Extremal covariant measurements

Giulio Chiribella[a] and Giacomo Mauro D'Ariano[b]
Istituto Nazionale di Fisica della Materia, Unità di Pavia, Dipartimento di Fisica "A. Volta," via Bassi 6, I-27100 Pavia, Italy and Department of Electrical and Computer Engineering, Northwestern University, Evanston, Illinois 60208

(Received 20 March 2006; accepted 15 August 2006; published online 28 September 2006)

We characterize the extremal points of the convex set of quantum measurements that are covariant under a finite-dimensional projective representation of a compact group, with action of the group on the measurement probability space which is generally nontransitive. In this case the POVM density is made of multiple orbits of positive operators, and, in the case of extremal measurements, we provide a bound for the number of orbits and for the rank of POVM elements. Two relevant applications are considered, concerning state discrimination with mutually unbiased bases and the maximization of the mutual information. © *2006 American Institute of Physics.* [DOI: 10.1063/1.2349481]

I. INTRODUCTION

A fundamental issue in the theory of quantum information[1] is the investigation of the ultimate precision limits for extracting classical information from a quantum system. Indeed, when the information is encoded on quantum states, its read-out suffers the intrinsically quantum limitation of discriminating among nonorthogonal states. One then needs to optimize the discrimination with respect to a given optimality criterion, which is dictated by the particular task for which the measurement is designed, or by the particular way the information is encoded over states. The good news is that, although the position of the problem has a limited generality due to the specific form of the optimality criterion, nevertheless for a large class of criteria the optimization method is given by a standard procedure. In such approach all possible measurements form a convex set (the convex combination of two measurements corresponding to the random choice between their apparatuses), and the optimization consists in maximizing a convex functional, e.g., the mutual information,[2,3] or to minimizing a concave functional, e.g., a Bayes cost,[4,5] over the convex set of measurements. Since the global maximum of a convex functional (or the minimum of a concave functional) is achieved over extremal points, the optimization can be restricted to the extremal elements of the set only.

In most situations of interest, the set of signal states on which the information is encoded is invariant under the unitary action of some group of physical transformations. The symmetry of the set of signal states is then reflected in a symmetry of the optimal measurements, which without loss of generality can be assumed to be *covariant*[5] with respect to the same group of transformations.

The problem of charactering extremal covariant measurements has been addressed in Refs. 6 and 7, however restricting the analysis to the case of group-action that is *transitive* on the probability space of measurement outcomes, namely any two points in the probability space are connected by some group element. The present paper completes the investigation by generalizing all results to the case of nontransitive group actions. Indeed the discrimination of states belonging to disjoint group orbits occurs in actual applications, and this situation has received little attention in the literature. Moreover, when classical information is encoded on quantum states it can be

[a] Electronic mail: chiribella@unipv.it
[b] Electronic mail: dariano@unipv.it

convenient to decode it with a measurement having outcomes that are not in one-to-one correspondence with the encoding states. This typically happens when the optimality criterion is nonlinear in the probabilities of measurement outcomes, as in the case of the mutual information.[8] In the presence of group symmetry, as recently noted by Decker,[9] even if the encoding states form a single group orbit, the maximization of the mutual information often selects covariant measurements with probability space that splits into disjoint orbits. It is then interesting to quantify the number of orbits needed for the maximization of the mutual information, or at least to give an upper bound for it. Indeed, as we will see in the present paper, the characterization of extremal covariant measurements also provides as a by-product an alternative and simpler derivation of the bound given in Ref. 9.

II. STATEMENT OF THE PROBLEM

In the general framework of quantum mechanics the state of a system is represented by a density operator ρ on a given Hilbert space $\mathcal{H}$, whereas the statistics of a measurement is described by a positive operator valued measure (POVM), which associates a positive semidefinite operator $P(B) \in \mathcal{B}(\mathcal{H})$ to any subset $B \in \sigma(\mathfrak{X})$ of the σ-algebra of events in the probability space $\mathfrak{X}$. The defining properties for a POVM are

$$0 \leqslant P(B) \leqslant \mathbb{1}, \quad \forall B \in \sigma(\mathfrak{X}), \tag{1}$$

$$P(\cup_{k=1}^{\infty} B_k) = \sum_{k=1}^{\infty} P(B_k), \quad \forall \{B_k\} \text{ disjoint} \tag{2}$$

$$P(\mathfrak{X}) = \mathbb{1}. \tag{3}$$

The probability of the event $B \in \sigma(\mathfrak{X})$ is then given by the Born rule

$$p(B) = \mathrm{Tr}[\rho P(B)]. \tag{4}$$

In this paper we will consider the case where the probability space $\mathfrak{X}$ supports the action of a compact group $\mathbf{G}$, namely any group element $g \in \mathbf{G}$ acts as a measurable automorphism of the probability space $\mathfrak{X}$, which maps $x \in \mathfrak{X}$ to $gx \in \mathfrak{X}$. If any two points $x_1, x_2 \in \mathfrak{X}$ are connected by some group element, i.e., $x_2 = gx_1$ for some $g \in \mathbf{G}$, the group action is called *transitive*. In this case, which is the most studied in the literature,[4,5] the whole probability space is the group orbit of an arbitrary point $x_0 \in \mathfrak{X}$, namely $\mathfrak{X} = \{gx_0 | g \in \mathbf{G}\}$. In this paper we will study the more general case where the group action is not transitive, and, accordingly, the probability space is not a single group orbit, but the union of a set of disjoint orbits, each one being labeled by an index $i \in \mathcal{I}$ for some set $\mathcal{I}$. For simplicity, we will assume the index set $\mathcal{I}$ to be finite.

