Family of generalized "pretty good" measurements and the minimal-error pure-state discrimination problems for which they are optimal

Carlos Mochon*
Institute for Quantum Information, California Institute of Technology, Pasadena, California 91125, USA
(Received 25 December 2005; published 22 March 2006)

Given a quantum pure state chosen from a set with some *a priori* probabilities, what is the optimal measurement needed to correctly guess the given state? We show that a good choice is the family of square-root or "pretty good" measurements, as each measurement in the family is optimal for at least one discrimination problem with the same quantum states but possibly different *a priori* probabilities. Furthermore, the map from measurement to discrimination problems can be explicitly described. In fact, for linearly independent states, every pair of discrimination problem and optimal measurement can be explicitly generated this way.

DOI: 10.1103/PhysRevA.73.032328 PACS number(s): 03.67.Lx

I. INTRODUCTION

Given a set of n pure states $|\psi_1\rangle,\ldots,|\psi_n\rangle$ and associated probabilities $p_1,\ldots,p_n$, the problem of minimal-error pure-state discrimination can be described by the following game: Alice chooses an index $i\in\{1,\ldots,n\}$ using the probability distribution $\{p_i\}$ and hands a copy of the state $|\psi_i\rangle$ to Bob, who must guess i by measuring the state. In the minimal error setting, we seek the positive operator valued measure (POVM) with elements $E_1,\ldots,E_N$ that maximizes the probability of success,

$$p_{\text{succ}}=\sum_{i=1}^{n} p_i\langle\psi_i|E_i|\psi_i\rangle. \tag{1}$$

There are a number of scenarios where such a problem arises. Many experimental setups require a measurement of an incoming signal that consists of a small number of nonorthogonal quantum states, and one seeks to implement the optimal distinguishing measurement. For such cases, an approximate numerical solution is often sufficient and one can be obtained by solving the associated semidefinite program (SDP) first identified by Yuen *et al.* [1].

However, the problem also arises in a number of quantum computation and communication contexts. Consider, for instance, the recent search for optimal measurements for hidden subgroup problems [2,3]. These problems involve exponentially many states to be distinguished, with typical thousand-bit oracles outputting one out of 2^{1000} states. Numerical searches for the optimal measurement are no longer tractable. Furthermore, in many cases what one seeks is the scaling of some parameter (say the failure probability) with the number of qubits. For such situations, a theoretical understanding of the problem is essential.

Often such discrimination problems are approached by guessing a measurement and then verifying its optimality. The square-root or "pretty good" measurement [4] was proven optimal for nonadaptive queries to a dihedral hidden subgroup oracle [3]. In fact, the pretty good measurement (PGM) is likely to be optimal for many of the the problems where a large amount of symmetry is available. However, when the PGM is not optimal, one is left with the following question: What other measurements should be considered? Is there a tool chest of measurements that can be tried?

For a given discrimination problem specified by *a priori* probabilities $p_1,\ldots,p_n$ and associated pure states $|\psi_1\rangle,\ldots,|\psi_n\rangle$, a good set of measurements is given by the PGMs associated with the same pure states but using different sets of probabilities $\tilde{p}_1,\ldots,\tilde{p}_n$. We shall refer to these measurements as the generalized PGMs.

Note that, strictly speaking, the PGM is not a measurement but rather a map from a set of states and probabilities into the set of measurements. In terms of the numbers $\{\tilde{p}_i\}$, the above map is just the standard PGM. However, the numbers $\{\tilde{p}_i\}$ should be considered as functions (to be discussed below) of the true *a priori* probabilities $\{p_i\}$. In such a case, the map from states and $\{p_i\}$ is no longer the usual PGM, and we refer to it as the generalized PGM to avoid confusion.

The main result of this paper is that each of the generalized PGMs is optimal for at least one discrimination problem defined with the same pure states, and a set of *a priori* probabilities given by the expression

$$p_i=\frac{C}{\langle\psi_i|\sqrt{\tilde{\rho}^{-1}}|\psi_i\rangle}, \tag{2}$$

where $\tilde{\rho}=\Sigma_i\tilde{p}_i|\psi_i\rangle\langle\psi_i|$ and C is a constant that normalizes the probabilities to sum to 1.

