

## MAXIMUM LIKELIHOOD ESTIMATION FOR A GROUP OF PHYSICAL TRANSFORMATIONS

GIULIO CHIRIBELLA\* and GIACOMO MAURO D'ARIANO†

*QUIT, Dipartimento di Fisica "A. Volta", Università di Pavia  
Via Bassi 6, Pavia, 27100, Italy*

*\*chiribella@fisicavolta.unipv.it*

*†dariano@unipv.it*

PAOLO PERINOTTI‡ and MASSIMILIANO F. SACCHI§

*QUIT, INFN, Via Bassi 6, Pavia, 27100, Italy*

*‡perinotti@fisicavolta.unipv.it*

*§msacchi@unipv.it*

Received 10 June 2005

The maximum likelihood strategy for the estimation of group parameters allows one to derive in a general fashion optimal measurements, optimal signal states, and their relations with other information theoretical quantities. These results provide deep insight into the general structure underlying optimal quantum estimation strategies. The entanglement between representation spaces and multiplicity spaces of the group action appears to be the unique kind of entanglement which is really useful for the optimal estimation of group parameters.

*Keywords:* Quantum estimation; group covariance; entanglement.

### 1. Introduction

Since the beginning of quantum estimation theory,<sup>1,2</sup> research on measurements and estimation strategies for the optimal detection of physical parameters has been a major focus. In particular, the case where the physical parameters to be estimated correspond to the unknown action of some symmetry group has received constant attention.<sup>3–8</sup> The reason for such interest, which in the last ten years has received strong motivation from the new field of quantum information, is the broad spectrum of applications of the area, ranging from quantum statistics to quantum cryptography, from the study of uncertainty relations to the design of high sensitivity measurements achieving the ultimate quantum limit.

Despite the long-standing attention to the problem, many new examples and applications of group parameter estimation have come up recently, and on the other hand, some controversial points have been clarified only in the very last few years.

In this rich variety of applications, it is somehow natural to look for a general point of view, suitable for capturing the main features of optimal estimation without entering the specific details of the symmetry group involved in the particular problem at hand.

In this paper, we will show how such general insight can be provided in a simple way in the *maximum likelihood approach*,<sup>9</sup> where measurements are designed to maximize the probability of estimating the true value of the unknown group parameter. In this approach, the relations between the quality of the estimation and other information theoretic properties, such as the Holevo  $\chi$ -quantity<sup>10</sup> and the dimension of the space spanned by a quantum state under the action of the group become straightforward.

The results of the maximum likelihood method recently allowed us to understand the crucial role of the equivalent irreducible representations of the group in the optimal estimation strategy,<sup>11</sup> giving a striking application of this mechanism with the solution of a long-standing controversy,<sup>12–16</sup> about the efficiency in the absolute transmission of a Cartesian reference frame. In that application, the technique of equivalent representations was the key idea for an efficient use of quantum resources.

Here, we will show in a general fashion that, far from being a technicality, the use of equivalent representations is synonymous to the use of the unique kind of entanglement which is really suitable for group parameter estimation. More precisely, such a kind of entanglement is the entanglement between spaces where the group acts irreducibly (*representation spaces*) and spaces where the group acts trivially (*multiplicity spaces*). This entanglement is related only to the group, without any relation with other “natural” tensor product structures that can be present in the system, e.g. when the system is made by a set of distinguishable particles.

The concept of representation/multiplicity is well known in the field of quantum error correction,<sup>17,18</sup> where the multiplicity spaces are referred to as *decoherence-free subspaces* and, more generally, *noiseless subsystems*. Moreover, the same concept has also recently found application in the context of quantum communication<sup>19</sup> and cryptography.<sup>20</sup> On the other hand, the application to quantum estimation of tools such as representation/multiplicity spaces and entanglement is a completely new issue.<sup>11,21</sup> In the maximum likelihood approach, the maximal entanglement between representation and multiplicity spaces appears to be the unifying feature of the optimal strategies for any group parameter estimation. Moreover, the validity of this result is not limited to the maximum likelihood estimation, and can be extended to estimation schemes where different figures of merit are considered.<sup>22</sup>

In Sec. 2, we will present the general approach to group parameter estimation. We will firstly start with a brief, self-contained introduction about group parameter estimation (Sec. 2.1), the maximum likelihood approach (Sec. 2.2), group theoretical tools (Sec. 2.3), and covariant measurements (Sec. 2.3). After that, we will derive in the general setting the optimal measurements (Sec. 2.5), and the optimal states (Sec. 2.6), emphasizing the role of equivalent representations and the relations with the Holevo  $\chi$ -quantity. Finally, we will conclude the section with a

discussion (Sec. 2.7) about the use of entanglement with an additional reference system, showing how to efficiently use such resource.

In Sec. 3, all the results will be generalized to infinite dimension (Sec. 3.1) and non-compact groups (Sec. 3.2). As an example of application, in Sec. 3.3, we provide a detailed analysis of the estimation of the joint displacement of two harmonic oscillators in phase space, underlying in this continuous variable context the connections with the well-known example by Gisin and Popescu about the estimation of a direction using parallel and anti-parallel spins.<sup>24</sup>

## 2. A General Approach to Group Parameter Estimation

### 2.1. Background

The issue of this paper is the problem of optimally discriminating a family of quantum signal states, which is the orbit generated by a given input state under the action of a group. In other words, if the input state is the density matrix  $\rho \in \mathcal{B}(\mathcal{H})$  on the Hilbert space  $\mathcal{H}$ , then we want to find the best estimation of the states in the orbit

$$\mathcal{O} = \{\rho_g = U_g \rho U_g^\dagger \mid g \in \mathbf{G}\}, \tag{1}$$

obtained by transforming the input state with the unitary representation  $\{U_g\}$  of the group  $\mathbf{G}$ .

In general, the points of the orbit are not in one-to-one correspondence with the elements of the group, since one can have  $\rho_{g_1} = \rho_{g_2}$  even for different  $g_1$  and  $g_2$ . However, in this paper we will consider for simplicity the case where the correspondence between the group and the orbit is one-to-one, since apart from a technical complication in the notation, the extension of the results to the general case is straightforward.