The simplest case of the nontransitive group action then arises when the probability space is the Cartesian product of the index set $\mathcal{I}$ with the compact group $\mathbf{G}$, i.e., $\mathfrak{X} = \mathcal{I} \times \mathbf{G}$. In this case, the action of a group element $h \in \mathbf{G}$ on a point $x = (i, g) \in \mathcal{I} \times \mathbf{G}$ is given by $hx = (i, hg)$. Measurements with outcomes in $\mathcal{I} \times \mathbf{G}$ naturally arise in the discrimination of a set of signal states which is the union of a certain number of disjoint group orbits, each orbit $\mathcal{O}_i$ being generated by the action of the group on a given initial state ρ_i, namely $\mathcal{O}_i = \{U_g \rho_i U_g^\dagger | g \in \mathbf{G}\}$ for some unitary representation $\mathsf{R}(\mathbf{G}) = \{U_g | g \in \mathbf{G}\}$. Precisely, if the *stability group* $\mathbf{G}_i = \{h \in \mathbf{G} | U_h \rho_i U_h^\dagger = \rho_i\}$ associated to any state ρ_i consists only of the identity element e, then there is a one-to-one correspondence between signal states and points of the probability space $\mathfrak{X} = \mathcal{I} \times \mathbf{G}$. In Sec. IV we will study in detail the case of POVMs with probability space $\mathfrak{X} = \mathcal{I} \times \mathbf{G}$.

If the stability groups associated to the initial states $\{\rho_i | i \in \mathcal{I}\}$ are nontrivial, namely $\mathbf{G}_i \neq \{e\}$ for some $i \in \mathcal{I}$, in order to have a one-to-one correspondence between signal states and measurement outcomes, one must consider the probability space $\mathfrak{X} = \cup_{i \in \mathcal{I}} \mathbf{G}/\mathbf{G}_i$, where $\mathbf{G}/\mathbf{G}_i$ denotes the quotient of $\mathbf{G}$ with respect to the equivalence relation "$g \sim g'$ if $g' = g \cdot h$ for some $h \in \mathbf{G}_i$." This more general case will be treated in Sec. V.

Definition 1 (*covariant POVMs*): *Let $\mathfrak{X}$ be a probability space supporting the group action $g: x \in \mathfrak{X} \mapsto gx \in \mathfrak{X}$. A POVM is covariant*[5] *if it satisfies the property*

$$P(B) = U_g^\dagger P(gB) U_g, \quad \forall\, B \in \sigma(\mathfrak{X}), \ \forall\, g \in \mathbf{G}, \tag{5}$$

where $gB \doteq \{gx | x \in B\}$.

In the case $\mathfrak{X} = \mathcal{I} \times \mathbf{G}$, it is simple to prove[10] that any covariant POVM admits an operator density $M(i,g)$ with respect to the (normalized) Haar measure dg 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[10]

$$M(i,g) = U_g A_i U_g^\dagger, \tag{6}$$

where $A_i \in \mathcal{B}(\mathcal{H})$ are Hermitian operators satisfying the constraints

$$A_i \geqslant 0, \quad \forall\, i \in \mathcal{I}, \tag{7}$$

$$\sum_{i \in \mathcal{I}} \int_{\mathbf{G}} \mathrm{d}g U_g A_i U_g^\dagger = \mathbb{1}. \tag{8}$$

Here and throughout the paper we adopt for the Haar measure the normalization

$$\int_{\mathbf{G}} \mathrm{d}g = 1. \tag{9}$$

According to the above discussion, any covariant POVM with probability space $\mathfrak{X} = \mathcal{I} \times \mathbf{G}$ is completely specified by a set of operators $\{A_i | i \in \mathcal{I}\}$, such that both constraints in Eqs. (7) and (8) are satisfied. Moreover, it turns out that it is very useful to represent such a vector of operators as a single block operator $A = \oplus_{i \in I} A_i$, acting on an auxiliary Hilbert space $\mathcal{H}_{\mathrm{aux}} \doteq \oplus_{i \in I} W_i$, where $\mathcal{W}_i \simeq \mathcal{H} \,\forall\, i \in \mathcal{I}$. In terms of the block operator $A \in \oplus_{i \in I} B(W_i)$ the two constraints Eq. (7) and Eq. (8) become

$$A \geqslant 0, \tag{10}$$

and

$$\mathcal{L}(A) = \mathbb{1}, \tag{11}$$

where $\mathcal{L}: \oplus_{i \in I} B(W_i) \to B(H)$ is the linear map

$$\mathcal{L}(A) \doteq \sum_{i \in \mathcal{I}} \int_{\mathbf{G}} \mathrm{d}g U_g A_i U_g^\dagger. \tag{12}$$

The two constraints (10) and (11) define such a convex subset of the space of block operators $\oplus_{i \in I} B(W_i)$, 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 C.

Proposition 1: *The convex set* C, *defined by the constraints* (10) and (11) *is compact in the operator norm.*

Proof: Since C is a subset of a finite dimensional vector space, it enough to show that C is bounded and closed. C is bounded, since for any $A \in \mathsf{C}$, one has $\|A\| \leqslant \mathrm{Tr}[A] = \Sigma_{i \in \mathcal{I}} \mathrm{Tr}[A_i] = \mathrm{Tr}[\mathcal{L}(A)] = d$ (using Eqs. (10) and (11)). Moreover, C is closed. In fact, if $\{A_n\}$ is a Cauchy sequence of points in C, then A_n converges to some block operator $A \in \oplus_{i \in I} B(W_i)$. We claim that A belongs to C. Of course, A satisfies condition (10). As regards condition (11), just notice that the $\mathcal{L}$ is continuous, being linear. Therefore, we have $\|\mathcal{L}(A) - \mathbb{1}\| = \|\mathcal{L}(A - A_n)\| \to 0$, namely A satisfies condition (11). ■

Observation 1: Since the convex set C is compact, it coincides with the convex hull of its extreme points, i.e., any element $A \in \mathsf{C}$ can be written as convex combination of extreme points. The classification of the extreme points of C will be given in Sec. IV.

