2 with $d = \dim(\mathcal{H})$, no covariant POVM is extremal in $\mathsf{M}(\Omega)$, due to Theorem 20 and Corollary 7. However, since in the optimization problems in presence of group symmetry one can restrict the attention to the convex subset of covariant POVMs (see Sec. 3.3), the characterization of the extremal points of such a subset becomes relevant.

In the following we will consider the case of nontransitive outcome space Ω, where Ω the union of a finite number of orbits generated by the action of the group $\mathbf{G}$. The group $\mathbf{G}$ will be assumed to be either finite or compact. According to Eq. (3.4), we will have

$$\Omega = \bigcup_{i \in \mathcal{I}} \mathcal{O}_i\ , \tag{8.38}$$

where $\mathcal{O}_i$ denotes a group orbit, labeled by an index i in some finite index set $\mathcal{I}$. In the most general case, the outcome space has not necessarily to be the union of a *finite* number of orbits, namely it might be the union of a countable or even continuous set of orbits. However, the same arguments used in the previous Section for the characterization of extremal POVMs allow one to prove that for finite dimensional Hilbert spaces $\mathcal{H} \cong \mathbb{C}^d$ an extremal covariant POVM is nonzero only over a finite set of orbits $\mathcal{I}$, with $|\mathcal{I}| \leq d^2$. Therefore, there is no loss of generality in studying the case of Eq. (8.38).

For simplicity, we will start from the case of trivial stability groups, i.e. the case where the points of each orbit $\mathcal{O}_i$ are in one-to-one correspondence with the elements of the group $\mathbf{G}$. In this case, the outcome space is the Cartesian product

$$\Omega = \mathcal{I} \times \mathbf{G} \ . \tag{8.39}$$

Finally, the results obtained for $\Omega = \mathcal{I} \times \mathbf{G}$ will be generalized to the case of nontrivial stability groups, where Ω is of the form of Eq. (8.38), with $\mathcal{O}_i \cong \mathbf{G}/\mathbf{G}_i$ for a set of stability groups $\mathbf{G}_i$.

8.3.1 The convex set of covariant POVMs with trivial stability groups

As a particular case of Proposition 9 of Chapter 1, in the case $\Omega = \mathcal{I} \times \mathbf{G}$ one can prove that any covariant POVM admits an operator density $M(i,g)$ with respect to the (normalized) Haar measure $\mathrm{d}g$ on the group $\mathbf{G}$, namely, if $B = (i, A)$, where $A \subseteq \mathbf{G}$ is a measurable subset, then $P(B) = \int_A \mathrm{d}g \, M(i,g)$. Moreover, such an operator density has necessarily the form

$$M(i,g) = U_g \ A_i \ U_g^\dagger \ , \tag{8.40}$$

where $A_i \in \mathcal{B}(\mathcal{H})$ are Hermitian operators satisfying the constraints

$$A_i \geq 0 \qquad \forall i \in \mathcal{I} \tag{8.41}$$

$$\sum_{i \in \mathcal{I}} \int_{\mathbf{G}} \mathrm{d}g \ U_g \ A_i \ U_g^\dagger = \mathbb{1} \ , \tag{8.42}$$

the latter coming form Eq. (3.23). According to the notation of this presentation, we adopt here for the Haar measure the normalization $\int_{\mathbf{G}} \mathrm{d}g = 1$.

According to the above discussion, any covariant POVM with probability space $\Omega = \mathcal{I} \otimes \mathbf{G}$ is completely specified by a set of operators $\{A_i \mid i \in \mathcal{I}\}$, such that both constraints in Eqs. (8.41) and (8.42) are satisfied. Moreover, it is very useful to represent such a vector of operators as a single block operator $A = \bigoplus_{i \in \mathcal{I}} A_i$, acting on an auxiliary Hilbert space $\mathcal{H}_{aux} \doteq \bigoplus_{i \in \mathcal{I}} \mathcal{W}_i$, where

$\mathcal{W}_i \cong \mathcal{H} \quad \forall i \in \mathcal{I}$. In terms of the block operator $A \in \bigoplus_{i\in\mathcal{I}} \mathcal{B}(\mathcal{W}_i)$ the two constraints Eq. (8.41) and Eq. (8.42) become

$$A \geq 0 \tag{8.43}$$

and

$$\mathcal{L}(A) = \mathbb{1} \ , \tag{8.44}$$

where $\mathcal{L} : \bigoplus_{i\in\mathcal{I}} \mathcal{B}(\mathcal{W}_i) \to \mathcal{B}(\mathcal{H})$ is the linear map

$$\mathcal{L}(A) \doteq \sum_{i\in\mathcal{I}} \int_{\mathbf{G}} \mathrm{d}g \ U_g A_i U_g^\dagger \ . \tag{8.45}$$

The two constraints (8.43) and (8.44) define a convex subset of the space of block operators $\bigoplus_{i\in\mathcal{I}} \mathcal{B}(\mathcal{W}_i)$, which is in one-to-one affine correspondence with the convex set of covariant POVMs. In the following, the convex set of block operators will be denoted by D.

Proposition 22. *The convex set* D*, defined by the constraints (8.43) and (8.44) is compact in the trace norm.*

Proof. Since D is a subset of a finite dimensional vector space, it enough to show that C is bounded and closed. D is bounded, since for any $A \in \mathsf{D}$, one has $||A||_1 = \sum_{i\in\mathcal{I}} \mathrm{Tr}[A_i] = \mathrm{Tr}[\mathcal{L}(A)] = d$ (using Eqs. (8.43) and (8.44)). Moreover, D is closed. In fact, if $\{A_n\}$ is a Cauchy sequence of points in D, then A_n converges to some block operator $A \in \bigoplus_{i\in\mathcal{I}} \mathcal{B}(\mathcal{W}_i)$. We claim that A belongs to D. Of course, A satisfies condition (8.43). As regards condition (8.44), just notice that the $\mathcal{L}$ is continuous, being linear. Therefore, we have $||\mathcal{L}(A) - \mathbb{1}||_1 = ||\mathcal{L}(A - A_n)||_1 \to 0$, namely A satisfies condition (8.44). ■

Notice that, since the convex set D is compact, it coincides with the convex hull of its extreme points, i.e. any element $A \in \mathsf{D}$ can be written as convex combination of extreme points.

The normalization of a covariant POVM, given by Eq. (8.44), can be rewritten in a simple form. In fact, due to the invariance of the Haar measure $\mathrm{d}g$, we have $[\mathcal{L}(A), U_g] = 0 \quad \forall g \in \mathbf{G}$, i.e. $\mathcal{L}(A)$ belongs to the commutant of $\{U_g\}$. Then, by exploiting Eq. (2.17), we can rewrite the normalization constraint (8.44) as

$$\sum_{i\in\mathcal{I}} \mathrm{Tr}[T^\mu_{kl} \ A_i] = d_\mu \ \delta_{kl} \qquad \forall \mu \in S, \quad \forall k,l = 1,\dots,m_\mu \ , \tag{8.46}$$

δ_{kl} denoting the Kronecker delta.

Again, this condition can be recasted into a compact form by introducing the auxiliary Hilbert space $\mathcal{H}_{aux} = \bigoplus_{i\in\mathcal{I}} \mathcal{W}_i$, with $\mathcal{W}_i \cong \mathcal{H} \quad \forall i \in \mathcal{I}$, and constructing a block operator with a repeated direct sum of the same operator T^{μ}_{kl}, i. e.

$$S^{\mu}_{kl} = \bigoplus_{i\in\mathcal{I}} S^{\mu}_{kli} \ , \qquad S^{\mu}_{kli} = T^{kl}_{\mu} \quad \forall i \in \mathcal{I}. \tag{8.47}$$

With this definition, Eq. (8.46) becomes

$$\mathrm{Tr}[S^{\mu}_{kl}\ A] = d_{\mu}\delta_{kl} \ , \qquad \forall \mu \in S, \quad \forall k,l = 1,\dots,m_{\mu} \ , \tag{8.48}$$

where A is the block operator $A = \bigoplus_{i\in\mathcal{I}}\ A_i$.

8.3.2 Extremal covariant POVMs with trivial stability groups

The characterization of the extremal covariant POVMs with probability space $\mathcal{I} \times \mathbf{G}$ can be given by exploiting the one-to-one affine correspondence between the convex set of covariant POVMs and the convex set C of block operators defined by the constraints (8.43) and (8.44), or, equivalently, by (8.43) and (8.48).

Definition 18. *A Hermitian block operator $P = \bigoplus_{i\in\mathcal{I}}\ P_i$ is a perturbation of $A \in \mathsf{D}$ if there exists an $\epsilon > 0$ such that $A + tP \in \mathsf{D}$ for any $t \in [-\epsilon, \epsilon]$.*

Recall that a point $A \in \mathsf{D}$ is extremal if and only if it admits only the trivial perturbation $P = 0$.

Lemma 14. *A block operator $P = \bigoplus_{i\in\mathcal{I}}\ P_i$ is a perturbation of $A \in \mathsf{D}$ if and only if*

$$\mathrm{Supp}(P) \subseteq \mathrm{Supp}(A) \tag{8.49}$$

$$\mathrm{Tr}[S^{\mu}_{kl}\ P] = 0 \qquad \forall \mu \in S, \quad \forall k,l = 1,\dots,m_{\mu} \tag{8.50}$$

Proof. According to Lemma 11, Condition (8.49) is equivalent to the existence of an $\epsilon > 0$ such that $A + tP \geq 0$ for all $t \in [-\epsilon, \epsilon]$. On the other hand, condition (8.50) is equivalent to require that $A + tP$ satisfies the normalization constraint (8.46) for all $t \in [-\epsilon, \epsilon]$. ■

Observation. Note that, due to the block form of both P and A, condition (8.49) is equivalent to

$$\mathrm{Supp}(P_i) \subseteq \mathrm{Supp}(A_i) \qquad \forall i \in \mathcal{I} \ . \tag{8.51}$$

Using the previous lemma, we can obtain a first characterization of extremality:

Theorem 22 (minimal support condition). *A point $A \in \mathsf{D}$ is extremal if and only if for any $B \in \mathsf{D}$,*

$$\mathrm{Supp}(B) \subseteq \mathrm{Supp}(A) \implies A = B \ . \tag{8.52}$$

Proof. Identical to the proof of Theorem 16. ■

Corollary 8. *If $A \in \mathsf{D}$ and $\mathrm{rank}(A) = 1$, then A is extremal.*

Proof. Since $\mathrm{rank}(A) = 1$, then, for any $B \in \mathsf{D}$, the condition $\mathrm{Supp}(B) \subseteq \mathrm{Supp}(A)$ implies $B = \lambda A$ for some $\lambda > 0$. Moreover, since both A and B are in D, from Eq. (8.48) we have $d_\mu = \mathrm{Tr}[S^\mu_{kk} B] = \lambda \mathrm{Tr}[S^\mu_{kk} A] = \lambda d_\mu$, whence necessarily $\lambda = 1$. Condition (8.52) then ensures that A is extremal. ■

A deeper characterization of extremal covariant POVMs can be obtained by using the following lemma.