In the case of one-to-one correspondence between signal states and elements of the group, the problem of state estimation becomes equivalent to estimating the action of a black-box that performs an unknown unitary transformation drawn from the set  $\{U_g \mid g \in \mathbf{G}\}$ . From this point of view, it is also important to find the optimal input states that allow one to discriminate the action of the unitary operators  $\{U_g\}$  in the best possible way.

Whatever point of view we choose, we always need to estimate the value of a group parameter. In order to do this, the most general estimation strategy allowed by quantum mechanics is described by a positive operator valued measure (POVM)  $M$ , which associates to any estimated parameter  $\hat{g} \in \mathbf{G}$  a positive semidefinite operator  $M(\hat{g})$  on  $\mathcal{H}$ , satisfying the normalization condition

$$\int_{\mathbf{G}} dg M(g) = \mathbb{1}, \tag{2}$$

$dg$  being the normalized invariant Haar measure on the group ( $\int_{\mathbf{G}} dg = 1$   $d(hg) = d(gh) = dg \forall g, h \in \mathbf{G}$ ). The probability density of estimating  $\hat{g}$  when the true value

of the parameter is  $g$  is then given by the usual Born rule:

$$p(\hat{g} | g) = \text{Tr}[\rho_g M(\hat{g})], \quad (3)$$

where  $\rho_g = U_g \rho U_g^\dagger$ . Note that here we are considering  $\mathbf{G}$  as a continuous group only for fixing notation, nevertheless — here and all throughout the paper —  $\mathbf{G}$  can have a finite number of elements, say  $|\mathbf{G}|$ , and in this case we have simply to replace integrals with sums and  $dg$  with  $1/|\mathbf{G}|$ .

In order to find an optimal estimation strategy, we need firstly to fix our optimality criterion. The most common way to do this is to weigh the estimation errors with some cost function<sup>1,2</sup>  $f(\hat{g}, g)$ , which assesses the cost of estimating  $\hat{g}$  when the true value is  $g$ . Once the cost function is fixed, we can adopt two different settings for the optimization, the Bayesian and the minimax. In the Bayesian setting, one assumes a prior distribution of the true values (which is usually uniform) and then the optimal estimation is that which minimizes the average cost, where the average is performed with respect to the probability distributions of both the estimated and the true values. On the other hand, in the minimax setting, no prior distribution of the true values is assumed, and one minimizes the maximum (over all possible true values) of the average cost, where now the average is done just over the estimated values.

## 2.2. *Maximum likelihood approach*

Many different criteria can be used to define what is an optimal estimation, each of them corresponding to a different choice of the cost function in the optimization procedure. In general, the choice of a cost function depends on the particular problem at hand. For example, if we need to estimate a state, a natural cost is the opposite of the fidelity between the estimated state and the true one, while, if we are interested in the estimation of a parameter, a more appropriate cost function would be the variance of the estimated values.

In this paper, however, since we want to investigate general properties of covariant estimation, we seek a criterion that maintains a clear meaning in the largest number of different situations. The simplest approach that allows a general analytical solution is given by the *maximum likelihood criterion*,<sup>1,2</sup> which corresponds to the maximization of the probability (probability density in the continuous case) that the estimated value of the unknown parameter actually coincides with its true value. In this case, the cost function is a Dirac-delta  $f(\hat{g}, g) = -\delta(\hat{g}, g)$  (Kronecker-delta in the finite case).

For finite groups, maximum likelihood is in some sense the most natural criterion. In fact, if we are trying to decide among a finite set of alternatives which is the true one, of course we would like to do this with the maximum probability of success. On the other hand, in the continuous case, the maximum likelihood approach can still be interpreted as the maximization of the probability that the estimated value lies in a narrow neighborhood of the true one.

### 2.3. Basic results from group theory

Here, we will recall some useful tools from group theory that we will exploit throughout the paper.

Consider a finite-dimensional Hilbert space  $\mathcal{H}$  and a unitary (or, more generally, projective) representation  $R(\mathbf{G}) = \{U_g \in \mathcal{B}(\mathcal{H}) \mid g \in \mathbf{G}\}$  of a compact Lie group  $\mathbf{G}$ . The Hilbert space can be decomposed into orthogonal subspaces in the following way:

$$\mathcal{H} \equiv \bigoplus_{\mu \in \mathbf{S}} \mathcal{H}_\mu \otimes \mathbb{C}^{m_\mu}, \tag{4}$$

where the sum runs over the set  $\mathbf{S}$  of irreducible representations of  $\mathbf{G}$  that appear in the Clebsch–Gordan decomposition of  $R(\mathbf{G})$ . The action of the group is irreducible in each representation space  $\mathcal{H}_\mu$ , while it is trivial in the multiplicity space  $\mathbb{C}^{m_\mu}$ , namely,

$$U_g \equiv \bigoplus_{\mu \in \mathbf{S}} U_g^\mu \otimes \mathbb{1}_{m_\mu}, \tag{5}$$

where each representation  $\{U_g^\mu\}$  is irreducible, and  $\mathbb{1}_d$  denotes the identity operator in a  $d$ -dimensional Hilbert space. Moreover, any operator  $O \in \mathcal{B}(\mathcal{H})$  in the commutant of  $R(\mathbf{G})$  — i.e. such that  $[O, U_g] = 0 \forall g \in \mathbf{G}$  — has the form

$$O = \bigoplus_{\mu \in \mathbf{S}} \mathbb{1}_{d_\mu} \otimes O_\mu, \tag{6}$$

where  $d_\mu$  is the dimension of  $\mathcal{H}_\mu$ , and  $O_\mu$  is a  $m_\mu \times m_\mu$  complex matrix. In particular, the group average  $\langle A \rangle_{\mathbf{G}} \equiv \int dg U_g A U_g^\dagger$  of a given operator  $A$  with respect to the invariant Haar measure  $dg$  is in the commutant of  $R(\mathbf{G})$ , and has the form

$$\langle A \rangle_{\mathbf{G}} = \bigoplus_{\mu \in \mathbf{S}} \mathbb{1}_{d_\mu} \otimes \frac{1}{d_\mu} \text{Tr}_{\mathcal{H}_\mu}[A], \tag{7}$$

where  $\text{Tr}_{\mathcal{H}_\mu}[A]$  is a short notation for  $\text{Tr}_{\mathcal{H}_\mu}[P_\mu A P_\mu]$ ,  $P_\mu$  denoting the orthogonal projector over the Hilbert subspace  $\mathcal{H}_\mu \otimes \mathbb{C}^{m_\mu}$  in the decomposition (4) of  $\mathcal{H}$ . Here and throughout the paper, we assume the normalization of the Haar measure:  $\int_{\mathbf{G}} dg = 1$ .