Observation 2: In this section and all throughout the paper, $\mathbf{G}$ is assumed to be a compact Lie group. Nevertheless, all results clearly hold also if $\mathbf{G}$ is a finite group, with cardinality $|\mathbf{G}|$. In this case, one only has to make the substitution $\int_{\mathbf{G}} \mathrm{d}g \to (1/|\mathbf{G}|)\Sigma_{g \in \mathbf{G}}$. Moreover, since now the probability space $\mathfrak{X}=\mathcal{I}\times\mathbf{G}$ is discrete, there is no need of introducing any operator density, and we simply have

$$P(i,g) = \frac{1}{|\mathbf{G}|} U_g A_i U_g^\dagger. \tag{13}$$

An example of covariant POVM with a finite symmetry group will be given in Sec. VI.

III. SOME RESULTS OF ELEMENTARY GROUP THEORY

Let $\mathbf{G}$ be a compact Lie group and let $\mathrm{d}g$ be the invariant Haar measure on $\mathbf{G}$, normalized such that $\int_{\mathbf{G}} \mathrm{d}g=1$. Consider a finite dimensional Hilbert space $\mathcal{H}$ and represent $\mathbf{G}$ on $\mathcal{H}$ by a unitary (generally projective) representation $\mathsf{R}(\mathbf{G})=\{U_g | g \in \mathbf{G}\}$. The collection of equivalence classes of irreducible representations which show up in the decomposition of $\mathsf{R}(\mathbf{G})$ will be denoted by S. Then $\mathcal{H}$ can be decomposed into the direct sum of orthogonal irreducible subspaces:

$$\mathcal{H} = \bigoplus_{\mu \in S} \bigoplus_{k=1}^{m_\mu} \mathcal{H}_k^\mu, \tag{14}$$

where the index μ labels equivalence classes of irreducible representations (irreps), while the index i is a degeneracy index labeling m_μ different equivalent representations in the class μ. Subspaces carrying equivalent irreps have all the same dimension d_μ and are connected by invariant isomorphisms, namely for any $k,l=1,\ldots,m_\mu$ there is an operator $T_{kl}^\mu \in \mathcal{B}(\mathcal{H})$ such that $\mathsf{Supp}(T_{kl}^\mu)=\mathcal{H}_l^\mu$, $\mathsf{Rng}(T_{kl}^\mu)=\mathcal{H}_k^\mu$, and $[T_{kl}^\mu, U_g]=0\ \forall g \in \mathbf{G}$. Due to Schur lemmas, any operator O in the commutant of the representation $\mathsf{R}(\mathbf{G})$ has the form:

$$O = \sum_\mu \sum_{k,l=1}^{m_\mu} \frac{\mathrm{Tr}[T_{lk}^\mu O]}{d_\mu} T_{kl}^\mu. \tag{15}$$

Using the above-presented formula, the normalization of a covariant POVM, given by Eq. (11), can be rewritten in a simple form. In fact, due to the invariance of the Haar measure $\mathrm{d}g$, we have $[\mathcal{L}(A), U_g]=0\ \forall g \in \mathbf{G}$, i.e., $\mathcal{L}(A)$ belongs to the commutant of $\mathsf{R}(\mathbf{G})$. Then, by exploiting Eq. (15), we rewrite the normalization constraint (11) as

$$\sum_{i \in \mathcal{I}} \mathrm{Tr}[T_{kl}^\mu A_i] = d_\mu \delta_{kl}, \quad \forall\ \mu \in S, \quad \forall\ k,l = 1, \ldots, m_\mu, \tag{16}$$

δ_{kl} denoting the Kronecker delta.

Again, this condition can be recast into a compact form by introducing the auxiliary Hilbert space $\mathcal{H}_{\mathrm{aux}}=\oplus_{i \in I} W_i$, with $\mathcal{W}_i \simeq \mathcal{H}\ \forall i \in \mathcal{I}$, and constructing a block operator with a repeated direct sum of the same operator T_{kl}^μ, i.e.,

$$S_{kl}^\mu = \bigoplus_{i \in \mathcal{I}} S_{kli}^\mu, \quad S_{kli}^\mu = T_\mu^{kl}, \quad \forall\ i \in \mathcal{I}. \tag{17}$$

With this definition, Eq. (16) becomes

$$\mathrm{Tr}[S_{kl}^\mu A] = d_\mu \delta_{kl}, \quad \forall\ \mu \in S, \quad \forall\ k,l = 1, \ldots, m_\mu, \tag{18}$$

where A is the block operator $A = \oplus_{i \in I} A_i$.

IV. EXTREMAL COVARIANT POVMs

This section contains the main result of the paper, namely the characterization of the extremal covariant POVMs with probability space $\mathcal{I}\otimes\mathbf{G}$. Such a characterization will 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 (10) and (11), or, equivalently, by (10) and (18).

Definition 2: *An Hermitian block operator* $P=\oplus_{i\in I}P_i$ *is a perturbation of* $A\in\mathsf{C}$ *if there exists an* $\epsilon>0$ *such that* $A+tP\in\mathsf{C}$ *for any* $t\in[-\epsilon,\epsilon]$.

Clearly, a point $A\in\mathsf{C}$ is extreme if and only if it admits only the trivial perturbation $P=0$.