Lemma 15. *Let A be a point of D, represented as*

$$A = \bigoplus_{i \in \mathcal{I}} X_i^\dagger X_i \ , \tag{8.53}$$

and define $\mathcal{H}_i = \mathrm{Rng}(X_i)$ the range of X_i. A block operator $P = \bigoplus_{i\in\mathcal{I}} P_i$ is a perturbation of A if and only if

$$P_i = X_i^\dagger \ Q_i \ X_i \qquad \forall i \in \mathcal{I} \ , \tag{8.54}$$

for some Hermitian $Q_i \in \mathcal{B}(\mathcal{H}_i)$, and

$$\sum_{i\in\mathcal{I}} \mathrm{Tr}[S^\mu_{kli} \ X_i^\dagger Q_i X_i] = 0 \ . \tag{8.55}$$

Proof. First of all, the form (8.54) is equivalent to condition (8.49). In fact, if P has the form (8.54), then clearly $\mathrm{Supp}(P) \subseteq \mathrm{Supp}(A)$. Viceversa, if we assume condition (8.49) and write $P = \bigoplus_{i\in\mathcal{I}} \ P_i$, we have necessarily $\mathrm{Supp}(P_i) \subseteq \mathrm{Supp}(X_i^\dagger X_i) = \mathrm{Supp}(X_i)$. Exploiting the singular value decomposition $X_i = \sum_{n=1}^{r_i} \lambda_n^{(i)} |w_n^i\rangle\langle v_n^i|$, where $\{|v_n^i\rangle\}$ and $\{|w_n^i\rangle\}$ are orthonormal bases for $\mathrm{Supp}(X_i)$ and $\mathrm{Rng}(X_i)$ respectively, we have that any Hermitian operator P_i satisfying $\mathrm{Supp}(P_i) \subseteq \mathrm{Supp}(X_i)$ has the form $P_i \ = \ \sum_{m,n} p^{(i)}_{mn} |v_m\rangle\langle v_n|$, whence it can be written as $P_i = X_i^\dagger Q_i X_i$, for some suitable Hermitian operator $Q_i \in \mathcal{B}(\mathrm{Rng}(X))$. Once the equivalence between the form (8.54) and condition (8.49) is established, relation (8.55) follows directly from Eq. (8.50). ■

Observation: According to the previous lemma, a perturbation of A is completely specified by a set of Hermitian operators $\{Q_i \in \mathcal{B}(\mathcal{H}_i) \mid i \in \mathcal{I}\}$, where $\mathcal{H}_i = \mathrm{Rng}(X_i)$. Such operators can be casted into a single block operator $Q \in \bigoplus_{i \in \mathcal{I}} \mathcal{B}(\mathcal{H}_i)$ by defining

$$Q = \bigoplus_{i \in \mathcal{I}} Q_i \ . \tag{8.56}$$

In terms of the block operator Q we have the following:

Lemma 16. *Let $A = \bigoplus_{i \in \mathcal{I}} X_i^\dagger X_i$ be a point of D. Define the block operators*

$$F_{kl}^{\mu} = \bigoplus_{i \in \mathcal{I}} X_i \ S_{kli}^{\mu} X_i^\dagger \ . \tag{8.57}$$

Then A admits a perturbation if and only if there exists an Hermitian block operator $Q \in \bigoplus_{i \in \mathcal{I}} \mathcal{B}(\mathcal{H}_i)$ such that

$$\mathrm{Tr}[F_{kl}^{\mu} Q] = 0 \ , \qquad \forall \mu \in \mathsf{S}, \ \forall k, l = 1, \dots, m_\mu \ . \tag{8.58}$$

Proof. Using the definition of F_{kl}^{μ} and the cyclic property of the trace, it is immediate to see the Eq. (8.58) is equivalent to Eq. (8.55). ■

The previous lemma enables us to characterize the extremal points of D.

Theorem 23 (Spanning set condition). *Let be $A = \bigoplus_{i \in \mathcal{I}} X_i^\dagger X_i$ be a point of D, and $\mathsf{F} = \{F_{kl}^{\mu} \mid \mu \in \mathsf{S}, k, l = 1, \dots, m_\mu\}$ be the set of block operators defined in Lemma 16. Then, A is extremal if and only if*

$$\mathrm{Span}(\mathsf{F}) = \bigoplus_{i \in \mathcal{I}} \mathcal{B}(\mathcal{H}_i) \ , \tag{8.59}$$

where $\mathcal{H}_i = \mathrm{Rng}(X_i)$.

Proof. A is extremal iff it admits only the trivial perturbation $P = 0$. Equivalently, due to Lemma 16, A is extremal iff the only Hermitian operator $Q \in \bigoplus_{i \in \mathcal{I}} \mathcal{B}(\mathcal{H}_i)$ that satisfies Eq. (8.58) is the null operator $Q = 0$. Let us decompose the Hilbert space $\mathcal{K} = \bigoplus_{i \in \mathcal{I}} \mathcal{B}(\mathcal{H}_i)$, as $\mathcal{K} = \mathrm{Span}(\mathsf{F}) \oplus \mathrm{Span}(\mathsf{F})^\perp$, where $\perp$ denotes the orthogonal complement with respect to the Hilbert-Schmidt product $(A, B) = \mathrm{Tr}[A^\dagger B]$. Then, A is extremal iff the only Hermitian operator in $\mathrm{Span}(\mathsf{F})^\perp$ is the null operator. This is equivalent to the condition $\mathrm{Span}(\mathsf{F})^\perp = \{0\}$, i.e. $\mathcal{K} = \mathrm{Span}(\mathsf{F})$. ■

Corollary 9. *Let $A = \bigoplus_{i \in \mathcal{I}} X_i^\dagger X_i$ be a point of D, and let define $r_i = \mathrm{rank}\, X_i$. If A is extremal, then the following relation holds*

$$\sum_{i \in \mathcal{I}} r_i^2 \leq \sum_{\mu \in \mathsf{S}} \ m_\mu^2 \ . \tag{8.60}$$

Proof. For an extreme point of D, relation (8.59) implies that the cardinality of the set F is greater than the dimension of $\mathcal{K} = \bigoplus_{i\in\mathcal{I}} \mathcal{B}(\mathcal{H}_i)$. Then, the upper bound (8.60) follows from $\dim \mathcal{K} = \sum_{i\in\mathcal{I}} r_i^2$ and from the fact that $|\mathsf{F}| = \sum_{\mu\in\mathsf{S}} m_\mu^2$. $\blacksquare$

Observation. If the group-representation $\{U_g\}$ is irreducible, than its Clebsch-Gordan decomposition contains only one term $\bar{\mu}$ with multiplicity $m_{\bar{\mu}} = 1$. Then, bound (8.60) becomes $\sum_{i\in\mathcal{I}} r_i^2 \leq 1$, namely for an extremal $A = \bigoplus_{i\in\mathcal{I}} A_i$, one has necessarily $\mathrm{rank}(A_{i_0}) = 1$ for some $i_0 \in \mathcal{I}$, and $A_i = 0$ for any $i \neq i_0$ (this is also a sufficient condition, due to Corollary 8). In terms of the corresponding covariant POVM $M(i,g) = U_g \; A_i \; U_g^\dagger$, one has $M(i,g) = 0$ for any $i \neq i_0$, i.e. the outcomes in the orbits $\mathcal{O}_i$ with $i \neq i_0$ correspond to null operators, namely they have zero probability for any state.

8.3.3 Extremal covariant POVMs with nontrivial stability groups

In the previous Subsection, we obtained a characterization of extremal covariant POVMs whose probability space is $\Omega = \mathcal{I} \times \mathbf{G}$ for some finite index set $\mathcal{I}$. The framework we outlined is suitable for a straightforward generalization to the case $\Omega = \cup_{i\in\mathcal{I}} \mathbf{G}/\mathbf{G}_i$, where $\mathbf{G}_i$ are compact subgroups of $\mathbf{G}$.

In this case, according to Proposition 9 of Chapter 1, it is possible to show that a covariant POVM P admits a density $M(x_i)$ such that for any measurable subset $B \subseteq \mathbf{G}/\mathbf{G}_i$ one has $P(B) \equiv P_i(B) \doteq \int_{B_i} \mathrm{d}x_i M(x_i)$, where $\mathrm{d}x_i$ is the group invariant measure on $\mathbf{G}/\mathbf{G}_i$. The form of the operator density is now

$$M(x_i) = U_{g_i(x_i)} \; A_i \; U_{g_i(x_i)}^\dagger \; , \tag{8.61}$$

where $A_i \geq 0$, and $g_i(x_i) \in \mathbf{G}$ is any representative element of the equivalence class $x_i \in \mathbf{G}/\mathbf{G}_i$. The normalization of the POVM is still given by Eq. (8.46). In addition, in order to remove the dependence of $M(x_i)$ from the choice of the representative $g_i(x_i)$, each operator A_i must satisfy the relation

$$[A_i, U_h] = 0 \quad \forall h \in \mathbf{G}_i \; . \tag{8.62}$$

The commutation constraint (8.62) can be simplified by decomposing each representation $\mathsf{R}(\mathbf{G}_i) = \{U_h \mid h \in \mathbf{G}_i\}$ into irreps

$$U_h = \bigoplus_{\nu\in\mathsf{S}_i} U_h^{\nu_i} \; \otimes \; 1\!\!1_{m_{\nu_i}} \; , \tag{8.63}$$

where m_{ν_i} denotes the multiplicity of the irrep ν_i, and S_i denotes the collection of all irreps contained in the decomposition of $\mathsf{R}(\mathbf{G}_i)$. This corresponds to the decomposition of the Hilbert space $\mathcal{H}$ as

$$\mathcal{H} = \bigoplus_{\nu_i \in \mathsf{S}_i} \mathcal{H}_{\nu_i} \otimes \mathbb{C}^{m_{\nu_i}} \ , \tag{8.64}$$

where $\mathcal{H}_{\nu_i}$ is a representation space, supporting the irrep ν_i, and $\mathbb{C}^{m_{\nu_i}}$ is a multiplicity space. In this decomposition, the commutation relation (8.62) is equivalent to the block form

$$A_i = \bigoplus_{\nu_i \in \mathsf{S}_i} \mathbb{1}_{\nu_i} \otimes A_{i,\nu_i} \ , \tag{8.65}$$

where $A_{i,\nu_i} \geq 0$ are operators acting on the multiplicity space $\mathbb{C}^{m_{\nu_i}}$.