**Remark 1. Entanglement between representation spaces and multiplicity spaces.** The choice of an orthonormal basis  $\mathbf{B}^\mu = \{|\phi_n^\mu\rangle \in \mathbb{C}^{m_\mu} \mid n = 1, \dots, m_\mu\}$  for a multiplicity space fixes a particular decomposition of the Hilbert space as a direct sum of irreducible subspaces:

$$\mathcal{H}_\mu \otimes \mathbb{C}^{m_\mu} = \bigoplus_{n=1}^{m_\mu} \mathcal{H}_n^\mu, \tag{8}$$

where  $\mathcal{H}_n^\mu \equiv \mathcal{H}_\mu \otimes |\phi_n^\mu\rangle$ . In this picture, it is clear that  $m_\mu$  is the number of different irreducible subspaces carrying the same representation  $\mu$ , each having dimension  $d_\mu$ .

Moreover, with respect to the decomposition (4), any pure state  $|\Psi\rangle \in \mathcal{H}$  can be written as

$$|\Psi\rangle = \bigoplus_{\mu \in \mathcal{S}} c_\mu |\Psi_\mu\rangle, \tag{9}$$

where  $|\Psi_\mu\rangle$  is a bipartite state in  $\mathcal{H}_\mu \otimes \mathbb{C}^{m_\mu}$  and  $\sum_{\mu \in \mathcal{S}} |c_\mu|^2 = 1$ . With respect to the direct sum decomposition (8), the Schmidt number of such a state is the minimum number of subspaces carrying the same representation  $\mu \in \mathcal{S}$  that are needed to decompose  $|\Psi\rangle$ .

**Remark 2. Maximum number of equivalent representations in the decomposition of a pure state.** The Schmidt number of any bipartite state  $|\Psi_\mu\rangle \in \mathcal{H}_\mu \otimes \mathbb{C}^{m_\mu}$  is always less than or equal to  $k_\mu = \min\{d_\mu, m_\mu\}$ . This means that any pure state can be decomposed using *no more* than  $k_\mu$  irreducible subspaces carrying the same representation  $\mu \in \mathcal{S}$ .

**2.4. Covariant measurements**

Since the set of states to be estimated is invariant under the action of the group, there is no loss of generality in assuming a covariant POVM, i.e. a POVM satisfying the property  $M(hg) = U_h M(g) U_h^\dagger$  for any  $g, h \in \mathbf{G}$ . In fact, it is well known that, for any possible POVM, there is always a covariant one with the same average cost,<sup>2</sup> this result holding both in the minimax approach and in the Bayesian approach with uniform prior distribution.

A covariant POVM has the form

$$M(g) = U_g \Xi U_g^\dagger, \tag{10}$$

where  $\Xi$  is a positive semidefinite operator. For covariant POVM's, exploiting the formula (7) for the group average, the normalization condition (2) can be translated into a simple set of conditions for the operator  $\Xi$ :

$$\text{Tr}_{\mathcal{H}_\mu}[\Xi] = d_\mu \mathbb{1}_{m_\mu}. \tag{11}$$

In this way, the optimization of a covariant POVM is reduced to the optimization of a positive operator satisfying the constraints (11).

**2.5. Optimal measurements**

Here, we derive for any given input state  $|\Psi\rangle \in \mathcal{H}$  the measurement that maximizes the probability (density) of estimating the true value of the unknown group parameter  $g \in \mathbf{G}$ . Note that, due to covariance, this probability has the same value for any group element:  $p(g|g) = \langle \Psi | \Xi | \Psi \rangle$ , according to Eqs. (3) and (10). In order to find the POVM, it is convenient to express the input state in the form (9), and

write each bipartite state  $|\Psi_\mu\rangle\rangle$  in the Schmidt form:

$$|\Psi_\mu\rangle\rangle = \sum_{m=1}^{r_\mu} \sqrt{\lambda_m^\mu} |\psi_m^\mu\rangle |\phi_m^\mu\rangle, \tag{12}$$

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

$$P_\Psi = \bigoplus_{\mu \in S} \sum_{m=1}^{r_\mu} \mathbb{1}_{d_\mu} \otimes |\phi_m^\mu\rangle\langle\phi_m^\mu|. \tag{13}$$

It projects onto the subspace  $\mathcal{H}_\Psi$  spanned by the orbit of the input state, this subspace being also the smallest invariant subspace containing the input state.

Clearly, the probability distribution of the outcomes of a covariant measurement  $M(\hat{g}) = U_{\hat{g}} \Xi U_{\hat{g}}^\dagger$  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 specify the operator  $P_\Psi \Xi P_\Psi$ . All covariant POVM's corresponding to the same operator will be equally optimal.

**Theorem 1. Optimal POVM.** *For a pure input state  $|\Psi\rangle$ , written according to Eqs. (9) and (12), the optimal covariant POVM in the maximum likelihood approach is given by*

$$P_\Psi \Xi P_\Psi = |\eta\rangle\langle\eta|, \tag{14}$$

where

$$|\eta\rangle = \bigoplus_{\mu \in S} \sqrt{d_\mu} e^{i \arg(c_\mu)} \sum_{m=1}^{r_\mu} |\psi_m^\mu\rangle |\phi_m^\mu\rangle. \tag{15}$$

The value of the likelihood for the optimal POVM is

$$p^{\text{Opt}}(g|g) = \left( \sum_{\mu \in S} |c_\mu| \sum_{m=1}^{r_\mu} \sqrt{\lambda_m^\mu d_\mu} \right)^2 \forall g. \tag{16}$$

**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 (12) 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  $\Xi$  implies

$$\langle \psi_m^\mu | \langle \phi_m^\mu | \Xi | \psi_m^\mu \rangle | \phi_m^\mu \rangle \leq \langle \phi_m^\mu | \text{Tr}_{\mathcal{H}_\mu} [\Xi] | \phi_m^\mu \rangle = d_\mu,$$

due to the normalization condition (11). By putting together these inequalities, we obtain the bound

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

holding for any possible POVM. It is immediate obvious that the covariant POVM given by Eqs. (14), (15) achieves the bound; hence, it is optimal.  $\square$

**Remark 3. Uniqueness of the optimal POVM.** Since the theorem specifies the optimal POVM only in the subspace spanned by the orbit of the input space, it follows that the optimal POVM is unique if and only if the orbit spans the whole Hilbert space. If it is not the case, one can arbitrarily complete the POVM given by (14) for the entire Hilbert space.