Lemma 1: *A block operator* $P=\oplus_{i\in I}P_i$ *is a perturbation of* $A\in\mathsf{C}$ *if and only if*

$$\mathsf{Supp}(P)\subseteq\mathsf{Supp}(A), \tag{19}$$

$$\mathrm{Tr}[S^{\mu}_{kl}P]=0, \quad \forall\,\mu\in S, \quad \forall\,k,l=1,\ldots,m_{\mu}. \tag{20}$$

Proof: Condition (19) is equivalent to the existence of an $\epsilon>0$ such that $A+tP\geqslant 0$ for all $t\in[-\epsilon,\epsilon]$ (see Lemma 1 of Ref. 7). On the other hand, condition (20) is equivalent to require that $A+tP$ satisfies the normalization constraint (16) for all $t\in[-\epsilon,\epsilon]$. ■

Observation: Note that, due to the block form of both P and A, condition (19) is equivalent to

$$\mathsf{Supp}(P_i)\subseteq\mathsf{Supp}(A_i), \quad \forall\,i\in\mathcal{I}. \tag{21}$$

Using the previous lemma, we can obtain a first characterization of extremality:

Theorem 1 (Minimal support condition): *A point* $A\in\mathsf{C}$ *is extremal if and only if for any* $B\in\mathsf{C}$,

$$\mathsf{Supp}(B)\subseteq\mathsf{Supp}(A)=\Rightarrow A=B. \tag{22}$$

Proof: Suppose A extremal. Then, if $\mathsf{Supp}(B)\subseteq\mathsf{Supp}(A)$, according to Lemma 1, $P=A-B$ is a perturbation of $A\in\mathsf{C}$. Hence, P must be zero. Conversely, if P is a perturbation of A, then $B=A+tP$ is an element of C for some $t\neq 0$. Due to Lemma 1, we have $\mathsf{Supp}(B)\subseteq\mathsf{Supp}(A)$. Then, condition (22) implies $B=A+tP=A$, i.e., $P=0$. Therefore, A is extremal. ■

Corollary 1: *If* $A\in\mathsf{C}$ *and* $\mathrm{rank}(A)=1$, *then A is extremal.*

Proof: Since $\mathrm{rank}(A)=1$, then, for any $B\in\mathsf{C}$, the condition $\mathsf{Supp}(B)\subseteq\mathsf{Supp}(A)$ implies $B=\lambda A$ for some $\lambda>0$. Moreover, since both A and B are in C, from Eq. (18) 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 (22) then ensures that A is extremal. ■

A deeper characterization of extremal covariant POVMs can be obtained by using the following lemma.

Lemma 2: *Let A be a point of* C, *represented as*

$$A=\bigoplus_{i\in\mathcal{I}}X_i^{\dagger}X_i, \tag{23}$$

and define $\mathcal{H}_i=\mathsf{Rng}(X_i)$ *the range of* X_i. *A block operator* $P=\oplus_{i\in I}P_i$ *is a perturbation of A if and only if*

$$P_i=X_i^{\dagger}Q_iX_i, \quad \forall\,i\in\mathcal{I}, \tag{24}$$

for some Hermitian $Q_i\in\mathcal{B}(\mathcal{H}_i)$, *and*

$$\sum_{i\in\mathcal{I}}\mathrm{Tr}[S^{\mu}_{kli}X_i^{\dagger}Q_iX_i]=0. \tag{25}$$

Proof: First of all, the form (24) is equivalent to condition (19). In fact, if P has the form (24), then clearly $\mathsf{Supp}(P)\subseteq\mathsf{Supp}(A)$. Conversely, if we assume condition (19) and write

$P=\oplus_{i\in I}P_i$, we have necessarily $\mathsf{Supp}(P_i)\subseteq\mathsf{Supp}(X_i^\dagger X_i)=\mathsf{Supp}(X_i)$. Exploiting the singular value decomposition $X_i=\Sigma_{n=1}^{r_i}\lambda_n^{(i)}|w_n^i\rangle\langle v_n^i|$, where $\{|v_n^i\rangle\}$ and $\{|w_n^i\rangle\}$ are orthonormal bases for $\mathsf{Supp}(X_i)$ and $\mathsf{Rng}(X_i)$ respectively, we have that any Hermitian operator P_i satisfying $\mathsf{Supp}(P_i)\subseteq\mathsf{Supp}(X_i)$ has the form $P_i=\Sigma_{m,n}p_{mn}^{(i)}|v_m\rangle\langle v_n|$, whence it can be written as $P_i=X_i^\dagger Q_i X_i$, for some suitable Hermitian operator $Q_i\in\mathcal{B}(\mathsf{Rng}(X))$. Once the equivalence between the form (24) and condition (19) is established, relation (25) follows directly from Eq. (20). ■

Observation: According to the previous lemma, a perturbation of A is completely specified by a set of Hermitian operators $\{Q_i\in\mathcal{B}(\mathcal{H}_i)|i\in\mathcal{I}\}$, where $\mathcal{H}_i=\mathsf{Rng}(X_i)$. Such operators can be cast into a single block operator $Q\in\oplus_{i\in I}B(H_i)$ by defining

$$Q=\bigoplus_{i\in\mathcal{I}}Q_i. \tag{26}$$

In terms of the block operator Q we have the following:

Lemma 3: *Let $A=\oplus_{i\in I}X_i^\dagger X_i$ be a point of C. Define the block operators*

$$F_{kl}^\mu=\bigoplus_{i\in\mathcal{I}}X_i S_{kli}^\mu X_i^\dagger. \tag{27}$$

Then A admits a perturbation if and only if there exists an Hermitian block operator $Q\in\oplus_{i\in I}B(H_i)$ such that

$$\mathrm{Tr}[F_{kl}^\mu Q]=0,\quad\forall\,\mu\in\mathsf{S},\quad\forall\,k,l=1,\ldots,m_\mu. \tag{28}$$

Proof: Using the definition of F_{kl}^μ and the cyclic property of the trace, it is immediate to see that Eq. (28) is equivalent to Eq. (25). ■

The previous lemma enables us to characterize the extremal points of C.