By defining $\omega = (i, \nu_i)$ and $\Omega = \cup_{i \in \mathcal{I}} \ S_i$, we can introduce an auxiliary Hilbert space, and associate to a covariant POVM the block operator

$$A = \bigoplus_{\omega \in \Omega} A_\omega \ , \tag{8.66}$$

where $A_\omega \doteq A_{i,\nu_i}$. Furthermore, we define the block operators

$$S^\mu_{kl} = \bigoplus_{\omega \in \Omega} S^\mu_{kl\omega} \ , \tag{8.67}$$

where now $S_{kl\omega} = \mathrm{Tr}_{\mathcal{H}_{\nu_i}}[\Pi_{\nu_i} T^\mu_{kl}]$. Here Π_{ν_i} denotes the projector onto $\mathcal{H}_{\nu_i} \otimes \mathbb{C}^{m_{\nu_i}}$, and $\mathrm{Tr}_{\mathcal{H}_{\nu_i}}$ denotes the partial trace over $\mathcal{H}_{\nu_i}$. With these definitions, the normalization of the POVM, given by Eq. (8.46), becomes equivalent to

$$\mathrm{Tr}[S^\mu_{kl} \ A] = \delta_{kl} \ d_\mu \ . \tag{8.68}$$

Now we call $\widetilde{\mathsf{D}}$ the convex set of block operators $A = \bigoplus_{\omega \in \Omega} A_\omega$, defined by the two conditions $A \geq 0$ and Eq. (8.68). Such a convex set is in one-to-one affine correspondence with the convex set of covariant POVMs with probability space $\Omega = \cup_{i \in \mathcal{I}} \mathbf{G}/\mathbf{G}_i$. Since the constraints defining $\widetilde{\mathsf{D}}$ are formally the same defining the convex set D, we can exploit the characterization of extremal points of the previous section. In particular, Corollary 9 becomes

Corollary 10. *Let $A = \bigoplus_{\omega \in \Omega} X^\dagger_\omega X_\omega$ be a point of $\widetilde{\mathsf{D}}$, and define $r_{i,\nu_i} \equiv r_\omega = \mathrm{rank}(X_\omega)$. If A is extremal, then the following relation holds:*

$$\sum_{i \in \mathcal{I}} \sum_{\nu_i \in \mathsf{S}_i} r^2_{i,\nu_i} \leq \sum_{\mu \in \mathsf{S}} m^2_\mu \ . \tag{8.69}$$

Observation. As in the case of Corollary 9, if the representation $\{U_g\}$ is irreducible, as a consequence of the bound about ranks, one obtains $\mathrm{rank}(A_{\omega_0}) = 1$ for some $\omega_0 \in \Omega$, and $A_\omega = 0$ for any $\omega \neq \omega_0$.

8.4 Extremal POVMs and optimization problems

8.4.1 Minimization of the error probability

Consider a communication scheme where a finite alphabet $\mathcal{I}$ is encoded into a set of signal states $\{\rho_i \mid i \in \mathcal{I}\}$, and the symbol "$i$" is emitted with probability p_i. In this case, it is important to find the best decoding strategy in order to minimize the probability of error in the readout. This optimization problem is equivalent to find the POVM P_i that maximizes the probability of successful discrimination L, given by

$$L = \sum_{i \in \mathcal{I}} p_i \; p(i|i) \ , \tag{8.70}$$

with $p(i|j) = \mathrm{Tr}[P_i \rho_j]$.

As it was firstly observed by Holevo, there are cases in which the optimal estimation strategy is not represented by an ordinary observable, namely the operators P_i in the optimal POVM are not orthogonal projectors. Taking into account the Naǐmark theorem (Theorem 1 of Chapter 1), this means that there are situations in which the estimation can improved by performing a nonlocal projective measurement (an "ordinary" observable) on the state $\rho_i \otimes \rho_0$ where ρ_0 is a fixed state of an ancillary system. This interesting feature was called *quantum information openness* [28]. The study of extremal POVMs gives the possibility of systematically constructing examples where the optimal POVM is not an observable. In fact, it is simple to prove that *any* extremal POVM is is the unique optimal POVM for the discrimination of a suitable set of signal states:

Proposition 23. *Let P_i be a POVM for the Hilbert space $\mathcal{H} \cong \mathbb{C}^d$ with outcome space $\mathcal{I}$, Π_i the projector onto the support of P_i, and $r_i = \mathrm{Tr}[P_i]$. Then, the POVM P_i maximizes the probability $L = \sum_i p_i \ \mathrm{Tr}[\rho_i P_i]$ for the set of states $\{\rho_i = 1/r_i \ \Pi_i\}$ with a priori probabilities $p_i = r_i/R \quad R = \sum_i r_i$. Moreover, if P_i is extremal, then P_i is the unique POVM that maximizes $\langle L \rangle$.*

Proof. For any POVM Q_i one has

$$L = 1/R \sum_i \mathrm{Tr}[Q_i \Pi_i] \leq 1/R \sum_i \mathrm{Tr}[Q_i] = d/R \ . \tag{8.71}$$

Since the choice $P_i = Q_i$ achieves the bound, the POVM P_i is optimal. Moreover, suppose that P_i is extremal. For any POVM Q_i, the equality

$L = d/R$ holds only if $\mathrm{Tr}[Q_i\Pi] = \mathrm{Tr}[Q_i]$, i.e. only if $\mathrm{Supp}(Q_i) \subseteq \mathrm{Supp}(P_i)$ for any i. Therefore, one has $\mathrm{Supp}(Q) \subseteq \mathrm{Supp}(P)$, where P and Q are the block operators $P = \bigoplus_{i\in\mathcal{I}} P_i$ and $Q = \bigoplus_{i\in\mathcal{I}} Q_i$. Since P_i is extremal, due to the minimal support condition of Theorem 16, one must have $Q = P$, i.e. $Q_i = P_i \quad \forall i$. ■

Since the projector valued measures are just a subset of the set of extremal POVMs, the above Proposition enables one to find a large number of cases of discrimination in which the maximum success probability cannot be achieved by an "ordinary" observable. A well known example of this situation [91] is the discrimination of the three pure states

$$\rho_1 = 1/2 \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}, \quad \rho_2 = 1/2 \begin{pmatrix} 1 & \omega \\ \omega^* & 1 \end{pmatrix}, \quad \rho_3 = 1/2 \begin{pmatrix} 1 & \omega^2 \\ \omega^{2*} & 1 \end{pmatrix}, \tag{8.72}$$

where $\omega = e^{(2\pi i)/3}$. If the three states have the same a priori probability $p_i = 1/3$, then the unique optimal POVM is given by $P_i = 2/3\rho_i \; i = 1, 2, 3$.

8.4.2 Maximum likelihood approach in the presence of symmetry

The characterization of the extremal covariant POVMs sheds light on some interesting features occurring in optimization problems. As a prototype of optimization problem we will consider the maximization of the likelihood for family of states of the form $\{\rho_g = U_g \rho U_g^\dagger \mid g \in \mathbf{G}\}$ generated by the action of a finite or compact group on the input state ρ in the finite-dimensional Hilbert space $\mathcal{H}$. For simplicity, we will consider the case of trivial stability group, in which the states of the orbit are in one-to-one correspondence with the elements of the group. Accordingly, we will search the best covariant POVM $P(\mathrm{d}g) = U_g \Xi U_g^\dagger \, \mathrm{d}g$ with outcome space $\Omega = \mathbf{G}$ in order to maximize the likelihood $p(g|g) = \mathrm{Tr}[\Xi\rho]$.

Similarly to the case of arbitrary POVMs, where each extremal POVM is the unique optimal one for a suitable family of states, in the covariant case any extremal POVM is the unique optimal one for the estimation of the orbit generated by a suitable input state.

Proposition 24. *Let be Ξ the seed of a covariant POVM. Denote by P the projector onto* $\mathrm{Supp}(\Xi)$ *and by* $r = \mathrm{Tr}[P]$. *Then the seed Ξ maximizes the likelihood for the input state $\rho = \frac{P}{r}$. Moreover, if Ξ is extremal, then it is the unique seed that maximizes the likelihood for the input state $\rho = \frac{P}{r}$.*

Proof. For any arbitrary seed ζ, the following bound holds:

$$L_\rho[\zeta] = \mathrm{Tr}[\rho\zeta] = \frac{\mathrm{Tr}[P\zeta]}{r} \leq \frac{\mathrm{Tr}[\zeta]}{r} = \frac{\dim(\mathcal{H})}{r}, \tag{8.73}$$

where the equality $\mathrm{Tr}[\zeta] = \dim(\mathcal{H})\ \forall\zeta$ is due to the normalization constraints (3.17). Clearly Ξ achieves this bound, then it is optimal. Moreover, suppose that Ξ is extremal. Notice that the inequality $\mathrm{Tr}[P\zeta] \leq \mathrm{Tr}[\zeta]$ becomes equality if and only of $\mathrm{Supp}(\zeta) \subseteq \mathrm{Supp}(\Xi)$, then, using Proposition (22), we can see that Ξ represents the unique optimal POVM. ■

Consider now a density matrix σ with support in the orthogonal complement of $\mathrm{Supp}(\Xi)$ and take the randomization

$$\rho = (1-\alpha)\frac{P}{r} + \alpha\sigma\ , \tag{8.74}$$

where $0 \leq \alpha \leq 1$ is some probability. In the following we prove that, for sufficiently small α, Ξ represents still the optimal POVM in maximum likelihood sense. In other words, in this case the extremal POVM represented by Ξ is stable under randomization and the same measuring apparatus can be used in principle for a larger class of mixed states.