**Remark 4. Square-root measurements.** The optimal POVM in Eqs. (14) and (15) coincides with the so-called “square-root measurement”.<sup>23</sup> In fact, such a measurement has a POVM with  $|\eta\rangle = F^{-\frac{1}{2}}|\psi\rangle$ , with the “frame operator”  $F$  given by  $F = \int dg U_g |\psi\rangle\langle\psi| U_g^\dagger$ . Using Eq. (7), one finds  $F = \bigoplus_\mu |c_\mu|^2 \sum_m \frac{\lambda_m^\mu}{d_\mu} \mathbb{1}_\mu \otimes |\phi_m^\mu\rangle\langle\phi_m^\mu|$ , and one can easily check that  $|\eta\rangle = F^{-\frac{1}{2}}|\psi\rangle$ .

### 2.6. Optimal input states

While in the previous paragraph, we assumed the input state to be given, and we were mainly interested in the problem of state estimation, here we will focus our attention on the problem of estimating the action of a black box that performs an unknown unitary transformation drawn from a group. From this point of view, our aim is now to determine which are the states in the Hilbert space that allow us to maximize the probability of successfully discriminating the unknown unitaries  $\{U_g\}$ .

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 that this bound can always be achieved by using suitable input states. Then, the optimal input states will be those that maximize the dimension of the subspace spanned by the orbit.

**Lemma 1.** *Let  $d_\Psi = \dim \text{Span}\{U_g|\Psi\rangle \mid g \in \mathbf{G}\}$  be the dimension of the subspace spanned by the orbit of the input state. Then*

$$d_\Psi = \sum_{\mu \in \mathcal{S}} d_\mu r_\mu, \tag{17}$$



where  $r_\mu$  is the Schmidt number of the bipartite state  $|\Psi_\mu\rangle\rangle$  (we define  $r_\mu = 0$  if  $|c_\mu| = 0$  in the decomposition (9)).

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

$$\int_{\mathbf{G}} dg U_g |\Psi\rangle\langle\Psi| U_g^\dagger = \bigoplus_{\mu \in \mathcal{S}} |c_\mu|^2 \mathbb{1}_{d_\mu} \otimes \frac{\text{Tr}_{\mathcal{H}_\mu} [|\Psi_\mu\rangle\rangle\langle\langle\Psi_\mu|]}{d_\mu},$$

the right-hand side coming from Eq. (7). Using the Schmidt form (12) of each bipartite state  $|\Psi_\mu\rangle\rangle$ , it follows that the dimension of the support is  $d_\Psi = \sum_{\mu \in \mathcal{S}} d_\mu r_\mu$ .  $\square$

**Theorem 2. Relation between likelihood and dimension.** *For any pure input state  $|\Psi\rangle \in \mathcal{H}$ , the following bound holds:*

$$p(g|g) \leq d_\Psi. \tag{18}$$

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

$$|\Psi\rangle = \frac{1}{\sqrt{d_\Psi}} \bigoplus_{\mu \in \mathcal{S}} \sqrt{d_\mu r_\mu} e^{i\theta_\mu} |\Psi_\mu\rangle\rangle, \tag{19}$$

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  $r_\mu$  and equal Schmidt coefficients ( $\lambda_m^\mu = 1/r_\mu$  for any  $m = 1, \dots, r_\mu$ ).

**Proof.** Exploiting Eq. (16), we have

$$p(g|g) \leq p^{\text{Opt}}(g|g) \tag{20}$$

$$= \left( \sum_{\mu} |c_\mu| \sum_{m=1}^{r_\mu} \sqrt{\lambda_m^\mu d_\mu} \right)^2 \tag{21}$$

$$\leq \left( \sum_{\mu} |c_\mu| \sqrt{r_\mu d_\mu} \right)^2 \tag{22}$$

$$\leq \sum_{\mu} r_\mu d_\mu = d_\Psi, \tag{23}$$

the inequalities (22) and (23) coming from the 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 (20) holds if we use the optimal POVM of Theorem 1. On the other hand, the Schwartz inequality (22) becomes an equality if and only if  $\lambda_m^\mu = 1/r_\mu$  for any  $m = 1, \dots, r_\mu$ . Finally, the last Schwartz inequality (23) becomes an equality if and only if  $|c_\mu| = \sqrt{r_\mu d_\mu / d_\Psi}$ . The requirements  $|c_\mu| = \sqrt{r_\mu d_\mu / d_\Psi}$  and  $\lambda_m^\mu = 1/r_\mu$  are satisfied only by states of the form (19).  $\square$

We can now answer to the question which are the best input states for discriminating a group of unitaries.

**Theorem 3. Optimal input states.** For any state  $\rho$  on  $\mathcal{H}$  and for any POVM, the likelihood is bounded from above by the quantity

$$L = \sum_{\mu \in S} d_\mu k_\mu, \tag{24}$$

where  $k_\mu \equiv \min\{d_\mu, m_\mu\}$ . The bound is achieved by pure states of the form

$$|\Psi\rangle = \frac{1}{\sqrt{L}} \bigoplus_{\mu \in S} \sqrt{d_\mu k_\mu} e^{i\theta_\mu} |E_\mu\rangle\rangle, \tag{25}$$

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.

**Proof.** Since the likelihood  $\mathcal{L}[\rho] = \text{Tr}[\rho \Xi]$  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. (18), the maximum likelihood is given by the maximum of  $d_\Psi$  over all pure states. Since the Schmidt number  $r_\mu$  cannot exceed  $k_\mu = \max\{d_\mu, m_\mu\}$ , we obtain that the maximum value is

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

According to Theorem 2, such a maximum is achieved by pure states of the form (25). □

The results of Theorems 1–3 have some important consequences.