Though certainly the matrix square-root cannot be solved numerically significantly faster than an SDP, the purpose of the above expression is to provide an analytical pairing of measurements with the problems for which they are optimal. Of course, a closed form for the inverse of the above map (i.e., $\{\tilde{p}_i\}$ in terms of $\{p_i\}$) would be more exciting. However, even the above form allows one to generate pairs of pure-state discrimination problems with their associated optimal measurement in terms of a closed-form analytical expression.

This is especially powerful for the case of linearly independent states, where we can prove that every pair of discrimination problem and optimal measurement can be gen-

*Electronic address: carlosm@theory.caltech.edu

erated in this fashion, that is, the space of generalized PGMs contains all optimal measurements for discrimination problems involving the same quantum states. In fact, when the states cannot be separated into orthogonal subspaces (see Definition 2), the mapping between probabilities $\{\tilde{p}_i\}$ defining the generalized measurement and *a priori* probabilities $\{p_i\}$ is one-to-one.

For linearly dependent states, the mapping is many-to-many and a generalization of Eq. (2) will be presented in Sec. III. Though we conjecture that all pairs of pure-state discrimination problems together with their optimal measurements can be generated from this map, we shall not prove it in this paper.

We note that while the work of this paper deals only with discrimination of pure quantum states, there exists a special class of mixed state discrimination that is also trivially covered: if the Hilbert space can be partitioned as a direct sum of subspaces $\mathcal{H}=\oplus_i \mathcal{H}_i$ such that the mixed states are simply mixtures of one pure state in each subspace

$$\rho_j = \sum_i a_i |\alpha_{i,j}\rangle\langle\alpha_{i,j}|, \tag{3}$$

where $|\alpha_{i,j}\rangle \in \mathcal{H}_i$ for all j, then the optimal discrimination strategy is to project into a subspace $\mathcal{H}_i$, and then use the optimal pure-state discrimination measurement for that subspace. This generalization is sufficient to cover the case of the dihedral hidden subgroup problem [3].

The rest of the paper is organized as follows: the remainder of the Introduction covers the history of the problem including a list of cases where the minimal-error state discrimination can be solved in a nice form. We also quickly review the optimality conditions for state discrimination.

Section II covers the case of linearly independent states, which is particularly simple because each problem has a unique optimal measurement that consists of a set of one-dimensional projectors. After a review of some of the properties of the optimal measurement for this case (many of which have appeared previously in the literature), we prove the main result: that the generalized PGMs are optimal. An alternative derivation which does not make use of the optimality conditions is given in Appendix A.

Section III covers the case of linearly dependent pure states. The basic principles are the same, but the mappings are no longer one-to-one, which introduces extra complications.

A. History

The problem of minimal-error state discrimination (including its extension to mixed states) dates back to the 1970s where independently Holevo [5] and Yuen *et al.* [1] identified a set of necessary and sufficient conditions for a POVM to be optimal (though only the latter paper proved that the conditions were sufficient, and related the problem to the theory of convex optimization).

There are a number of cases for which the problem has been solved. The problem of distinguishing between two (pure or mixed) states was solved by Helstrom [6]. Yuen *et al.* [1] solved the case of a set of equiprobable pure states that contain a subset that sums to the identity (i.e., if $\Sigma_i a_i|\psi_i\rangle\langle\psi_i| = I$ for some $a_i \geq 0$, then the set $E_i = a_i|\psi_i\rangle\langle\psi_i|$ is an optimal measurement). Andersson *et al.* [7] solved the case involving the three states $|0\rangle$ and $|\pm\rangle = \cos\theta|0\rangle \pm \sin\theta|1\rangle$, where the last two states have equal *a priori* probabilities.