Theorem 2 (Spanning set condition): *Let $A=\oplus_{i\in I}X_i^\dagger X_i$ be a point of C, and $\mathsf{F}=\{F_{kl}^\mu|\mu\in\mathsf{S},k,l=1,\ldots,m_\mu\}$ be the set of block operators defined in Lemma 3. Then, A is extremal if and only if*

$$\mathsf{Span}(\mathsf{F})=\bigoplus_{i\in\mathcal{I}}\mathcal{B}(\mathcal{H}_i), \tag{29}$$

where $\mathcal{H}_i=\mathsf{Rng}(X_i)$.

Proof: A is extremal iff it admits only the trivial perturbation $P=0$. Equivalently, due to Lemma 3, A is extremal iff the only Hermitian operator $Q\in\oplus_{i\in I}B(H_i)$ that satisfies Eq. (28) is the null operator $Q=0$. Let us decompose the Hilbert space $\mathcal{K}=\oplus_{i\in I}B(H_i)$, as $\mathcal{K}=\mathsf{Span}(\mathsf{F})\oplus\mathsf{Span}(\mathsf{F})^\perp$, 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 $\mathsf{Span}(\mathsf{F})^\perp$ is the null operator. This is equivalent to the condition $\mathsf{Span}(\mathsf{F})^\perp=\{0\}$, i.e., $\mathcal{K}=\mathsf{Span}(\mathsf{F})$. ■

Corollary 2: *Let $A=\oplus_{i\in I}X_i^\dagger X_i$ be a point of C, and let define $r_i=\mathrm{rank}(X_i)$. If A is extremal, then the following relation holds*

$$\sum_{i\in\mathcal{I}}r_i^2\leqslant\sum_{\mu\in\mathsf{S}}m_\mu^2. \tag{30}$$

Proof: For an extreme point of C, relation (29) implies that the cardinality of the set F is greater than the dimension of $\mathcal{K}=\oplus_{i\in I}B(H_i)$. Then, the upper bound (30) follows from $\dim\mathcal{K}=\Sigma_{i\in\mathcal{I}}r_i^2$ and from the fact that $|\mathsf{F}|=\Sigma_{\mu\in\mathsf{S}}m_\mu^2$. ■

Observation: If the group-representation $\mathsf{R}(\mathbf{G})$ is irreducible, then its Clebsch-Gordan decomposition contains only one term $\bar{\mu}$ with multiplicity $m_{\bar{\mu}}=1$. Then, bound (30) becomes $\Sigma_{i\in\mathcal{I}}r_i^2\leqslant 1$, namely for an extremal $A=\oplus_{i\in I}A_i$, 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 1). In terms of the corresponding covariant POVM $M(i,g)=U_gA_iU_g^\dagger$, one has $M(i,g)=0$ for any $i\neq i_0$, i.e., corresponding to events in the probability space that never occur.

V. EXTREMAL COVARIANT POVMs IN THE PRESENCE OF NONTRIVIAL STABILITY GROUPS

In Sec. IV, we obtained a characterization of extremal covariant POVMs whose probability space is $\mathfrak{X}=\mathcal{I}\times\mathbf{G}$ for some finite index set $\mathcal{I}$. The framework we outlined is suitable for a straightforward generalization to the case $\mathfrak{X}=\cup_{i\in\mathcal{I}}\mathbf{G}/\mathbf{G}_i$, where $\mathbf{G}_i$ are compact subgroups of $\mathbf{G}$.

In this case, 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

$$M(x_i)=U_{g_i(x_i)}A_iU^{\dagger}_{g_i(x_i)},\tag{31}$$

where $A_i\geqslant 0$, 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. (16). 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

$$[A_i,U_h]=0,\qquad\forall\, h\in\mathbf{G}_i.\tag{32}$$

The commutation constraint (32) can be simplified by decomposing each representation $\mathsf{R}(\mathbf{G}_i)=\{U_h|h\in\mathbf{G}_i\}$ into irreps

$$U_h=\bigoplus_{\nu\in\mathsf{S}_i}U_h^{\nu_i}\otimes\mathbb{1}_{m_{\nu_i}},\tag{33}$$

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

$$\mathcal{H}=\bigoplus_{\nu_i\in\mathsf{S}_i}\mathcal{H}_{\nu_i}\otimes\mathbb{C}^{m_{\nu_i}},\tag{34}$$

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 (32) is equivalent to the block form

$$A_i=\bigoplus_{\nu_i\in\mathsf{S}_i}\mathbb{1}_{\nu_i}\otimes A_{i,\nu_i},\tag{35}$$

where $A_{i,\nu_i}\geqslant 0$ 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

$$A=\bigoplus_{\omega\in\Omega}A_\omega,\tag{36}$$

where $A_\omega\doteq A_{i,\nu_i}$. Furthermore, we define the block operators

$$S^{\mu}_{kl}=\bigoplus_{\omega\in\Omega}S^{\mu}_{kl\omega},\tag{37}$$

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. (16), becomes equivalent to

$$\mathrm{Tr}[S^{\mu}_{kl}A]=\delta_{kl}d_{\mu}.\tag{38}$$

Now we call D the convex set of block operators $A=\oplus_{\omega\in\Omega}A_\omega$, defined by the two conditions $A\geqslant 0$ and Eq. (38). Such a convex set is in one-to-one affine correspondence with the convex set of covariant POVMs with probability space $\mathfrak{X}=\cup_{i\in\mathcal{I}}\mathbf{G}/\mathbf{G}_i$. Since the constraints defining D are formally the same defining the convex set C, we can exploit the characterization of extremal points of the previous section. In particular, Corollary 2 becomes

Corollary 3: *Let* $A=\oplus_{\omega\in\Omega}X^{\dagger}_{\omega}X_{\omega}$ *be a point of* D, *and define* $r_{i,\nu_i}\equiv r_{\omega}=\mathrm{rank}(X_{\omega})$. *If A is extremal, then the following relation holds:*

$$\sum_{i\in\mathcal{I}}\sum_{\nu_i\in\mathsf{S}_i} r^2_{i,\nu_i}\leqslant\sum_{\mu\in\mathsf{S}} m^2_{\mu}. \tag{39}$$

Observation: As in the case of Corollary 2, if the representation $\mathsf{R}(\mathbf{G})$ 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 $\omega\neq\omega_0$.