**Consequence 1 (Each irreducible subspace contributes to the likelihood with its dimension).** According to Eq. (25), 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 related to the dimension  $d_\mu$  by Eqs. (17), (18), and (24). In other words, the maximum likelihood approach allows one 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. From this point of view, the interpretation of the well-known example<sup>24</sup> about the quantum information of two parallel versus anti-parallel spin 1/2 particles is clear: for parallel spins the input state lies completely in the triplet (symmetric) subspace, while for anti-parallel it has a nonzero component also on the singlet. Evaluating the likelihood with Eq. (16), we have indeed  $p(g|g)^{\text{Opt}} = (1 + \sqrt{3})^2/2 \approx 3.73$  for anti-parallel spins, instead of  $p(g|g)^{\text{Opt}} = 3$  for parallel ones. Notice, however, that the latter is not the optimal input state in the maximum likelihood approach, which instead has coefficients  $c_\mu = \sqrt{r_\mu d_\mu / d_\Psi}$ , with  $\mu = 0$  denoting the singlet and  $\mu = 1$  the triplet. The largest  $r_\mu$  are given by  $r_\mu = 1$ , whence  $c_\mu = \sqrt{d_\mu / d_\Psi}$ , namely  $c_0 = 1/2$  and  $c_1 = \sqrt{3}/2$ , giving likelihood  $p(g|g)^{\text{Opt}} = 4$ . It is possible to show<sup>9</sup> that this optimal input state

can be chosen as a factorized state, such a state being the tensor product of two *mutually unbiased* spin states.<sup>9</sup>

**Consequence 2 (Key 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 (24) 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  $\mu$ , for any  $\mu$  in the Clebsch–Gordan decomposition.

**Consequence 3 (Maximization of the Holevo  $\chi$ -quantity).** The optimal states in the maximum likelihood approach are those which maximize the Holevo  $\chi$ -quantity,<sup>10</sup> which in the group covariant case is defined as

$$\chi_{\mathbf{G}}(\rho) = S\left(\int dg U_g \rho U_g^\dagger\right) - \int dg S(U_g \rho U_g^\dagger), \tag{26}$$

$S(\rho) = -\text{Tr}[\rho \log(\rho)]$  being the von Neumann entropy. In fact, for pure input states  $\rho = |\Psi\rangle\langle\Psi|$ , the  $\chi$ -quantity is the entropy of the average state:  $\chi_{\mathbf{G}}(\rho) = S(\langle\rho\rangle_{\mathbf{G}})$ . Using Eq. (7), we have

$$\langle\rho\rangle_{\mathbf{G}} = \bigoplus_{\mu \in \mathcal{S}} |c_\mu|^2 \frac{\mathbb{1}_{d_\mu}}{d_\mu} \otimes \text{Tr}_{\mathcal{H}_\mu}[|\Psi_\mu\rangle\langle\Psi_\mu|].$$

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

$$\chi_{\mathbf{G}}(\rho) \leq \log d_\Psi, \tag{27}$$

and that the bound is attained by states of the form (19). Finally, the maximum over all pure states is

$$\chi_{\mathbf{G}}(\rho) = \log L, \tag{28}$$

achieved by states of the form (25). In this way, the likelihood is directly related to the  $\chi$ -quantity, providing an upper bound to the amount of classical information that can be extracted from the orbit of the input state.

### 3. Internal Versus External Entanglement

Up to now, we have looked for the optimal input states in the Hilbert space of the system undergoing the unknown group transformation. From this point of view, the entanglement between representation and multiplicity spaces was just a kind of *internal entanglement* between two *virtual subsystems*<sup>18</sup> with Hilbert spaces given by the representation and the multiplicity spaces, respectively.

Suppose now that we can exploit an additional entangled resource, i.e. we can entangle the system that undergoes the unknown group transformation with an additional external system, which acts as a reference. In this case, we have to consider the tensor product Hilbert space  $\mathcal{H} \otimes \mathcal{H}_{\mathcal{R}}$ , where the group acts via the

representation  $\{U'_g = U_g \otimes \mathbb{1}_{\mathcal{R}} \mid g \in \mathbf{G}\}$ . From the point of view of the group structure, the only effect of the reference system is simply to increase the multiplicity of the irreducible representations. In fact, the Clebsch–Gordan decompositions of  $\{U_g\}$  and  $\{U'_g\}$  contain exactly the same irreducible representations, while the new multiplicities are  $m'_\mu = m_\mu \cdot d_{\mathcal{R}}$ , where  $d_{\mathcal{R}} = \dim \mathcal{H}_{\mathcal{R}}$ . Therefore, from the sole consideration of the decomposition of the Hilbert space, we obtain the following:

**Theorem 4. Use of external entanglement.** *Once  $m'_\mu = m_\mu \cdot d_{\mathcal{R}} \geq d_\mu$  for any irreducible representation  $\mu \in \mathbf{S}$ , any further increase of the dimension of the reference system is useless.*

*In particular, if  $d_\mu \geq m_\mu \forall \mu \in \mathbf{S}$ , there is no need for entanglement with an external system.*

**Proof.** In the decomposition (9) of a pure state, the rank of any bipartite state cannot exceed  $k'_\mu = \min\{d_\mu, m'_\mu\}$ . Therefore, once  $m'_\mu \geq d_\mu$  for any  $\mu$ , the orbit of any pure state can always be embedded in a subspace where all the multiplicities are equal to the dimensions. □

In other words, once the saturation  $m'_\mu \geq d_\mu$  is reached, there is no need for increasing the dimension for the reference system.

**Corollary 1. Dimension of the reference system.** *The maximum dimension of an external system that is useful for estimation is*

$$\bar{d}_{\mathcal{R}} = \max \left\{ \left\lceil \frac{d_\mu}{m_\mu} \right\rceil \mid \mu \in \mathbf{S} \right\}, \tag{29}$$

where the “ceiling”  $\lceil x \rceil$  denotes the minimum integer greater than  $x$ .