There are also a large number of cases for which the regular PGM is optimal: for instance, a set of symmetric pure states such that $|\psi_i\rangle = U^i|\psi_0\rangle$ and $U^n = I$ [8]. This has been generalized to mixed states and larger symmetries [9,10]. A necessary and sufficient condition for the regular PGM to be optimal for linearly independent pure states was derived in Ref. [11], which states that the diagonal elements of the square root of the Gram matrix must all be equal.

Furthermore, Eldar and Forney [12] have shown that the regular PGM is always optimal when maximizing the "least-squares error" which for linearly independent pure states can be written as $\Sigma_i p_i \mathrm{Re}[\langle E_i|\psi_i\rangle]$ for a POVM composed of one-dimensional projectors $|E_i\rangle$. A similar result was proven by Holevo [13].

There are also a number of properties of the optimal measurement that are known. Kennedy [14] proved that the optimal measurement for a set of linearly independent pure states is a von Neumann measurement. The results have been generalized to the case of mixed states that can be decomposed as a mixture of linearly independent states [15]. On the other hand, Hunter [16] has pointed out that there are cases where the optimal strategy is simply to choose the most likely state without making any measurement, though such a situation can only occur when discriminating mixed states or when all the pure states are identical. Finally, a number of numerical and iterative approximation algorithms for state discrimination exist, including one by Helstrom [17], though their performance relative to standard numerical algorithms for SDPs has not been examined.

Many variations of the state discrimination problem have also been studied in the literature, including the case where one wishes to maximize the mutual information between Alice and Bob, and the unambiguous case where Bob may only output if he is certain of being correct (and otherwise outputs nothing). We refer the reader to the review article by Bergou *et al.* [18] for a survey of some of these alternatives.

B. Optimality conditions

Though the original optimality conditions were found in Refs. [1,5], we shall follow the modern treatment from Refs. [2,19]. The problem of minimal-error pure-state discrimination is the solution to the SDP,

$$\text{Maximize:} \quad \sum_i p_i \langle\psi_i|E_i|\psi_i\rangle, \tag{4}$$

$$\text{Subject to:} \quad E_i \geq 0 \quad \text{for all } i, \tag{5}$$

$$\sum_i E_i = I, \tag{6}$$

where the maximization is to be carried out over positive semidefinite matrices $E_1, \ldots, E_n$. The dual SDP is given by

$$\text{Minimize:} \quad \mathrm{Tr}[Z], \tag{7}$$

$$\text{Subject to:} \quad Z \geqslant p_i|\psi_i\rangle\langle\psi_i| \quad \text{for all } i, \tag{8}$$

where the minimization variable Z is a positive semidefinite matrix.

Both problems are strictly feasible and there is no duality gap. Furthermore, a pair of optimal solutions for the SDPs must satisfy the complementary slackness conditions

$$(Z - p_i|\psi_i\rangle\langle\psi_i|)E_i = E_i(Z - p_i|\psi_i\rangle\langle\psi_i|) = 0 \tag{9}$$

for all i. By summing the left hand side over i, it is clear that if $E_1, \ldots, E_n$ is an optimal solution to the primal problem; then an optimal solution of the dual problem is given by $\Sigma_i p_i|\psi_i\rangle\langle\psi_i|E_i$.

To verify that a POVM is optimal, the necessary and sufficient conditions are therefore the following:

(i) $\Sigma_i p_i|\psi_i\rangle\langle\psi_i|E_i$ is Hermitian,

(ii) $\Sigma_i p_i|\psi_i\rangle\langle\psi_i|E_i \geqslant p_j|\psi_j\rangle\langle\psi_j|$ for all j.

The above equations guarantee that $Z=\Sigma_i p_i|\psi_i\rangle\langle\psi_i|E_i$ is positive semidefinite and dual feasible. Optimality is then guaranteed because the primal and dual SDP solutions have matching values.

II. LINEARLY INDEPENDENT STATES

We begin by analyzing the problem of minimum-error pure-state discrimination for a set of linearly independent states. This case is particularly simple as the optimal measurement is unique and consists of a von Neumann measurement involving n one-dimensional projectors.