VI. APPLICATIONS

Here we give two examples of the use of the characterization of extremal POVMs in the solution of concrete optimization problems.

A. State discrimination with mutually unbiased bases

1. Two Fourier transformed bases

Here we consider a case of state discrimination where the set of signal states is the union of two mutually unbiased bases (MUBs),[11] related by Fourier transform. Precisely, let $\mathcal{H}$ be a d-dimensional Hilbert space, and consider the orthornormal bases $\mathcal{B}_1=\{|n\rangle\,|\,n=0,\ldots,d-1\}$ and $\mathcal{B}_2=\{|e_n\rangle\,|\,n=0,\ldots,d-1\}$, where $|e_n\rangle=(1/\sqrt{d})\Sigma_{m=0}^{d-1}\omega^{mn}|m\rangle$, $\omega=\exp(2\pi i/d)$. $\mathcal{B}_1$ and $\mathcal{B}_2$ are mutually unbiased, namely $|\langle m|e_n\rangle|^2=1/d$ for any m,n. Consider the two sets of states defined by $\mathcal{S}_1=\{\rho_{1n}=|n\rangle\langle n|\,|\,n=0,\ldots,d-1\}$ and $\mathcal{S}_2=\{\rho_{2n}=|e_n\rangle\langle e_n|\,|\,n=0,\ldots,d-1\}$. Now the problem is to determine with minimum error probability the state of the system, which is randomly prepared either in a state of $\mathcal{S}_1$ with probability p/d, or in a state of $\mathcal{S}_2$ with probability $(1-p)/d$.

Exploiting the results of the present paper it is immediate to find the measurement that minimizes the error probability. In fact, let us consider the irreducible representation of the group $\mathbf{G}=\mathbb{Z}_d\times\mathbb{Z}_d$ given by

$$\mathsf{R}(\mathbf{G})=\left\{U_{pq}=\sum_{n=0}^{d-1}\omega^{qn}|n\oplus p\rangle\langle n|,(p,q)\in\mathbb{Z}_d\times\mathbb{Z}_d\right\}, \tag{40}$$

where $\oplus$ denotes addition modulo d. Then, the sets $\mathcal{S}_1$ and $\mathcal{S}_2$ are the group orbits of the initial states ρ_{10} and ρ_{20}, respectively. Moreover, the states ρ_{10} and ρ_{20} have nontrivial stability groups $\mathbf{G}_1$ and $\mathbf{G}_2$, defined by the unitaries $\mathsf{R}(\mathbf{G}_1)=\{U_{0q}|q\in\mathbb{Z}_d\}$ and $\mathsf{R}(\mathbf{G}_2)=\{U_{p0}|p\in\mathbb{Z}_d\}$. Therefore, signal states are in one-to-one correspondence with points of the probability space $\mathfrak{X}=\mathbf{G}/\mathbf{G}_1\cup\mathbf{G}/\mathbf{G}_2$, such points being denoted by couples (i,n) where $i\in\{1,2\}$ and $n\in\mathbb{Z}_d$. For the discrimination we can consider without loss of generality a covariant POVM, of the form of Eq. (31), where now the group element g is the couple $(p,q)\in\mathbb{Z}_d\times\mathbb{Z}_d$. Moreover, since the probabilities are linear in the POVM, in the minimization of the error probability we can restrict the attention to extremal covariant POVMs. Now, the representation $\mathsf{R}(\mathbf{G})$ is irreducible, whence Corollary 3 requires either $A_1=0$ or $A_2=0$ in Eq. (31). This means that either the states in $\mathcal{S}_1$ or the states in $\mathcal{S}_2$ are never detected. Moreover, since the states within a given set, either $\mathcal{S}_1$ or $\mathcal{S}_2$, are orthogonal, they can be perfectly distinguished among themselves. Therefore, the optimal POVM is $P^{(1)}(i,n)=\delta_{i1}|n\rangle\langle n|$ if $p\geqslant 1/2$, and $P^{(2)}(i,n)=\delta_{i2}|e_n\rangle\langle e_n|$ otherwise. In particular, if $p=1/2$, an experimenter who tries to discriminate states of two Fourier transformed bases cannot do anything better than randomly choosing one of the orthogonal measurements $P^{(1)}$ and $P^{(2)}$. This is the working principle of the BB84 crypthographic protocol.[13]

2. Mutually unbiased bases in prime dimension

If the dimension of the Hilbert space $\mathcal{H}$ is a prime number, then there are $d+1$ MUBs that are generated by the irreducible representation

$$\mathsf{R}(\mathbf{G}) = \left\{ U_{pq} = \sum_{n=0}^{d-1} \omega^{qn} |n \oplus p\rangle\langle n|, (p,q) \in \mathbb{Z}_d \times \mathbb{Z}_d \right\}$$

via the construction by Wootters and Fields[11] (see also Ref. 12).

In this case, the result of the previous paragraph can be immediately generalized to a case of state discrimination with more than two MUBs. Again, due to the irreducibility of the representation $\mathsf{R}(\mathbf{G})$, an extremal POVM is the group orbit of a single operator. Therefore, denoting by $\mathcal{S}_i$ the set of states associated to the basis $\mathcal{B}_i$, and by p_i/d the probability of extracting a state from $\mathcal{S}_i$ $(\Sigma_{i\in\mathcal{I}} p_i = 1)$, we have that the covariant POVM which discriminates the signal states with minimum error probability is the orthogonal measurement onto the basis $\mathcal{B}_{\bar{l}}$ such that $p_{\bar{l}} = \max_{l\in\mathcal{I}}\{p_l\}$.