This mechanism of *saturation of the multiplicities* can be simply quantified in terms of the likelihood. In fact, the improvement coming from the external reference system can be evaluated using Theorem 3, yielding the value of the likelihood for the optimal input state:  $L' = \sum_{\mu \in \mathbf{S}} d_\mu k'_\mu$ , where  $k'_\mu = \min\{d_\mu, m'_\mu\}$ . The upper bound that can be achieved with the use of a reference system as in Corollary 1 is then

$$L_{\max} = \sum_{\mu \in \mathbf{S}} d_\mu^2. \tag{30}$$

## 4. Generalization to Infinite Dimension and Non-Compact Groups

### 4.1. Compact groups in infinite dimension

The main problem with infinite dimension comes from the fact that, in some cases, the optimal states of Sec. 2.6 are not normalizable. From a physical point of view, this means that one has to approximate them with normalized states in some reasonable way, fixing additional constraints, such as the energy constraint. Clearly, the best approximation depends on the particular problem at hand.

In a similar way, the POVM elements  $P(g) = U_g \Xi U_g^\dagger$  in general are not operators. For example, the well-known optimal POVM for the estimation of the phase of the radiation field is given by

$$P(\phi) = |e(\phi)\rangle\langle e(\phi)|, \tag{31}$$

where  $|e(\phi)\rangle = \sum_{n=0}^\infty e^{in\phi}|n\rangle$  are the so-called Susskind–Glogower vectors.<sup>25</sup> Since such vectors are not normalizable, the POVM elements  $P(\phi)$  are not operators acting in the Hilbert space  $\mathcal{H}$ . For this reason, in infinite dimensions one should substitute the positive operator  $P(g)$  with a positive form  $\pi_g$ , defined by  $\pi_g(|\phi\rangle, |\psi\rangle) = \langle\phi|P(g)|\psi\rangle$ .

However, except for this technicality, all results of Sec. 2.5 concerning optimal POVM’s are essentially valid in infinite dimensions.

#### 4.2. Non-compact groups

The generalization of our method to the case of non-compact groups is more involved than for compact groups in infinite dimension. Nevertheless, such a generalization is crucial for many physically meaningful cases, e.g. the estimation of displacement or of squeezing parameters in quantum optics.

In the following, we will consider the case of *unimodular groups*, i.e. groups for which the left-invariant measure  $d_Lg$  ( $d_Lhg = d_Lg \ \forall g, h \in \mathbf{G}$ ) and the right-invariant one  $d_Rg$  ( $d_Rgh = d_Rg \ \forall g, h \in \mathbf{G}$ ) coincide. We will then define the invariant Haar measure as  $dg = d_Lg = d_Rg$ . Notice that, from the Bayesian point of view, it is no longer possible to assume that the group parameters are distributed according to such a measure, since the uniform measure over a non-compact group is non-normalizable.

In general, for non-compact groups the irreducible representations contained in the Clebsch–Gordan decomposition may form a continuous set. To deal with such a situation, one should replace in the Secs. 2.5 and 2.6 direct sums with direct integrals. For example, the decomposition of the Hilbert space (4) would be rewritten as<sup>a</sup>

$$\mathcal{H} = \int_{\mathbf{S}}^\oplus m(d\mu)\mathcal{H}_\mu \otimes \mathcal{M}_\mu, \tag{32}$$

$m(d\mu)$  being a measure over the set of irreducible representations, and  $\mathcal{M}_\mu$  denoting the multiplicity space. In the following, we will not carry on this rather technical generalization, leaving the case of a direct integral of irreducible representations only to a specific example (see next paragraph). We will instead consider the simplest case of group representations that can be decomposed in a *discrete series* of

---

<sup>a</sup>This decomposition holds for the whole class of *locally compact type I groups*, which covers all the possible cases that are relevant for physical applications. Exceptions to this decomposition fall in the somewhat exotic class of *wild groups*.

irreducible components. In other words, we will assume that it is still possible to write the Clebsch–Gordan decomposition

$$U_g = \bigoplus_{\mu \in S} U_g^\mu \otimes \mathbb{1}_{\mathcal{M}_\mu}, \tag{33}$$

where the set  $S$  is discrete.

Finally, we require any irreducible representation  $\{U_g^\mu\}$  in the Clebsch–Gordan series to be *square summable*, which, in the in case of unimodular groups, is equivalent to the property

$$\int dg |\langle \psi_\mu | U_g | \phi_\mu \rangle|^2 < \infty \quad \forall |\psi_\mu\rangle, |\phi_\mu\rangle \in \mathcal{H}_\mu. \tag{34}$$

Under these hypotheses, the results of Theorems 1–3 can be immediately extended to non-compact groups. In fact, in this case we can exploit a simple generalization of the formula (7) for the group average, which allows us the use all the results of Secs. 2.5 and 2.6 by just substituting the dimensions  $d_\mu$  of the irreducible subspaces with their *formal dimensions*.

**Proposition 1.** *Let be  $\{U_g\}$  a discrete series of square-summable representations of a unimodular group. Then, the group average  $\langle A \rangle_{\mathbf{G}}$  of a given operator  $A$  is given by*

$$\langle A \rangle_{\mathbf{G}} = \bigoplus_{\mu \in S} \mathbb{1}_{\mathcal{H}_\mu} \otimes \frac{\text{Tr}_{\mathcal{H}_\mu}[A]}{d_\mu}, \tag{35}$$

where the formal dimension  $d_\mu$  is defined as

$$d_\mu = \left( \int dg |\langle \psi_\mu | U_g^\mu | \phi_\mu \rangle|^2 \right)^{-1}, \tag{36}$$