We begin by introducing some notation which will be used in the rest of this section and, where appropriate, in the case of linearly dependent states. We shall then review some of the properties of the optimal measurement and the PGM for linearly independent states, finally proving the main result in Sec. II D. An alternate derivation of the same result is given in Appendix A.

A. Notation

Let $|\psi_1\rangle, \ldots, |\psi_n\rangle$ be a set of states that are chosen by Alice with probabilities $p_1, \ldots, p_n$, respectively. We assume $p_i > 0$ for all i and that the states are linearly independent and span the Hilbert space $\mathcal{H}$ that contains them.

Sampling from the above distribution produces the density operator

$$\rho = \sum_i p_i|\psi_i\rangle\langle\psi_i|. \tag{10}$$

Associated with the above problem is also a (scaled) Gram matrix defined by

$$G_{i,j} = \langle i|G|j\rangle = \sqrt{p_i p_j}\langle\psi_i|\psi_j\rangle. \tag{11}$$

The Gram matrix should be thought of, not as an operator on $\mathcal{H}$, but rather on a different n-dimensional vector space. The latter space has a natural basis that is invariant under basis changes in $\mathcal{H}$.

The two spaces are related by the linear map

$$M = \sum_i \sqrt{p_i}|\psi_i\rangle\langle i|, \quad \text{i.e.,} \quad M|i\rangle = \sqrt{p_i}|\psi_i\rangle. \tag{12}$$

And both ρ and G can be written in terms of this map,

$$\rho = MM^\dagger, \quad G = M^\dagger M. \tag{13}$$

Although M is non-Hermitian, we can use the singular value decomposition to write

$$M = UDV^{-1}, \tag{14}$$

where U and V are unitary operators, and D is diagonal with non-negative entries. In fact, because the vectors were required to be linearly independent, G is a positive definite operator and so is D.

With the above notation, we can write

$$\rho = UD^2U^{-1}, \quad G = VD^2V^{-1}, \tag{15}$$

$$\sqrt{\rho} = UDU^{-1}, \quad \sqrt{G} = VDV^{-1}. \tag{16}$$

At this point we are implicitly using the same Hilbert space for both ρ and G, however this should not lead to any confusion below.

Similar notation can also be introduced when the states are not linearly independent, but in that case M is no longer a square matrix and G is not invertible. Furthermore, some of the occurrences of D have to be padded by zeros in order to have the right matrix size.

Returning to the case of linearly independent states, we will be interested in the equivalent matrices for the state discrimination problem with the same states but (nonzero) probabilities $\tilde{p}_1, \ldots, \tilde{p}_n$, to be fixed later. We define $\tilde{\rho}$, $\tilde{G}$, $\tilde{M}$, $\tilde{U}$, $\tilde{D}$, and $\tilde{V}$ as above so that

$$\tilde{M} = \tilde{U}\tilde{D}\tilde{V}^{-1} = \sum_i \sqrt{\tilde{p}_i}|\psi_i\rangle\langle i|, \tag{17}$$

$$\tilde{\rho} = \tilde{M}\tilde{M}^\dagger = \tilde{U}\tilde{D}^2\tilde{U}^{-1} = \sum_i \tilde{p}_i|\psi_i\rangle\langle\psi_i|, \tag{18}$$

$$\tilde{G} = \tilde{M}^\dagger\tilde{M} = \tilde{V}\tilde{D}^2\tilde{V}^{-1} = \sum_{i,j} (\sqrt{\tilde{p}_i\tilde{p}_j}\langle\psi_i|\psi_j\rangle)|i\rangle\langle j|. \tag{19}$$

Finally, we introduce the diagonal probability matrices

$$P = \sum_i p_i|i\rangle\langle i|, \quad \tilde{P} = \sum_i \tilde{p}_i|i\rangle\langle i|, \tag{20}$$

which will help us relate the pairs of operators. For instance,

$$M = \tilde{M}\sqrt{P\tilde{P}^{-1}}, \quad G = \sqrt{P\tilde{P}^{-1}}\tilde{G}\sqrt{P\tilde{P}^{-1}}. \tag{21}$$

Note, however, that in general cases $\sqrt{G}$ has no simple linear expression in terms of $\sqrt{\tilde{G}}$.