3. Mutually unbiased bases in dimension p^r

In the case of Hilbert space dimension $d = p^r$, where p is prime number, $d+1$ MUBs can be constructed by introducing a projective representation of the Abelian group $\widetilde{\mathbf{G}} = \mathbb{F}_d \times \mathbb{F}_d$, where $\mathbb{F}_d$ is the finite field of cardinality d, considered here as an additive group. In order to apply the results of the paper to this case, we first outline the method for constructing MUBs presented in Ref. 12, to which we refer for details and for the explicit proofs.

Consider an orthonormal basis for $\mathcal{H}$, denoted as $\{|n\rangle | n \in \mathbb{F}_d\}$, in which basis elements are labeled by elements of the field. Then, introduce the projective representation

$$\mathsf{R}(\widetilde{\mathbf{G}}) = \{U_p V_q | (p,q) \in \mathbb{F}_d \times \mathbb{F}_d\}, \tag{41}$$

where U_p, V_q are the unitary operators uniquely defined by the relations

$$U_p|n\rangle = |n+p\rangle,$$

$$V_q|n\rangle = \langle q,n\rangle |n\rangle. \tag{42}$$

Here, $\langle a,b\rangle \doteq \chi(a\cdot b)$, where $\chi(x)$ is any nontrivial character of the additive group $\mathbb{F}_d$, and $a+b$ $(a\cdot b)$ denote the addition (product) in the finite field $\mathbb{F}_d$. With the above definition $\langle a,b\rangle$ is a symmetric bicharacter for the additive group $\mathbb{F}_d$, namely $|\langle a,b\rangle| = 1, \langle a,b\rangle = \langle b,a\rangle$, and $\langle a,b+c\rangle = \langle a,b\rangle\langle a,c\rangle$, for any $a,b,c \in \mathbb{F}_d$. By definition (42), the operators U_p, V_q commute up to a phase, namely

$$V_q U_p = \langle p,q\rangle U_p V_q. \tag{43}$$

To construct $d+1$ MUBs, it is useful to introduce $d+1$ sets of the unitary operators, each set being labeled by an index $i \in \mathbb{F}_d \cup \{\infty\}$ (∞ is just a label which denotes an additional value, not in $\mathbb{F}_d$, of the index i). The $d+1$ sets of unitary operators are defined by

$$W(i,j) \doteq \begin{cases} \alpha(i,j) U_j V_{i\cdot j}, & i \in \mathbb{F}_d \\ V_j, & i = \infty, \end{cases} \tag{44}$$

where $\alpha(i,j)$ are suitable phase factors (see Ref. 12), chosen in such a way that, for any fixed i, the operators $W(i,j)$ form a unitary representation of the additive group $\mathbb{F}_d$, namely

$$W(i,j)W(i,l) = W(i,j+l), \quad \forall\, j,l \in \mathbb{F}_d. \tag{45}$$

Since the group $\mathbb{F}_d$ is Abelian, for fixed i the operators $W(i,j)$ can be diagonalized on the same basis, denoted by $\mathcal{B}_i$. The above construction guarantees that the bases $\{\mathcal{B}_i | i \in \mathbb{F}_d \cup \{\infty\}\}$ are all mutually unbiased. Moreover, the one-dimensional projector $P(i,k)$ onto the kth element of the basis $\mathcal{B}_i$ can be written as[12]

$$P(i,k) = d^{-1} \sum_{j\in\mathbb{F}_d} \overline{\langle j,k\rangle} W(i,j). \tag{46}$$

Now we exploit the above-noted construction to show that, for any $i \in \mathbb{F}_d \cup \{\infty\}$, the set of states $\mathcal{S}_i = \{\rho_{ik} = P_{ik} | k \in \mathbb{F}_d\}$ is the orbit of the initial state $\rho_{i0} = P_{i0}$ under the action of the representation $\mathsf{R}(\widetilde{G})$.

For $i \in \mathbb{F}_d$, we have indeed

$$\begin{aligned} U_p V_q P(i,k) V_q^\dagger U_p^\dagger &= d^{-1} \sum_{j \in \mathbb{F}_d} \overline{\langle j,k \rangle} \alpha(i,j) U_p V_q U_j V_{k \cdot j} V_{-q} U_{-p} \\ &= d^{-1} \sum_{j \in \mathbb{F}_d} \overline{\langle j,k \rangle} \langle j,q \rangle \langle i \cdot j, -p \rangle \alpha(i,j) U_j V_{i \cdot j} \\ &= d^{-1} \sum_{j \in \mathbb{F}_d} \overline{\langle j,k - q + i \cdot p \rangle} W(i,j) = P(i,k - q + i \cdot p), \end{aligned} \tag{47}$$

where we used Eqs. (46), (44), (43), and the properties $\langle a,-b \rangle = \overline{\langle a,b \rangle}, \langle a,b+c \rangle = \langle a,b \rangle \langle a,c \rangle$, and $\langle a, b \cdot c \rangle = \langle a \cdot b, c \rangle$. Similarly, for $i = \infty$ we obtain

$$U_p V_q P(\infty,k) V_q^\dagger U_p^\dagger = P(\infty,k+p). \tag{48}$$

Notice that from Eqs. (47), (48) it also follows that for any $i \in \mathbb{F}_d \cup \{\infty\}$, the stability group of $\rho_{i0} = P_{i0}$ is the additive group $\mathbb{F}_d$, which is projectively represented by the unitaries $\{U_p V_{i \cdot p} | p \in \mathbb{F}_d\}$ for $i \in \mathbb{F}_d$, and by the unitaries $\{V_q | q \in \mathbb{F}_d\}$ for $i = \infty$.