Proposition 25. *Consider the randomized state ρ in (8.74) and denote by $\bar{q}$ the maximum eigenvalue of σ. If $\alpha < \frac{1}{1+r\bar{q}}$, then Ξ is the unique seed which maximizes the likelihood for the state ρ of Eq. (8.74).*

Proof. Let us denote by Q the projection onto $\mathrm{Supp}(\sigma)$, then the following bound holds for any seed ζ:

$$\begin{aligned}
\mathcal{L}_\rho[\zeta] &= \frac{(1-\alpha)}{r}\mathrm{Tr}[P\zeta] + \alpha\mathrm{Tr}[\sigma\zeta] && (8.75)\\
&\leq \frac{(1-\alpha)}{r}\mathrm{Tr}[P\zeta] + \alpha\bar{q}\mathrm{Tr}[Q\zeta] && (8.76)\\
&\leq \frac{(1-\alpha)}{r}\mathrm{Tr}[(P+Q)\zeta] && (8.77)\\
&\leq \frac{(1-\alpha)}{r}\mathrm{Tr}[\zeta] = \frac{(1-\alpha)}{r}\dim(\mathcal{H})\ . && (8.78)
\end{aligned}$$

This bound is achieved by Ξ, proving optimality. Notice that Ξ is the unique optimal seed: namely the equality in (8.77) is attained if and only if $\mathrm{Tr}[Q\zeta] = 0$, that is when $\mathrm{Supp}(Q) \subseteq \mathrm{Ker}(\zeta)$, while in (8.78) equality is attained if and only if $\mathrm{Supp}(\zeta) \subseteq \mathrm{Supp}(P) \oplus \mathrm{Supp}(Q)$. Therefore the bound is achieved if and only if $\mathrm{Supp}(\zeta) \subseteq \mathrm{Supp}(P) = \mathrm{Supp}(\Xi)$, implying $\zeta = \Xi$.

8.4.3 Maximization of the mutual information in the presence of symmetry

A frequent problem in quantum communication is to find the POVM P_i, $i \in \mathcal{I}$, that maximizes the mutual information for a given set of signal states

$\mathcal{S} = \{\rho_j \mid j \in \mathcal{J}\}$, each of them emitted with probability p_j. Denoting by $p(i,j) = p_j \mathrm{Tr}[M_i \rho_j]$ the joint probability of the outcome j with the state ρ_i, and by $p_1(i)$ and $p_2(j)$ its marginals, the mutual information is defined as

$$I = H[p] - H[p_1] - H[p_2] \ , \tag{8.79}$$

where $H[q] \doteq \sum_i -q_i \log q_i$ is the Shannon entropy. According to Subsection 3.3.3, when the set of signal states is invariant under the action of some finite group $\mathbf{G}$ and all states in the same group orbit have the same probability, one can without loss of generality restrict the search for the optimal POVM among covariant POVMs with probability space $\Omega = \mathcal{I} \otimes \mathbf{G}$, for some finite index set $\mathcal{I}$[31, 90]. However, differently from the case of state discrimination, the points of the probability space do not need to be in one-to-one correspondence with the signal states. Therefore, the set $\mathcal{I}$ is not specified *a priori.*

Combining our characterization of extremal covariant POVMs with two basic properties of the mutual information (for the proofs, see Ref.[31]), we can readily obtain a bound about the cardinality of the index set $\mathcal{I}$. The two mentioned properties are

- The mutual information is a convex functional of the POVM.
- In the maximization of the mutual information, one can consider without loss of generality POVMs made of rank-one operators.

Consider a covariant POVM $P(i,g) = \frac{1}{|\mathbf{G}|} U_g A_i U_g^\dagger$. Due to convexity, we can we can restrict the attention to extremal covariant POVMs. Then, from Corollary 9, we have the bound $\sum_{i \in \mathcal{I}} \mathrm{rank}(A_i)^2 \leq \sum_{\mu \in \mathsf{S}} m_\mu^2$. Since the POVM operators can be chosen with unit rank, we also have that the number of nonzero operators A_i must be smaller than $\sum_{\mu \in \mathsf{S}} m_\mu^2$. Therefore, we can assume without loss of generality

$$\mid \mathcal{I} \mid \leq \sum_{\mu \in \mathsf{S}} m_\mu^2 \ . \tag{8.80}$$

This provides an alternative derivation of the bound given in Ref.[90]. Finally, if the representation $\{U_g\}$ is irreducible, the bound gives $|\mathcal{I}| = 1$, namely the probability space is $\Omega \cong \mathbf{G}$, according to the classic result of [31].

Chapter 9

Epilogue: estimation vs distribution of information

The determination of optimal estimation strategies is not only useful for the assessment of the ultimate precision limits in the extraction of classical information encoded into quantum signals, but also for the investigation about the very nature of quantum information. In particular, quantum estimation is directly related to the distribution of information to a large number of users. An arbitrary quantum channel that distributes information to M users in a permutationally invariant way can be indeed approximated by a classical incoherent scheme where some parameters of the input system are estimated and, conditionally to the estimate, M copies of a suitable input state are produced. With an optimized choice of the estimation strategy, it is possible to implement a classical scheme where, from the point of view of each single user, the distance between the output of the original channel and the output of the approximating scheme asymptotically vanishes with the scaling $1/M$, or even faster.

9.1 Cloning and state estimation

Perhaps the most distinctive feature of quantum information is the fact that it is intrinsically private, as it cannot be arbitrarily copied and distributed to many users [92, 93]. The optimal cloning of pure states, corresponding to the channel that produces from N copies of an unknown quantum state the best possible approximation of $M > N$ copies, is the simplest example of distribution of quantum information to a number of users. In the literature, the problem of optimal cloning has been studied in many variants, especially in the presence of symmetry, i.e. when the set of states to be cloned is the group orbit of a given input state [94, 95, 96, 97, 98, 99], or, more generally, of a set of input states [100]. A common feature in all known examples is that implementing the optimal cloning requires a coherent interaction of the input copies with a set of ancillae, namely an information processing which is typically quantum. A classical incoherent scheme, where some information is extracted from the input systems via a measurement and the output copies are reprepared conditionally to the measurement outcome is generally suboptimal: In order to optimally clone an unknown state, it is generally better to let the system interact properly with the ancillae than to try to estimate its state.

In the cloning of pure states, it has been noted in all known examples that, when the number M of output copies becomes large, the performances of the optimal cloning tend to decrease, asymptotically reaching the value that they have in the incoherent classical scheme "estimate the input state + prepare in the output M copies of the estimated state". This fact led to the conjecture that asymptotically optimal cloning and optimal state estimation are two equivalent problems, in the sense that in the limit of infinite output copies the optimal cloning is achieved by the classical scheme "measure and prepare" [101, 102]. Recently, an argument for the proof of this result has been presented by Bae and Acìn in Ref. [103], using two results regarding entanglement sharing [104] and entanglement breaking channels [105], respectively. This argument regards symmetric cloning channels with infinite output copies, the word "symmetric" meaning here that the output states are invariant under permutations of the output Hilbert spaces. The line of demonstration consists in first proving that for infinite output copies a symmetric cloning, when restricted at the single copy level, must be an entanglement breaking channel, and than in using the fact that any entanglement breaking channel can be realized by a classical scheme "measure and prepare".

The argument of Ref. [103] proves that any symmetric cloning with $M = \infty$ output copies can be realized by a classical incoherent scheme "measure and prepare". This result is conceptually important, since it highlights the connection between the symmetry of the output states and the realization of a channel by a classical scheme. Nevertheless, it refers to a situation which is somewhat artificial, since a real experiment never produces an infinity of clones. Such an argument cannot be immediately adapted to treat the case of finite M and to assess the quality of the classical scheme as an approximation of quantum cloning. An open question then remains: How fast quantum cloning converges to state estimation?

In the following, we present a recent achievement about this issue [24], which proves that at the single-copy level for any cloning channel there is always a classical scheme that produces the same output states within an accuracy $1/M$.

9.2 Channels for symmetric distribution of information

Quantum cloning is a special example of channel that distributes of quantum information to many users. In the literature, different examples of distribution of information have been also considered such as, for example, quantum broadcasting [106], which is the analogue of cloning for mixed states, and superbroadcasting [107, 108, 109], which is a process where the information contained in a mixed state is distributed and, at the same time, the purity of the single user output states is increased. In the case of superbroadcasting of qubits, in Ref. [110] it has been noticed that the single copy fidelity for the quantum channel decreases with the number of users, asymptotically reaching the lower bound given by the fidelity of the optimal estimation of the Bloch vector direction. This example provides a generalization of the relation between cloning and state estimation to mixed states: The optimal superbroadcasting channel is asymptotically realized by a the measurement of physical parameters, followed by a suitable repreparation.

Here we will focus our attention on a general class of quantum channels that distribute information, i.e. of channels that transform states on a given input space $\mathcal{H}_{in}$ to states on the $M-$fold tensor product $\mathcal{H}_{out} = \mathcal{H}^{\otimes M}$, where $\mathcal{H}$ is the single-user output space. According to the general formalism of quantum mechanics, any such channel is described by a completely positive trace-preserving map $\mathcal{E} : \mathcal{S}(\mathcal{H}_{in}) \longrightarrow S(\mathcal{H}^{\otimes M})$ [111]. Moreover, we require the channels to distribute information symmetrically among all users, i.e. for any input state ρ on $\mathcal{H}_{in}$, the state $\mathcal{E}(\rho)$ is invariant under permutations of the M output spaces. We name a channel with the above properties a *channel for symmetric distribution of information* (*SDI-channel*, for short).

For an SDI-channel, invariance under permutations implies that any group of k users receive the same state

$$\rho_{out}^{(k)} = \mathrm{Tr}_{M-k}[\mathcal{E}(\rho)] \ , \tag{9.1}$$

Tr_n denoting partial trace over n output spaces, no matter which ones. In particular, each single user receives the same state $\rho_{out}^{(1)} = \mathrm{Tr}_{M-1}[\mathcal{E}(\rho)]$.