B. Uniqueness of the map

In this section, we will review some of the properties of the map that takes a discrimination problem to its corresponding optimal measurement. In particular, we prove that

if we fix the states to be discriminated (and assume they are irreducible, a property to be defined below), then there is a one-to-one map between the set of *a priori* probabilities and the set of optimal measurements.

Lemma 1. When the states $\{|\psi_i\rangle\}$ are linearly independent and the *a priori* probabilities are nonzero, then the optimal measurement for minimum-error state discrimination satisfies

(a) $\langle\psi_i|E_i|\psi_i\rangle \neq 0$ for all i,

(b) $E_i^2 = E_i$ and $\mathrm{Tr}\, E_i = 1$ for all i,

(c) the optimal measurement is unique.

Proof. Let $|\psi_1^d\rangle, \ldots, |\psi_n^d\rangle$ be the (unnormalized) vectors dual to $|\psi_1\rangle, \ldots, |\psi_n\rangle$ so that

$$\langle\psi_j^d|\psi_i\rangle = \delta_{i,j}. \tag{22}$$

Their existence is guaranteed by the linear independence of the original vectors.

We begin by proving (a). Assume that k is such that $\langle\psi_k|E_k|\psi_k\rangle = 0$. Define Π to project onto the space spanned by the $n-1$ vectors $\{|\psi_i\rangle\}$ excluding $|\psi_k\rangle$. Then $I-\Pi$ is the projectors onto the orthogonal space (i.e., $I-\Pi = |\psi_k^d\rangle\langle\psi_k^d|/\langle\psi_k^d|\psi_k^d\rangle$). We introduce new elements as follows:

$$E_i' = \Pi E_i \Pi + \delta_{i,k}(I-\Pi). \tag{23}$$

These clearly form a POVM, and $\langle\psi_i|E_i'|\psi_i\rangle = \langle\psi_i|E_i|\psi_i\rangle$ for $i \neq k$, whereas $\langle\psi_k|E_k'|\psi_k\rangle \geqslant \langle\psi_k|(I-\Pi)|\psi_k\rangle > 0$ and therefore the original POVM was not optimal.

Part (b) is simply the result of Kennedy [14] that proves that all optimal POVMs for linearly independent pure states must consist of projectors. Note that often the result is cited by saying that there exists an optimal POVM consisting of projectors, but the preceding stronger statement (i.e., that only projectors can be optimal) follows from his proof too, which for completeness we summarize below:

The mutual slackness condition acting on $|\psi_k^d\rangle$,

$$E_j\left(\sum_i p_i E_i|\psi_i\rangle\langle\psi_i| - p_j|\psi_j\rangle\langle\psi_j|\right)|\psi_k^d\rangle = 0, \tag{24}$$

leads to

$$E_j(p_k E_k|\psi_k\rangle) = \delta_{j,k}(p_k E_k|\psi_k\rangle). \tag{25}$$

From (a) we know that the vectors $E_k|\psi_k\rangle$ are all nonzero. Given that, the above equation implies that the vectors are orthogonal (i.e., for $j \neq k$ $E_k|\psi_k\rangle \neq 0$, $E_j|\psi_j\rangle \neq 0$ are eigenvectors of the Hermitian operator E_k with different eigenvalues) and hence the POVM consists of one-dimensional projectors.

To prove (c) we assume that there exists two optimal POVMs: $E_1, \ldots, E_n$ and $E_1', \ldots, E_n'$. Since the optimality constraints are all linear, their convex combination must be an optimal measurement as well. Unless both POVM's are identical, their convex combination will include at least one element that is not a projector. By part (b), all optimal POVMs contain only projectors. ■

Definition 2. We say that $|\psi_1\rangle, \ldots, |\psi_n\rangle$ are reducible if there exists a pair of projectors $\{\Pi, I-\Pi\}$ such that for every i either $\Pi|\psi_i\rangle = |\psi_i\rangle$ or $(I-\Pi)|\psi_i\rangle = |\psi_i\rangle$ holds.