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

**Proof.** Since the group average  $\langle A \rangle_{\mathbf{G}}$  of an operator is in the commutant of the representation (33) it has the form  $\langle A \rangle_{\mathbf{G}} = \bigoplus_{\mu} \mathbb{1}_{\mathcal{H}_\mu} \otimes A_\mu$ , for some suitable operators  $A_\mu$  acting in the multiplicity space. Taking the expectation value with respect to a normalized vector  $|\psi_\mu\rangle \in \mathcal{H}_\mu$ , we obtain  $A_\mu = \langle \psi_\mu | \langle A \rangle_{\mathbf{G}} | \psi_\mu \rangle = \text{Tr}_{\mathcal{H}_\mu}[A \langle B_\mu \rangle_{\mathbf{G}}]$ , where  $B_\mu = |\psi_\mu\rangle \langle \psi_\mu| \otimes \mathbb{1}_{\mathcal{M}_\mu}$ . Now, since the group average  $\langle B \rangle_{\mathbf{G}}$  is in the commutant of  $\{U_g\}$ , and since  $|\psi_\mu\rangle \in \mathcal{H}_\mu$ , we have  $\langle B \rangle_{\mathbf{G}} = 1/d_\mu \mathbb{1}_{\mathcal{H}_\mu} \otimes \mathbb{1}_{\mathcal{M}_\mu}$  for some constant  $d_\mu$ . The constant  $d_\mu$  is simply evaluated by taking the expectation value of  $\langle B \rangle_{\mathbf{G}}$  with respect to a normalized vector  $|\phi_\mu\rangle | \alpha_\mu \rangle \in \mathcal{H}_\mu \otimes \mathcal{M}_\mu$ .  $\square$

**Remark.** The *formal dimension* of Eq. (36) 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.

**4.3. An application: Two identical and two conjugated coherent states**

Here, we give two examples of 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 group of 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  $\frac{d^2\alpha}{\pi}$  (here, we put the constant  $\pi$  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. 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 on quantum information carried by parallel and anti-parallel spins.<sup>24</sup> This analogy, already noticed in the study of the optimal “phase conjugation map” by Cerf and Iblisdir,<sup>26</sup> will be analyzed here in detail from the general point of view of group parameter estimation.

4.3.1. *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

$$\mathcal{S} = \{|\alpha\rangle|\alpha\rangle \in \mathcal{H}^{\otimes 2} \mid \alpha \in \mathbb{C}\}, \tag{37}$$

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

$$D(\alpha)^{\otimes 2} = V^\dagger D(\sqrt{2}\alpha) \otimes \mathbb{1}V, \tag{38}$$

where  $V = \exp[-\frac{\pi}{4}(a_1^\dagger a_2 - a_1 a_2^\dagger)]$  ( $a_1$  and  $a_2$  denoting annihilation operators for the first and the second oscillator, respectively). 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

$$d = \left( \int_{\mathbb{C}} \frac{d^2\alpha}{\pi} | \langle 0|D(\sqrt{2}\alpha)|0\rangle|^2 \right)^{-1} = 2, \tag{39}$$

given by Eq. (36). Moreover, according to Eq. (8), a possible decomposition of the tensor product Hilbert space into irreducible subspaces is given by any set of the form

$$\mathcal{H}_n = V^\dagger \mathcal{H} \otimes |\phi_n\rangle, \tag{40}$$

where  $\{|\phi_n\rangle | n \in 0, 1, \dots\}$  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 1, we have that for the state  $|\Psi\rangle$  the optimal-likelihood covariant POVM must have  $\Xi$  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. (12)]. 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  $\alpha$ . For example, we can take the POVM

$$M(\alpha) = 2D(\alpha)^{\otimes 2} (V^\dagger|\mathbb{1}\rangle\rangle\langle\langle\mathbb{1}|V) D(\alpha)^{\dagger\otimes 2}, \tag{41}$$

where the unitary  $V$  is defined as above, and  $|\mathbb{1}\rangle\rangle$  is the vector  $|\mathbb{1}\rangle\rangle = \sum_{n=0}^\infty |n\rangle|n\rangle$ . 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

$$\tilde{M}(\alpha) = 2D(\alpha)^{\otimes 2} (V^\dagger|0\rangle\langle 0| \otimes \mathbb{1}V) D(\alpha)^{\dagger\otimes 2}. \tag{42}$$

In a quantum optical setup, this POVM corresponds to performing firstly a heterodyne measurement on each oscillator, thus obtaining two different estimates  $\alpha_1$  and  $\alpha_2$  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  $\alpha$  is given in both cases by the Gaussian

$$p(\hat{\alpha} | \alpha) = 2e^{-2|\hat{\alpha} - \alpha|^2} \tag{43}$$

(normalized with respect to the invariant measure  $\frac{d^2\alpha}{\pi}$ ). The value of the likelihood is  $p(\alpha | \alpha) = 2$ , according to our general formula (16).

**Remark. Improving the likelihood with squeezing.** The maximum likelihood can be improved using the doubly-squeezed state

$$|\Psi_x\rangle = V^\dagger \sqrt{1-x^2} \sum_{n=0}^\infty x^n |n\rangle|n\rangle, \tag{44}$$



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

$$p(\alpha | \alpha) = 2\frac{1+x}{1-x}. \tag{45}$$

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

### 4.3.2. Conjugated coherent states

Now, the family of signal states is

$$\mathcal{S} = \{|\alpha\rangle|\alpha^*\} | \alpha \in \mathbb{C}\}, \tag{46}$$

where complex conjugation is defined with respect to the basis  $\{|n\rangle | n = 0, 1, \dots\}$ . These states are generated from the input state  $|\Psi\rangle = |0\rangle|0\rangle$  by the action of the representation  $\{D(\alpha) \otimes D(\alpha^*) | \alpha \in \mathbb{C}\}$ .

Unfortunately, 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

$$D(\alpha) \otimes D(\alpha^*)|D(\beta)\rangle\rangle = e^{\alpha\beta^* - \alpha^*\beta}|D(\beta)\rangle\rangle. \tag{47}$$

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

$$\mathcal{H} \otimes \mathcal{H} = \int_{\mathbb{C}}^{\oplus} \frac{d^2\beta}{\pi} \mathcal{H}_\beta. \tag{48}$$

In the same way as in (6), an operator  $O \in \mathcal{B}(\mathcal{H} \otimes \mathcal{H})$  in the commutant of the representation can be written as

$$O = \int_{\mathbb{C}}^{\oplus} \frac{d^2\beta}{\pi} \mathbb{1}_\beta o(\beta), \tag{49}$$

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

In this particular example, it is easy to extend the results of Secs. 2.5 and 2.6 to the case of a direct integral of irreducible representations. In fact, using functional calculus, we can generalize the formula (7) for the group average.