A special family in the class of SDI-channels is the one formed by the *classical SDI-channels*, that consist in measuring the input and broadcasting the measurement outcome, with each user preparing locally the same state conditionally to the measurement outcome. The classical SDI-channels have the special form

$$\widetilde{\mathcal{E}}(\rho) = \sum_i \text{Tr}[P_i \rho]\ \rho_i^{\otimes M}\ , \tag{9.2}$$

where the POVM operators $\{P_i\}$ represent the estimation strategy performed on the input ($P_i \geq 0, \quad \sum_i P_i = \mathbb{1}_{in}$), and ρ_i is the state prepared conditionally to the outcome i. The classical SDI-channels do not distribute genuine quantum information, but only only classical information—i.e. the information about the outcome "i"—,which is subsequently encoded in a quantum output system.

9.3 Distribution of information to many users

The aim of this Section is to show that any SDI-channel can be approximated by a classical SDI-channel, within an accuracy that increases with the number of output copies. The accuracy of the approximation will be quantified by the trace-norm distance

$$\|\rho_{out}^{(1)} - \tilde{\rho}_{out}^{(1)}\|_1 = \text{Tr}|\rho_{out}^{(1)} - \tilde{\rho}_{out}^{(1)}| \tag{9.3}$$

between the single user output states. This quantity is interesting since it governs the distinguishability of states [1]: The minimum error probability p_{err} in distinguishing between two equally probable states ρ_1 and ρ_2 is indeed given by

$$p_{err} = \frac{1}{2} - \frac{1}{4}\|\rho_1 - \rho_2\|_1\ , \tag{9.4}$$

and for small distances it approaches the random guess value $p_{err} = 1/2$. In our case, a small distance $\|\rho_{out}^{(1)} - \tilde{\rho}_{out}^{(1)}\|_1$ means that a single user have a little chance of distinguishing between the outputs of the two channels $\mathcal{E}$ and $\widetilde{\mathcal{E}}$ by any measurement on his local state. Similarly, the trace norm distance $\|\rho_{out}^{(k)} - \tilde{\rho}_{out}^{(k)}\|_1$ is a useful quantity to discuss the presence of multipartite entanglement in the output states of the channel $\mathcal{E}$: Since the state $\tilde{\rho}_{out}^{(k)}$ coming from $\widetilde{\mathcal{E}}$ in Eq.(9.2) is separable, a small distance means that any group of k users has a very little chance of detecting entanglement.

The approximation of SDI-channels can be simply derived from the invariance of their output states under permutations. Permutationally invariant states have been thoroughly studied in the research about quantum de Finetti Theorem [112], and especially in its finite versions[1] [114, 115], where

[1] The expression "finite quantum de Finetti theorem" is used rather commonly in the literature. However, it is worth stressing that the finite version of classical de Finetti theorem is not by de Finetti, but by Diaconis and Freedman (Ref.[113]).

the goal is to approximate a permutationally invariant state ρ on $\mathcal{H}^{\otimes M}$ with a mixture of identically prepared states $\tilde{\rho} = \sum_i \; p_i \rho_i^{\otimes M}$. In particular, as we will see in the following, the recent techniques of Ref. [115] provide a very useful tool to prove our results.

For simplicity, we first start by considering the special case of SDI-channel with output states in the totally symmetric subspace $\mathcal{H}_+^{\otimes M} \subset \mathcal{H}^{\otimes M}$, which is the case, for example, of the optimal cloning of pure states. In order to approximate channels we use the following version of finite quantum de Finetti Theorem, which is proved with the same techniques of Ref.[115], with a slight improvement of the bound given therein:

Lemma 17. *For any state ρ on $\mathcal{H}_+^{\otimes M} \subset \mathcal{H}^{\otimes M}$, consider the separable state*

$$\tilde{\rho} = \int \mathrm{d}\psi \; p(\psi) \; |\psi\rangle\langle\psi|^{\otimes M} \; , \tag{9.5}$$

where the probability distribution $p(\psi)$ is given by

$$p(\psi) = \mathrm{Tr}\left[\Pi_\psi \; \rho\right], \quad \Pi_\psi = d_M^+ \; |\psi\rangle\langle\psi|^{\otimes M}, \tag{9.6}$$

where $\mathrm{d}\psi$ denotes the normalized Haar measure over the pure states $|\psi\rangle \in \mathcal{H}$, and $d_M^+ = \dim(\mathcal{H}_+^{\otimes M})$. Then, one has

$$\|\rho^{(k)} - \tilde{\rho}^{(k)}\|_1 \leq 4 s_{M,k}, \quad s_{M,k} \doteq 1 - \sqrt{\frac{d_{M-k}^+}{d_M^+}} \; , \tag{9.7}$$

$\rho^{(k)}$ denoting the reduced state $\rho^{(k)} = \mathrm{Tr}_{M-k}[\rho]$.

Proof. The identity in the totally symmetric subspace $\mathcal{H}_+^{\otimes n} \subset \mathcal{H}^{\otimes n}$ can be written as

$$\mathbb{1}_n^+ = d_n^+ \int \mathrm{d}\psi \; P_n(\psi) \; , \tag{9.8}$$

where $P_n(\psi) = |\psi\rangle\langle\psi|^{\otimes n}$. Using Eq.(9.8) with $n = M - k$, we can write $\rho^{(k)} = d_+^{M-k} \int \mathrm{d}\psi \; \rho_k(\psi)$, where $\rho_k(\psi) = \mathrm{Tr}_{M-k}\left[\rho \; \mathbb{1}^{\otimes k} \otimes P_{M-k}(\psi)\right]$. On the other hand, the reduced state $\tilde{\rho}^{(k)}$ can be written as

$$\tilde{\rho}^{(k)} = d_M^+ \int \mathrm{d}\psi \; P_k(\psi) \; \rho_k(\psi) \; P_k(\psi) \; . \tag{9.9}$$

Then, the difference between $\rho^{(k)}$ and $\tilde{\rho}^{(k)}$, denoted by $\Delta^{(k)}$, is given by

$$\Delta^{(k)} = d_{M-k}^+ \int \mathrm{d}\psi \left[\rho_k(\psi) - \frac{d_M^+}{d_{M-k}^+} P_k(\psi) \rho_k(\psi) P_k(\psi)\right] .$$

Notice that the integrand on the r.h.s. has the form $A-BAB$, with $A(\psi)=\rho_k(\psi)$ and $B(\psi)=\sqrt{d^+_M/d^+_{M-k}}\; P_k(\psi)$. Using the relation

$$A-BAB = A(\mathbb{1}-B)+(\mathbb{1}-B)A-(\mathbb{1}-B)A(\mathbb{1}-B) \tag{9.10}$$

we obtain

$$\Delta^{(k)} = d^+_{M-k}\left(C+C^\dagger-D\right)\ , \tag{9.11}$$

where

$$C \ = \ \int \mathrm{d}\psi\ A(\psi)\,[\mathbb{1}-B(\psi)]\ , \tag{9.12}$$

$$D \ = \ \int \mathrm{d}\psi\ [\mathbb{1}-B(\psi)]\ A(\psi)\ [\mathbb{1}-B(\psi)]\ . \tag{9.13}$$

The operator C is easily calculated using the relation

$$\begin{aligned}\int \mathrm{d}\psi\ \rho_k(\psi)\ P_k(\psi) &= \int \mathrm{d}\psi \mathrm{Tr}_{M-k}[\rho\ P_M(\psi)]\\ &= \frac{\mathrm{Tr}_{M-k}[\rho]}{d^+_M} = \frac{\rho^{(k)}}{d^+_M}\ ,\end{aligned}$$

which follows from Eq. (9.8) with $n=M$. In this way we obtain $C=s_{M,k}/d^+_{M-k}\ \rho^{(k)}$. Since C is nonnegative, we have $\|C\|_1=\mathrm{Tr}[C]=s_{M,k}/d^+_{M-k}$. Moreover, due to definition (9.13) also D is nonnegative, then we have $\|D\|_1=\mathrm{Tr}[D]=\mathrm{Tr}[C+C^\dagger]$, as follows by taking the trace on both sides of Eq.(9.11). Thus, the norm of D is $\|D\|_1=2\|C\|_1$. Finally, taking the norm on both sides of Eq. (9.11), and using triangular inequality we get $\|\Delta^{(k)}\|\leq 4d^+_{M-k}\ \|C\|_1 = 4s_{M,k}$, that is bound (9.7). ■

Since the dimension of the totally symmetric subspace $\mathcal{H}_+^{\otimes n}$ is given by $d^+_n=\binom{d+n-1}{n}$, for $M\gg kd$ the ratio d^+_{M-k}/d^+_M tends to $1-\frac{k(d-1)}{M}$. Therefore, Lemma 17 yields

$$\|\rho^{(k)}-\tilde\rho^{(k)}\|_1 \leq \frac{2(d-1)k}{M}, \quad M\gg kd, \tag{9.14}$$

i.e. the distance between $\rho^{(k)}$ and the separable state $\tilde\rho^{(k)}$ vanishes as k/M.