If a set of states is reducible, then the optimal distinguishing measurement can be implemented by first applying the POVM $\{\Pi, I-\Pi\}$ and then solving the remaining problems on each subspace. Therefore, we shall be mainly interested in irreducible cases.

Lemma 3. Given a set of irreducible linearly independent states $|\psi_1\rangle, \ldots, |\psi_n\rangle$ and a POVM $E_1, \ldots, E_n$, there is at most one set of *a priori* nonzero probabilities $p_1, \ldots, p_n$ such that the measurement is optimal for the minimum-error state discrimination problem.

Proof. For the POVM to be optimal, it must consist of one-dimensional projectors, so we choose a basis such that $E_i = |i\rangle\langle i|$. In this basis, the expression $Z = \Sigma_k p_k|\psi_k\rangle\langle\psi_k|E_k$ has components $Z_{i,j} \equiv \langle i|Z|j\rangle = p_j\langle i|\psi_j\rangle\langle\psi_j|j\rangle$. The requirement that Z be Hermitian implies that

$$p_j\langle i|\psi_j\rangle\langle\psi_j|j\rangle = (p_i\langle j|\psi_i\rangle\langle\psi_i|i\rangle)^* \tag{26}$$

for all i and j. Recall that the optimal solution satisfies $\langle\psi_i|E_i|\psi_i\rangle = |\langle i|\psi_i\rangle|^2 \neq 0$ and that we are looking for solutions with nonzero $\{p_i\}$. Therefore, either the above equation determines the ratio p_i/p_j or we have $\langle i|\psi_j\rangle = \langle j|\psi_i\rangle = 0$.

Consider the n node graph, where two vertices i, j are connected if $\langle i|\psi_j\rangle \neq 0$ (which implies $\langle j|\psi_i\rangle \neq 0$ if the POVM is optimal for a set of nonzero *a priori* probabilities). If there exists a path from a node i to a (not necessarily adjacent) node j, then the above equation fixes the ratio of p_i/p_j. If the graph is connected, then all ratios are fixed, and the nonzero probabilities $p_1, \ldots, p_n$ must be unique.

If the graph is not connected, let $\mathcal{C}$ be the vertices of one component. Defining $\Pi = \Sigma_{i\in\mathcal{C}}|i\rangle\langle i|$, the projectors $\{\Pi, I-\Pi\}$ would prove that that the states $\{|\psi_i\rangle\}$ are reducible contrary to our assumptions. ■

Note that in the proof of the above lemma we offer a different map from optimal measurements for linearly independent states into the set of *a priori* probabilities for which they are optimal. However, in this case the space of possible measurements includes the $O(n^2)$-dimensional space of linearly independent bases, out of which only an $O(n)$-dimensional subspace is optimal. The formalism of PGMs presented below has the advantage that it immediately identifies this optimal subspace.

C. The generalized PGM

Given a set of states $|\psi_1\rangle, \ldots, |\psi_n\rangle$ that span their Hilbert space and positive numbers $\tilde{p}_1, \ldots, \tilde{p}_n$, define the operators

$$\tilde{E}_i = \tilde{p}_i\sqrt{\tilde{\rho}^{-1}}|\psi_i\rangle\langle\psi_i|\sqrt{\tilde{\rho}^{-1}}. \tag{27}$$

As we shall mostly be interested in a fixed set of states while varying the numbers $\tilde{p}_i$, we shall refer to the above as the (generalized) PGM associated with $\{\tilde{p}_i\}$.

We review below a few of the properties of the (generalized) PGM.

Lemma 4. The operators $\tilde{E}_1, \ldots, \tilde{E}_n$ form a POVM. Furthermore, when the states $|\psi_1\rangle, \ldots, |\psi_n\rangle$ are linearly independent, the measurement consist of n orthogonal projectors and can be simulated by the application of the unitary $\tilde{V}\tilde{U}^{-1}$ followed by a measurement in the computational basis.