In the problem of state discrimination where the state ρ_{ik} is randomly drawn from the set $\mathcal{S}_i$ with probability p_i / d, we can then use the results about extremal covariant POVMs with nontrivial stability group to find the minimum error POVM. Again, since the representation $\mathsf{R}(\widetilde{\mathbf{G}})$ is irreducible,[14] an extremal POVM must be the group orbit of a single operator. The optimal POVM for state discrimination is then the orthogonal measurement onto the basis $\mathcal{B}_{\bar{l}}$ which occurs with highest probability $p_{\bar{l}} = \max_i \{p_i\}$.

B. Maximization of the mutual information

A frequent problem in quantum communication is to find the POVM $P_i, i \in \mathcal{I}$, that maximizes the mutual information with a given set of signal states $\mathcal{S} = \{\rho_j | j \in \mathcal{J}\}$. Denoting by p_j the probability of the signal state ρ_j, by $q_i = \Sigma_{j \in \mathcal{J}} p_j \mathrm{Tr}[M_i \rho_j]$ the overall probability of the outcome i, and by $p_{ij} = p_j \mathrm{Tr}[M_i \rho_j]$ the joint probability of the outcome j with the state ρ_i, the mutual information is defined as

$$I = H(\{p_{ij}\}) - H(\{p_i\}) - H(\{q_j\}), \tag{49}$$

where $H(\{p_i\}) \doteq \Sigma_i - p_i \log(p_i)$ is the Shannon entropy. As in the minimization of a Bayes cost,[4,5] 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 $\mathfrak{X} = \mathcal{I} \otimes \mathbf{G}$, for some finite index set $\mathcal{I}$.[15,9] 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 the following basic properties of the mutual information (for the proofs, see Ref. 15), we can readily obtain a bound about the cardinality of the index set $\mathcal{I}$.

Property 1: *The mutual information is a convex functional of the POVM.*

Property 2: *In the maximization of the mutual information, one can consider without loss of generality POVMs made of rank-one operators.*

Consider a covariant POVM $P(i,g) = (1/(|\mathbf{G}|) U_g A_i U_g^\dagger$. Due to Property 1, in the maximization of the mutual information we can consider extremal covariant POVMs. Then, from Corollary 2, we have the bound $\Sigma_{i\in\mathcal{I}}\text{rank}(A_i)^2 \leqslant \Sigma_{\mu\in\mathsf{S}} m_\mu^2$. Due to Property 2, this also implies that the number of (rank-one) operators A_i must be smaller than $\Sigma_{\mu\in\mathsf{S}} m_\mu^2$. Therefore, we can assume without loss of generality

$$|\mathcal{I}| \leqslant \sum_{\mu\in\mathsf{S}} m_\mu^2. \tag{50}$$

This provides an alternative derivation of the bound given in Ref. 9. Finally, if the representation $\mathsf{R}(\mathbf{G})$ is irreducible, the bound gives $|\mathcal{I}|=1$, namely the probability space is $\mathfrak{X}\simeq\mathbf{G}$, according to the classic result of Ref. 15.

ACKNOWLEDGMENTS

We acknowledge the referee for an observation that enriched the paper, by stimulating us to treat in detail the problem of state discrimination with $d+1$ MUBs in the general case of Hilbert space dimension p^r. This work has been supported by Ministero Italiano dell'Università e della Ricerca (MIUR) through FIRB (bando 2001) and PRIN 2005.

[1] I. L. Chuang and M. A. Nielsen, *Quantum Information and Quantum Computation* (Cambridge University Press, Cambridge, 2000).

[2] A. S. Holevo, J. Multivariate Anal. **3**, 337 (1973).

[3] A. Peres, *Quantum Theory: Concepts and Methods* (Kluwer Academic, Dordrecht, 1993), pp. 279–289.

[4] C. W. Helstrom, *Quantum Detection and Estimation Theory* (Academic, New York, 1976).

[5] A. S. Holevo, *Probabilistic and Statistical Aspects of Quantum Theory* (North Holland, Amsterdam, 1982).

[6] G. M. D'Ariano, J. Math. Phys. **45**, 3620 (2004).

[7] G. Chiribella and G. M. D'Ariano, J. Math. Phys. **45**, 4435 (2004).

[8] P. Shor, in *Quantum Communication, Computing, and Measurement 3*, edited by P. Tombesi and O. Hirota (Kluwer, Dordrecht, 2001); LANL e-print quant-ph/0009077.

[9] T. Decker, eprint quant-ph/0509122.

[10] The proof of this statement is the straightforward generalization of the corresponding proof for transitive group actions (see Ref. 5, pp. 166–169).

[11] W. K. Wootters and B. D. Fields, Ann. Phys. (N.Y.) **191**, 363 (1989).

[12] K. R. Parthasarathy, eprint quant-ph/0408069.

[13] C. H. Bennett and G. Brassard, in *Proceedings of the IEEE International Conference on Computers, Systems and Signal Processing, Bangalore, India* (IEEE, New York, 1984), pp. 175–179.

[14] If an operator O commutes with all the unitaries U_pV_q in the representation $\mathsf{R}(\widetilde{\mathbf{G}})$, then it must commute at least with U_p and V_q. In other words, O must be diagonal both on the eigenvectors of U_p and on the eigenvectors of V_q. But these two bases are mutually unbiased, whence the only possibility is O proportional to the identity, i.e., the representation $\mathsf{R}(\widetilde{\mathbf{G}})$ is irreducible.

[15] E. B. Davies, IEEE Trans. Inf. Theory **24**, 596 (1978).