With the above Lemma, we are ready to prove the approximation Theorem for SDI-channels with output in the totally symmetric subspace:

Theorem 24. *Any SDI-channel $\mathcal{E}$ with output states in the totally symmetric subspace $\mathcal{H}_+^{\otimes M}\subset\mathcal{H}^{\otimes M}$ can be approximated by a classical channel*

$$\widetilde{\mathcal{E}}(\rho) = \int \mathrm{d}\psi\ \mathrm{Tr}[P_\psi\rho]\ |\psi\rangle\langle\psi|^{\otimes M}\ , \tag{9.15}$$

where P_ψ is a quantum measurement ($P_\psi \geq 0$ and $\int \mathrm{d}\psi \; P_\psi = \mathbb{1}_{in}$). For large M, the accuracy of the approximation is

$$\|\rho_{out}^{(k)} - \tilde{\rho}_{out}^{(k)}\|_1 \leq \frac{2(d-1)k}{M}, \quad M \gg kd. \tag{9.16}$$

Proof. Consider the channel $\mathcal{E}_*$ in the Heisenberg picture, defined by the relation $\mathrm{Tr}[O\mathcal{E}(\rho)] = \mathrm{Tr}[\mathcal{E}_*(O)\rho]$ for any state ρ on $\mathcal{H}_{in}$ and for any operator O on $\mathcal{H}_{out}$. Since the channel $\mathcal{E}$ is trace-preserving, $\mathcal{E}_*$ is identity-preserving, namely $\mathcal{E}_*(\mathbb{1}_{out}) = \mathbb{1}_{in}$. Applying Lemma 1 to the output state $\rho_{out} = \mathcal{E}(\rho)$, we get $\tilde{\rho}_{out} = \int \mathrm{d}\psi \; \mathrm{Tr}[\Pi_\psi \mathcal{E}(\rho)] \; |\psi\rangle\langle\psi|^{\otimes M}$. Since $\mathrm{Tr}[\Pi_\psi\mathcal{E}(\rho)] = \mathrm{Tr}[\mathcal{E}_*(\Pi_\psi)\rho]$, by defining $P_\psi \doteq \mathcal{E}_*(\Pi_\psi)$, we immediately obtain that $\tilde{\rho}_{out} = \widetilde{\mathcal{E}}(\rho)$, with $\widetilde{\mathcal{E}}$ as in Eq. (9.15). The operators $\{P_\psi\}$ represent a quantum measurement on $\mathcal{H}_{in}$, since they are obtained by applying a completely positive identity-preserving map to Π_ψ, which is a measurement on $\mathcal{H}_{out}$. Finally, the bound (9.16) then follows from Eq. (9.14). ■

The above Theorem proves that for large M the quantum information distributed to a single user can be efficiently replaced by the classical information about the measurement outcome ψ. In fact, the single user output states of the channels $\mathcal{E}$ and $\widetilde{\mathcal{E}}$ become closer and closer—and therefore less distinguishable—as M increases. For large M, the error probability in distinguishing between $\rho_{out}^{(1)}$ and $\tilde{\rho}_{out}^{(1)}$ has to satisfy the bound

$$p_{err} \geq \frac{1}{2} - \frac{d-1}{2M} \;, \tag{9.17}$$

namely it approaches 1/2 at rate M^{-1}. For example, for qubits Eq. (9.17) gives already with $M = 10$ an error probability $p_{err} \geq 0.45$, quite close to the error probability of a purely random guess. More generally, the bound (9.16) implies that for any group of k users there is almost no entanglement in the state $\rho_{out}^{(k)}$, since it is close to a completely separable state. As the number of users grows, multipartite entanglement vanishes at any finite order: only k-partite entanglement with $k = \mathcal{O}(M)$ can survive.

Theorem 24 shows in particular that quantum cloning can be approximated via quantum state estimation, thus providing the generalization of the result of Ref. [103] to the case of finite number of output copies.

All results obtained for SDI-channels with output in the totally symmetric subspace can be easily extended to arbitrary SDI-channels, exploiting the fact that any permutationally invariant state can be purified to a totally symmetric one [115]:

Lemma 18. *Any permutationally invariant state ρ on $\mathcal{H}^{\otimes M}$ can be purified to a state $|\Phi\rangle \in \mathcal{K}_+^{\otimes M} \subset \mathcal{K}^{\otimes M}$, where $\mathcal{K} = \mathcal{H}^{\otimes 2}$.*

Once the state ρ has been purified, we can apply Lemma 17 to the state $|\Phi\rangle$, thus approximating its reduced states. The reduced states of ρ are then obtained by taking the partial trace over the ancillae used in the purification. This implies the following

Lemma 19. *For any permutationally invariant state ρ on $\mathcal{H}^{\otimes M}$, purified to $|\Phi\rangle \in \mathcal{K}_+^{\otimes M}$, $\mathcal{K} = \mathcal{H}^{\otimes 2}$, consider the separable state*

$$\tilde{\rho} = \int \mathrm{d}\Psi \ p(\Psi) \ \rho(\Psi)^{\otimes M} \tag{9.18}$$

where $\mathrm{d}\Psi$ is the normalized Haar measure over the pure states $|\Psi\rangle \in \mathcal{K}$, $\rho(\Psi)$ is the reduced state $\rho(\Psi) = \mathrm{Tr}_{\mathcal{H}}\left[|\Psi\rangle\langle\Psi|\right]$, and $p(\Psi)$ is the probability distribution given by $p(\Psi) = \mathrm{Tr}[\Pi_\Psi |\Phi\rangle\langle\Phi|]$, with $\Pi_\Psi = D_M^+ \ |\Psi\rangle\langle\Psi|^{\otimes M}$, $D_M^+ = \dim(\mathcal{K}_+^{\otimes M})$. Then, one has

$$\|\rho_{k,A} - \tilde{\rho}_{k,A}\|_1 \leq 4 S_{M,k}, \quad S_{M,k} \doteq 1 - \sqrt{\frac{D_{M-k}^+}{D_M^+}} \ . \tag{9.19}$$

Proof. Applying Lemma 1 to $\tau = |\Phi\rangle\langle\Phi|$, we get the state

$$\tilde{\tau} = \int \mathrm{d}\Psi \ p(\Psi) \ |\Psi\rangle\langle\Psi|^{\otimes M} \ . \tag{9.20}$$

The state $\tilde{\rho}$ is then obtained by tracing out the ancillae used in the purification, namely it is given by Eq.(9.18). Since partial traces can only decrease the distance, the bound (9.19) immediately follows from the bound (9.7). ■

It is then immediate to obtain the following:

Theorem 25. *Any SDI-channel $\mathcal{E}$ can be approximated by a classical channel*

$$\widetilde{\mathcal{E}}(\rho) = \int \mathrm{d}\Psi \ \mathrm{Tr}[P_\Psi \rho] \ \rho(\Psi)^{\otimes M} \ , \tag{9.21}$$

where P_Ψ is a quantum measurement, namely $P_\Psi \geq 0$ and $\int \mathrm{d}\Psi \ P_\Psi = \mathbb{1}_{in}$. For large M, the accuracy of the approximation is

$$\|\rho_{out}^{(k)} - \tilde{\rho}_{out}^{(k)}\|_1 \leq \frac{2(d^2-1)k}{M}, \quad M \gg kd^2. \tag{9.22}$$

This Theorem extends Theorem 24 and all its consequences to the case of arbitrary SDI-channels. In particular, it proves that asymptotically the optimal cloning of mixed state can be efficiently simulated via mixed states estimation. The results of the measurement P_Ψ are indeed in correspondence with pure states on $\mathcal{H} \otimes \mathcal{H}$, and, therefore, with mixed states on $\mathcal{H}$. Accordingly, the knowledge of the classical result Ψ is enough to reproduce efficiently the output of the optimal cloning machine.

Notice the dependence on the dimension of the single user's Hilbert space in both Theorems 24 and 25: increasing d makes the bounds (9.16) and (9.22) looser, leaving more room to cloning/broadcasting of genuine quantum nature. Rather surprisingly, instead, the efficiency of the above approximations *does not depend* on the dimension of the full input Hilbert space, e. .g. it doesn't depend on the number N of the input copies of an SDI-channel. No matter how large is the physical system carrying the input information, if there are many users at the output there is no advantage of quantum over classical information processing. Accordingly, the previous results can be applied to channels from $\mathcal{H}^{\otimes N}$ to $\mathcal{H}^{\otimes M}$, even with $M < N$. As long as $M \gg kd^2$ any such channel can be efficiently replaced by a classical one. In particular, this argument holds also for the purification of quantum information [117, 118]: if M is enough large, any strategy for quantum purification can be approximated by a classical measure-and-prepare scheme. Only for small M one can have a genuine quantum purification.

In this Chapter we have considered the general class of quantum channels that equally distribute information among M users, showing that for large M any such channel can be efficiently approximated by a classical one, where the input system is measured, the measurement outcome is broadcast, and each user prepares locally the same state accordingly. The approximating channel can be regarded as the concatenation of a *quantum-to-classical* channel (the measurement), followed by a *classical-to-quantum* channel (the local preparation). Actually, the latter channel is needed only for the sake of comparison with the original quantum transformation to be approximated, since, due to the data processing inequality, this additional stage can only decrease the amount of information contained in the classical probability distribution of measurement outcomes. Therefore, asymptotically, there is no genuine distribution of quantum information, but just an announcement of the classical information extracted by a measurement. In other words, we cannot distribute more information about a quantum system than what we are able to read out via quantum estimation.

Chapter 10

Conclusions

The aim of this presentation was to give a systematic treatment about the estimation of physical parameters identifying the action of a given symmetry group. The problem has an exceedingly broad spectrum of applications in the fields of quantum communication and cryptography, quantum metrology, high-precision interferometry, and in many others. This motivated the search for a general point of view, able to give a clear insight into a variety of particular cases.

The research program of this presentation started from the derivation in Chapter 4 of the optimal estimation strategies and of the optimal signal states for the estimation of an unknown group transformation in the maximum likelihood approach. In the sense used in the presentation, the maximum likelihood approach consists in the maximization of the probability (probability density for continuous groups) that the estimated transformation coincides with the true one. This approach shed light on a general structure which underlies optimal estimation strategies, such a structure being related to the so called representation and multiplicity spaces in the tensor product structure (TPS) induced by the group. The maximum likelihood analysis shows indeed that the probability (density) of a correct estimation is larger, the more the entanglement between representation spaces and multiplicity spaces is exploited. Accordingly, the optimal input states are a coherent superposition of maximally entangled states in the group-induced TPS.

As a kind of long example of application of the maximum likelihood approach, Chapter 5 treated in detail the discrimination of a finite number of unitary transformations (gates) that form a group representation. In this case, it was shown that, by applying the same unknown transformation for a sufficient number of times on an entangled input state, it is always possible to discriminate the gates with zero error probability. This situation of perfect discriminability arises when the tensor representation involved in the problem contains all the irreps of the given finite group, and for each irrep the multiplicity is larger than or equal to the dimension of the representation space. According to the maximum likelihood approach, perfect discrimination is obtained by testing the unknown gates on states that exploit the maximum possible amount of entanglement between each representation space and a subspace of the corresponding multiplicity space.

The results obtained in the maximum likelihood approach were extended to other physically meaningful criteria in Chapter 6. This was done by introducing a class of cost functions, that was named the "generalized Holevo class" and consists in invariant cost functions with negative Fourier coefficients. When optimality is defined as the minimization of the Bayes cost (or even of the worst case cost) with respect to any such function, the optimal measurement for the estimation is the one already obtained in the maximum likelihood approach. The optimal input states to test the unknown transformation are of the same form as in the maximum likelihood approach, i.e. a coherent superposition of maximally entangled states in the tensor product structure induced from the group, but with a dependence on the choice of the cost function that appears in the coefficients of the superposition. The general results of Chapter 6, which cover the case of any finite and compact group, are the core of the whole presentation.