**Proposition 2.** *The average  $\langle A \rangle_{\mathbb{C}}$  of an operator over the representation  $\{D(\alpha) \otimes D(\alpha^*) \mid \alpha \in \mathbb{C}\}$  is*

$$\langle A \rangle_{\mathbb{C}} = \int_{\mathbb{C}}^{\oplus} \frac{d^2\beta}{\pi} \mathbb{1}_{\beta} \text{Tr}_{\mathcal{H}_{\beta}}[A], \tag{50}$$

where  $\text{Tr}_{\mathcal{H}_{\beta}}[A] = \langle\langle D(\beta)|A|D(\beta)\rangle\rangle$ .

This expression for the group average is equivalent to that of Eq. (7) modulo the obvious substitutions:

$$\begin{cases} \bigoplus_{\mu \in S} \rightarrow \int_{\mathbb{C}}^{\oplus} \frac{d^2\beta}{\pi}, \\ d_{\mu} \rightarrow d_{\beta} = 1 \quad \forall \beta \in \mathbb{C}. \end{cases} \tag{51}$$

The optimal POVM is obtained by Theorem 1 by making these substitutions. We just need to decompose the input state on the irreducible subspaces, i.e.

$$|0\rangle|0\rangle = \int_{\mathbb{C}}^{\oplus} \frac{d^2\beta}{\pi} e^{-|\beta|^2/2} |D(\beta)\rangle, \tag{52}$$

and then take the optimal POVM given by the operator  $\Xi = |\eta\rangle\langle\eta|$  in Eqs. (14) and (15), which in the present case becomes

$$|\eta\rangle = \int_{\mathbb{C}}^{\oplus} \frac{d^2\beta}{\pi} |D(\beta)\rangle. \tag{53}$$

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

$$M(\alpha) = D(\alpha) \otimes D(\alpha^*) |\eta\rangle\langle\eta| D(\alpha)^{\dagger} \otimes D(\alpha^*)^{\dagger}, \tag{54}$$

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

$$p(\hat{\alpha} \mid \alpha) = 4e^{-4|\hat{\alpha}-\alpha|^2} \tag{55}$$

(normalized with respect to the invariant measure  $\frac{d^2\alpha}{\pi}$ ).

Notice that the value of the likelihood  $p(\alpha \mid \alpha) = 4$  could also be calculated directly using the formula (16), which now reads

$$p(\alpha \mid \alpha) = \left( \int_{\mathbb{C}} \frac{d^2\beta}{\pi} e^{-|\beta|^2/2} \right)^2 = 4. \tag{56}$$

Comparing the optimal distribution (55) for two conjugated coherent states with the corresponding one for two identical coherent states (43), we can observe that the variance has been reduced by one half, while the likelihood has become twice as high. It is interesting to note the remarkable analogy between this example in continuous variables and the example by Gisin and Popescu<sup>24</sup> on the quantum 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 set-up 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  $\alpha$  and  $\alpha^*$  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-1/2 particles the “spin-flip” operation is unitarily equivalent to the complex conjugation, whence we can also regard 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 also provide in both cases 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 (16), 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.

## Acknowledgments

This work has been supported by the FET European Networks on Quantum Information and Communication Contract IST-2000-29681:ATESIT, by MIUR 2003-Cofinanziamento, and by INFN PRA-2002-CLON.

## References

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. S. Massar and S. Popescu, *Phys. Rev. Lett.* **74** (1995) 1259.
4. M. Hayashi, *J. Phys. A* **31** (1998) 4633.
5. R. D. Gill and S. Massar, *Phys. Rev. A* **61** (2002) 042312.
6. D. Bruss, A. K. Ekert and C. Macchiavello, *Phys. Rev. Lett.* **81** (1998) 2598.
7. V. Bužek, R. Derka and S. Massar, *Phys. Rev. Lett.* **82** (1999) 2207.
8. G. M. D’Ariano, C. Macchiavello and P. Perinotti, *Phys. Rev. A* **72** (2005) 042327.
9. G. Chiribella, G. M. D’Ariano, P. Perinotti and M. F. Sacchi, *Phys. Rev. A* **70** (2004) 062105.
10. A. S. Holevo, in *Proc. 2nd Japan–USSR Symp. Probability Theory*, eds. G. Maruyama and J. V. Prokhorov, Lecture Notes in Mathematics, Vol. 330 (Springer-Verlag, Berlin, 1973), p. 104.
11. G. Chiribella, G. M. D’Ariano, P. Perinotti and M. F. Sacchi, *Phys. Rev. Lett.* **93** (2004) 180503.
12. A. Peres and P. F. Scudo, *Phys. Rev. Lett.* **87** (2001) 167901.

13. E. Bagan, M. Baig and R. Muñoz-Tapia, *Phys. Rev. Lett.* **87** (2001) 257903.
14. A. Peres and P. F. Scudo, *J. Mod. Opt.* **49** (2002) 1235.
15. E. Bagan, M. Baig and R. Muñoz-Tapia, *Phys. Rev. A* **69** (2004) 050303.
16. E. Bagan, M. Baig and R. Muñoz-Tapia, *Phys. Rev. A* **70** (2004) 030301(R).
17. E. Knill, R. Laflamme and L. Viola, *Phys. Rev. Lett.* **84** (2000) 2525.
18. P. Zanardi, *Phys. Rev. Lett.* **87** (2001) 077901.
19. S. D. Bartlett, T. Rudolph and R. W. Spekkens, *Phys. Rev. Lett.* **91** (2003) 027901.
20. S. D. Bartlett, T. Rudolph and R. W. Spekkens, *Phys. Rev. A* **70** (2004) 032307.
21. M. Hayashi, quant-ph/0407053.
22. G. Chiribella, G. M. D'Ariano and M. F. Sacchi, *Phys. Rev. A* **72** (2005) 042338.
23. P. Hausladen and W. K. Wootters, *J. Mod. Opt.* **41** (1994) 2385.
24. N. Gisin and S. Popescu, *Phys. Rev. Lett.* **83** (1999) 432.
25. L. Susskind and J. Glogower, *Physics* **1** (1964) 49.
26. N. J. Cerf and S. Iblisdir, *Phys. Rev. A* **64** (2001) 032307.