Chapter 7 was devoted to the exploration of the estimation problem in the case of nonunimodular groups, for which a single Haar measure does not exists. This case, that was not covered neither by Chapter 4 nor by Chapter 6, brought to light some unexpected features which are characteristic for nonunimodular groups, such as the apparently paradoxical fact that in the maximum likelihood the true value is not the most likely one. The general results, here presented in the maximum likelihood approach, allowed one to treat the remarkable example of the joint estimation of displacement and squeezing in the radiation field, finding for suitable input states a remarkable relation between the uncertainties in the optimal joint estimation and the uncertainties in the optimal separate measurements.

The analysis of Chapter 7 closed the part of the thesis related to the explicit construction of optimal estimation strategies for the minimization of the average value of a given cost function (in the Bayesian and/or in the minimax approach). The generalization to different optimization criteria—such as the maximization of the mutual information—and to more complex setups—such that the estimation of signal states that form the union of different orbits—is a rather hard topic, whose solution is not possible analytically, or, at least, has not been possible up to now. Chapter 8 was devoted to the characterization of the extremal POVMs for finite dimensional quantum systems. This analysis serves as an intermediate step for the optimization in the mentioned cases, where the convexity of the figure of merit allows one to restrict the attention to the extreme points of suitable convex sets, such as the convex set of POVMs with given outcome space, or, in the presence of symmetry, the convex subset of covariant POVMs. Some connections between the characterization of the extremal POVMs and the solution of optimization problems have been presented, relating extremality with properties such as the uniqueness and the stability of the optimal measurements, and presenting some simple examples of state estimation with mixed states. Chapter 8 contains also a result which is interesting by itself, namely the fact that the extremal POVMs for a finite dimensional system have always a finite number of effective outcomes. This result suggests that, for convex optimization criteria, such as the minimization of the Bayes cost or the maximization of the mutual information, one can always find, if needed, an optimal POVM with a finite number of outcomes.

The presentation is concluded in Chapter 9 with a brief *excursus* in the field of quantum information processing, included here to provide, after the derivation of the optimized strategies of the Chapters 4, 6, and 7, an additional *a posteriori* motivation for the interest in the estimation of quantum signals. The presented results showed indeed that the estimation of suitable physical parameters can be used to approximate efficiently any quantum channel that distributes quantum information to a large number of users. This implies that, asymptotically, one cannot distribute more information about a quantum system than one is able to extract from it.

Chapter 11

Acknowledgments

This thesis stems from a very special and lucky context, which is the research activity of the Quantum Information Theory Group of Pavia University, led by Prof. Mauro D'Ariano. Completing the PhD studies within the group has been a great opportunity of enriching my education, due to the atmosphere of continuous discussion, to the presence of many seminars and small courses during the year, and to the possibility of attending international conferences. The first acknowledgment of the thesis is naturally for Mauro, for his enthusiasm in research, his ability in transmitting it, and his precious suggestions, and for all the people of the group with which I shared both research and friendship during my PhD studies: Massimiliano Sacchi and Paolo Perinotti (who greatly contributed to my training in research since the diploma thesis), Chiara Macchiavello, Francesco Buscemi, Lorenzo Maccone and Michael Keyl.

I would like to thank Stephen Bartlett, the official referee of this thesis for his very careful reading and his gratifying appreciation. His ability in going through all the details of the work is an admirable example of how the work of a referee should be. The final version of this presentation greatly benefited from his suggestions and from his encouragement.

The PhD period was a nice opportunity of exchanging ideas and points of view with several people around the world. Among them, I'm especially grateful to Martin Rötteler, for many nice discussions (some of them also appear in Chapter 5 of this thesis) and for his friendship. I also would like to thank Robert Spekkens for his encouragements in my research and for his gratifying interest in this thesis. Finally, it is a pleasure to thank Masahito Hayashi for many stimulating discussions and for his invitation to visit Erato Project during summer 2005, and also Laura Cattaneo and Sergio Albeverio for their invitation to Bonn University, where I enjoyed a nice atmosphere and pleasant discussions.

Concluding this thesis, I also would like to aknowledge all professors of the PhD courses I have attended in Pavia, for the opportunity of scientific enrichment they gave me. In particular, I'm grateful to Annalisa Marzuoli and Giorgio Zambotti, for their interest and for the pleasure of many discussions.

Grazie alla Roby, per la sua amicizia, il suo sincero affetto e la sua irresistibile simpatia. Un grande ringraziamento è riservato a Domenico, Francesco, Antonio, Alberto, Gabriele, Alessandro e Giovanni, i miei amici più cari in questi ultimi anni.

Grazie a Giada, per le nostre piacevoli chiacchierate e per la sua pazienza nell'ascoltare tutte le puntate di una tragicomica epopea, a Chiara e Francesca per la loro simpatia, le belle serate al Melograno e le raffinate degustazioni del pregiatissimo "Er Grei".

Un caloroso ringraziamento va poi tutti i compagni del condominio Orchidea—nell'ordine: Giovanni, Marco, Antonio, Stefano, Fidel, Paco Pegna, Lapo Gianni, Guglielmo di Rubruck, Jacopo Ortis e Giovanni da Pian del Carpine—agli esuberanti vicini, allo pterodattilo e alla bambina del piano di sopra.

Grazie ai miei genitori Edoardo e Gaia, per il loro affetto, il loro entusiasmo nei momenti belli e il loro sostegno in quelli meno belli. Ringraziamenti speciali sono poi riservati a Valentina, per i suoi illuminanti consigli, il suo modo impareggiabile di interpretare il ruolo di sorella, per il suo brio, la sua sensibilità, la sua pazienza, la sua fantasia, e numerose altre doti che richiederebbero molto spazio per essere elencate, e ad Alice, entusiasta mascotte della nostra famiglia e gelosa custode dell'unità del branco.

Grazie ai miei nonni di Mantova, Carlo e Santina, e di Salina, Saverio e Maria, per essere sempre così calorosamente affettuosi. Grazie all'*adorata tzia* Emma, alla Eva dal piè veloce, e a Diana la perugina.

Grazie a Valentina[1], tesoro dietro l'arcobaleno, per il suo amore che, entrato improvvisamente nella mia vita, la ha resa più bella di quanto io stesso avrei potuto immaginare.

[1] Come sarà sicuramente chiaro al lettore perspicace, il termine "Valentina" indica qui una persona diversa dall'impareggiabile sorella di più sopra.

Bibliography

[1] C. W. Helstrom, *Quantum detection and estimation theory* (Academic Press, New York, 1976).

[2] A. S. Holevo, *Probabilistic and Statistical Aspects of Quantum Theory* (North Holland, Amsterdam, 1982).

[3] M.A. Nielsen and I.L. Chuang, *Quantum computation and quantum informartion* (Cambridge University Press, 2000).

[4] S. D. Bartlett, T. Rudolph, and R. W. Spekkens, Phys. Rev. Lett. **91**, 027901 (2003).

[5] S. D. Bartlett, T. Rudolph, and R. W. Spekkens, Phys. Rev. A **70**, 032307 (2004).

[6] E. Bagan, M. Baig, A. Brey, R. Munoz-Tapia and R. Tarrach, Phys. Rev. Lett. **85**, 5230 (2000).

[7] A. Peres and P. F. Scudo, Phys. Rev. Lett. **86**, 4160 (2001).

[8] G. Chiribella, G. M. D'Ariano, P. Perinotti, and M. F. Sacchi, Phys. Rev. Lett. **93**, 180503 (2004).

[9] S. Bartlett, T. Rudolph, and R. W. Spekkens, quant-ph/0610030.

[10] V. Bužek, R. Derka, and S. Massar, Phys. Rev. Lett. **82**, 2207, (1999).

[11] V. Giovannetti, S. Lloyd, and L. Maccone, Nature **412**, 417 (2001).

[12] S. Helgason, *Differential Geometry, Lie Groups and Symmetric Spaces* (Academic Press, New York, 1978).

[13] D.P. Zhelobenko, *Compact Lie Groups and Their Representations* (American Mathematical Society, Providence, 1973).

[14] A. O. Barut and R. Raczka, *Theory of Group Representations and Applications* (World Scientific, 1986).

[15] G. Chiribella, G. M. D'Ariano, P. Perinotti, and M.F. Sacchi, Phys. Rev. A **70**, 061205 (2004).

[16] G. Chiribella, G. M. D'Ariano, P. Perinotti, and M.F. Sacchi, Int. J. Quant. Inf. **4**, 453 (2006).

[17] G. Chiribella, G. M. D'Ariano, and M. Rötteler, paper in preparation.

[18] G. Chiribella, G. M. D'Ariano, and M. F. Sacchi, Phys. Rev. A **72**, 042336 (2005).

[19] G. Chiribella, G. M. D'Ariano, and M. F. Sacchi, Phys. Rev. A **73**, 062103 (2006).

[20] E. Arthurs and J. L. Kelly, Bell Syst. Techn. Journ. **40**, 725 (1965).

[21] G. Chiribella, G. M. D'Ariano, and M. F. Sacchi, J. Phys. A , J. Phys. A **39**, 2127 (2006).

[22] G. Chiribella and G. M. D'Ariano, J. Math. Phys. **45**, 4435 (2004).

[23] G. Chiribella and G. M. D'Ariano, J. Math. Phys. **47**, 092107 (2006).

[24] G. Chiribella and G. M. D'Ariano, quant-ph/0608007.

[25] G. Ludwig, *Measuring and preparing process*, Lecture Notes in Phys. **29**, 122 (1974).

[26] A. S. Holevo, J. Multivariate Anal. **3**, 337 (1973).

[27] E. B. Davies and J. T. Lewis, *Comm. Math. Phys.* **17**, 239 (1970).

[28] A. S. Holevo, Trudy Moscow Mat. Obšč. **26**, 133 (1972), AMS. Trans (1974).

[29] J. von Neumann, *Mathematical Foundations of Quantum Mechanics* (Princeton University Press, Princeton, NJ, 1955).

[30] M. A. Naǐmark, Izv. Akad. Nauk SSSR, Ser. Mat. **3**, 277 (1940).

[31] E. B. Davies, IEEE Trans. Inform. Theory **24**, 596 (1978).

[32] P. Zanardi and M. Rasetti, Phys. Rev. Lett. **79**, 3306 (1998).

[33] D. A. Lidar, I. L. Chuang, and K. B. Whaley, Phys. Rev. Lett. **81**, 2594 (1998).

[34] E. Knill, R. Laflamme, and L. Viola, Phys. Rev. Lett. **84**, 2525 (2000).

[35] M. Rötteler and T. Beth, quant-ph/9812070.

[36] I. Daubechies, *Ten lectures on Wavelets* (SIAM, Philadelphia, 1992).

[37] A. Grossmann, J. Morlet, and T. Paul, J. Math. Phys. **26**, 10 (1985).

[38] E. P. Wigner, *Symmetries and reflections* (Indiana Univ. Press, Bloomington-London, 1970).

[39] E. B. Davies, *Quantum theory of open systems* (Acad. Press, London, 1976).

[40] M. Ozawa, in *Research Reports on Information Sciences, Series A: Mathematical Sciences, N. 74* (Department of Information Sciences Tokyo Institute of Technology, 1980).

[41] P. Hausladen and W. K. Wootters, J. Mod. Opt. **41**, 2385 (1994).

[42] J. Von Korff and J. Kempe, Phys. Rev. Lett. **93**, 260502 (2004).

[43] A. S. Holevo, in G. Maruyama and J. V. Prokhorov, *Proceeding of the Second Japan-USSR Symposium on Probability Theory*, Lecture notes in Mathematics, vol. 330 p. 104, (Springer-Verlag, Berlin, 1973).

[44] N. Gisin and S. Popescu, Phys. Rev. Lett. **83**, 432 (1999).

[45] N. J. Cerf and S. Iblisdir, Phys. Rev. A **64**, 032307 (2001).

[46] A. Acín, Phys. Rev. Lett. **87**, 177901 (2001).

[47] G. M. D'Ariano, P. Lo Presti, and M.G. A. Paris, Phys. Rev. Lett. **87**, 270404 (2001)

[48] R. Duan, Y. Feng, and M. Ying, eprint quant-ph/0601150.

[49] V. Giovannetti, S. Lloyd, L. Maccone, Phys. Rev. Lett. **96**, 010401 (2006).

[50] S. L. Braunstein, Nature **440**, 617 (2006).

[51] A. Chefles, Phys. Rev. A **65**, 052314, (2002).

[52] E. Knill, Los Alamos National Laboratory Report LAUR-96-2807 (1996).

[53] A. Klappenecker and M. R otteler, IEEE Trans. Inform. Theory, **48**, 2392 (2002).

[54] C. H. Bennett and S. J. Wiesner, Phys. Rev. Lett. **69**, 2881 (1992).

[55] P. W. Shor, p. 124 in *Proceedings pf the 35th Annual Symposium of the Foundations of Computer Science*, ed. S. Goldwasser (IEEE Computer Society Press, Los Alamitos, CA, 1994).

[56] L. K. Grover, p. 212 in *Proceedings, 28th Annual ACM Symposium on the Theory of Computing (STOC)*, (May 1996), eprint quant-ph/9605043; Phys. Rev. Lett. **79**, 325 (1997).

[57] C. H. Bennett, G. Brassard, C. Crepeau, R. Jozsa, A. Peres, and W. K. Wootters, Phys. Rev. Lett. **70**, 1895 (1993).

[58] A. K. Ekert, Phys. Rev. Lett. **67**, 661 (1991); Nature **358**, 14 (1992).

[59] V. Giovannetti, S. Lloyd, and L. Maccone, Science **306**, 1330 (2004).

[60] A. Acín, E. Jane, and G. Vidal, Phys. Rev. A **64**, 050302(R) (2001).

[61] A. Fujiwara, Phys. Rev. A **65**, 012316 (2001).

[62] M. A. Ballester, Phys. Rev. A **70**, 032310 (2004).

[63] P. Zanardi, Phys. Rev. Lett. **87** 077901 (2001).

[64] G. M. D'Ariano, C. Macchiavello, and M. F. Sacchi, Phys. Lett. A **248**, 103 (1998).

[65] A. S. Holevo, Rep. Math. Phys. **16**, 385 (1979).

[66] G. M. D'Ariano, P. Lo Presti, and M. F. Sacchi, Phys. Lett. A **272**, 32 (2000).

[67] G. Chiribella and G. M. D'Ariano, J. Math. Phys. **45**, 4435 (2004).

[68] G. M. D'Ariano, C. Macchiavello, and P. Perinotti, Phys. Rev. A **72**, 042327 (2005).

[69] V. Bužek, R. Derka, and S. Massar, in M. Hayashi, *Asymptotic theory of quantum statistical inference* (World Scientific, Singapore, 2005).

[70] M. Hayashi, Phys. Lett. A **354**, 183 (2006).

[71] E. Bagan, M. Baig, and R. Muñoz-Tapia, Phys. Rev. Lett. **87**, 257903 (2001).

[72] A. Peres and P. F. Scudo, J. Mod. Opt.**49**, 1235 (2002).

[73] J. Kahn, eprint quant-ph/0603115.

[74] G. Chiribella, L. Maccone, and P. Perinotti, quant-ph/0608042.

[75] M. A. Ballester, Phys. Rev. A **69**, 022303 (2004).

[76] A. S. Holevo, Rep. Math. Phys. **13**, 287 (1978).

[77] C. G. Bollini and L. E. Oxman, Phys. Rev. A **47**, 2339 (1993).

[78] L. Susskind and J. Glogower, Physics **1**, 49 (1964).

[79] G. J. Milburn, W.-Y. Chen, and K. R. Jones, Phys. Rev. A **50**, 801 (1994).

[80] M. Duflo and C. C. Moore, J. Funct. Anal. **21**, 209 (1976).

[81] A. L. Carey, Bull. Austral. Math. Soc. **15**, 1 (1976).

[82] U. Cattaneo, J. Math. Phys. **23**, 659 (1982).

[83] M. Ban, K. Kurukowa, R. Momose, and O. Hirota, Int. J. Theor. Phys. **36**, 1269 (1997).

[84] M. Sasaki, A. Carlini, and A. Chefles, J. Phys. A: Math. Gen. **34**, 7017 (2001).

[85] Y. C. Eldar and G. D. Forney, IEEE Trans. IT **47**, 858 (2001).

[86] R. T. Rockafeller, *Convex analysis* (Priceton University Press, Priceton, NJ, 1970).

[87] K. R. Parthasaraty, Inf. Dim. Anal. **2**, 557 (1999).

[88] E. Størmer, in A. Hartkamper and H. Neumann, *Foundations of Quantum Mechanics and Ordered Linear Spaces* (Springer, Berlin 1974).

[89] G. M. D'Ariano, P. Lo Presti, and P. Perinotti, J. Phys. A: Math. Gen. **38**, 5979 (2005).

[90] T. Decker, eprint quant-ph/0509122.

[91] A. S. Holevo, Probab. Math. Statist. **3**, 113 (1982).

[92] W. K. Wootters and W. H. Zurek, Nature **299**, 802 (1982).

[93] D. Dieks, Phys. Lett. A **92**, 271 (1982).

[94] V. Bužek and M. Hillery, Phys. Rev. A **54**, 1844 (1996).

[95] N. Gisin and S. Massar, Phys. Rev. Lett. **79**, 2153 (1997).

[96] R. F. Werner, Phys. Rev. A **58**, 1827 (1998).

[97] N. J. Cerf, A. Ipe, and X. Rottenberg, Phys. Rev. Lett. **85**, 1754 (2000).

[98] N. J. Cerf and S. Iblisdir, Phys. Rev. Lett. **87**, 247903 (2001).

[99] G. M. D'Ariano and C. Macchiavello, Phys. Rev. A **87**, 042306 (2003).

[100] G. Chiribella, G. M. D'Ariano, P. Perinotti, and N. J. Cerf, Phys. Rev. A **72**, 042336 (2005).

[101] N. Gisin and S. Massar, Phys. Rev. Lett. **79**, 2153 (1997); D. Bruß, A. Ekert and C. Macchiavello, Phys. Rev. Lett. **81**, 2598 (1998).

[102] M. Keyl, in http://www.imaph.tu-bs.de/qi/problems/.

[103] J. Bae and A. Acín, Phys. Rev. Lett. **97** 030402 (2006)

[104] R. F. Werner, Lett. Math. Phys. **17**, 359 (1989).

[105] M. Horodecki, P. W. Shor and M. B. Ruskai, Rev. Math. Phys **15**, 629 (2003).

[106] H. Barnum, C. M. Caves, C. A. Fuchs, Phys. Rev. Lett. **76** , 2818 (1996).

[107] G. M. D'Ariano, C. Macchiavello, and P. Perinotti, Phys. Rev. Lett. **95**, 060503 (2005).

[108] F. Buscemi, G. M. D'Ariano, C. Macchiavello, and P. Perinotti, Phys. Rev. A **74**, 042309 (2006).

[109] G. M. D'Ariano, P. Perinotti, and M. F. Sacchi, New J. Phys. **8**, 99 (2006).

[110] G. Chiribella, G. M. D'Ariano, C. Macchiavello, P. Perinotti, and F. Buscemi, Phys. Rev. A **74** (2006).

[111] K. Kraus, Lecture notes in Physics **190**, Springer-Verlag (1983).

[112] C. M. Caves, C. A. Fuchs, and R. Schack, J. Math. Phys. **43**, 4537 (2002).

[113] P. Diaconis and D. Freedman, Ann. Prob. **8**, 745 (1980).

[114] R. König and R. Renner, J. Math. Phys. **46**, 122108 (2005).

[115] M. Christandl, R. Koenig, G. Mitchison, and R. Renner, eprint quant-ph/0602130.

[116] M. Keyl and R. F. Werner, J. Math. Phys. **40**, 3283 (1999).

[117] J.I. Cirac, A.K. Ekert, and C. Macchiavello, Phys. Rev. Lett. **82**, 4344 (1999).

[118] M. Keyl and R.F. Werner, Annales Henri Poincaré **2**, 1 (2001).