Proof. Because the states span the Hilbert space and the numbers $\tilde{p}_i$ are positive, the operator $\tilde{\rho}$ is positive definite.

Hence $\tilde{\rho}$ is invertible and the operators $\tilde{E}_i$ are well defined. Furthermore, they are positive semidefinite as they are proportional to projectors onto the unnormalized states $\sqrt{\tilde{\rho}^{-1}}|\psi_i\rangle$.

To prove that $\{\tilde{E}_i\}$ form a POVM, we just need to verify that they sum to the identity

$$\sum_i \tilde{E}_i = \sqrt{\tilde{\rho}^{-1}}\tilde{\rho}\sqrt{\tilde{\rho}^{-1}} = I. \tag{28}$$

Note that the above is true, even if the numbers $\{\tilde{p}_i\}$ do not sum to 1.

Focusing on the case of linearly independent states, we have

$$\sqrt{\tilde{p}_i}\sqrt{\tilde{\rho}^{-1}}|\psi_i\rangle = \sqrt{\tilde{\rho}^{-1}}\tilde{M}|i\rangle = (\tilde{U}\tilde{D}^{-1}\tilde{U}^{-1})(\tilde{U}\tilde{D}\tilde{V}^{-1})|i\rangle = \tilde{U}\tilde{V}^{-1}|i\rangle \tag{29}$$

and hence $\tilde{E}_i = \tilde{U}\tilde{V}^{-1}|i\rangle\langle i|\tilde{V}\tilde{U}^{-1}$, proving the remainder of the lemma. ■

D. Main result

Theorem 5. The PGM associated with a set of positive numbers $\tilde{p}_1,\ldots,\tilde{p}_n$ and states $|\psi_1\rangle,\ldots,|\psi_n\rangle$ is the optimal measurement for the minimum-error state discrimination problem with the same states and *a priori* probabilities

$$p_i = \frac{C}{\langle\psi_i|\sqrt{\tilde{\rho}^{-1}}|\psi_i\rangle} = \frac{C\tilde{p}_i}{\langle i|\sqrt{\tilde{G}}|i\rangle}, \tag{30}$$

where C is the normalization constant needed to make the *a priori* probabilities sum to 1.

Proof. Let

$$Z = \sum_i p_i|\psi_i\rangle\langle\psi_i|E_i = \sum_i p_i\tilde{p}_i|\psi_i\rangle\langle\psi_i|\sqrt{\tilde{\rho}^{-1}}|\psi_i\rangle\langle\psi_i|\sqrt{\tilde{\rho}^{-1}}. \tag{31}$$

We need to verify that Z is Hermitian and satisfies the constraints $Z \geqslant p_i|\psi_i\rangle\langle\psi_i|$ for all i.

Using

$$p_i = \frac{C}{\langle\psi_i|\sqrt{\tilde{\rho}^{-1}}|\psi_i\rangle}, \tag{32}$$

which is well defined because $\tilde{\rho}$ is positive definite, we obtain

$$Z = C\left(\sum_i \tilde{p}_i|\psi_i\rangle\langle\psi_i|\right)\sqrt{\tilde{\rho}^{-1}} = C\sqrt{\tilde{\rho}}, \tag{33}$$

which is clearly Hermitian.

All that remains is to check the condition $Z \geqslant p_i|\psi_i\rangle\langle\psi_i|$ for $i=1,\ldots,n$. We use the fact that if Z is positive definite, then

$$Z \geqslant p_i|\psi_i\rangle\langle\psi_i| \Leftrightarrow I \geqslant p_i\sqrt{Z^{-1}}|\psi_i\rangle\langle\psi_i|\sqrt{Z^{-1}} \Leftrightarrow 1 \geqslant p_i\langle\psi_i|Z^{-1}|\psi_i\rangle, \tag{34}$$

where in the last step we used the fact that the matrix $\sqrt{Z^{-1}}|\psi_i\rangle\langle\psi_i|\sqrt{Z^{-1}}$ has only one nonzero eigenvalue. Using our expression for Z, we find

$$p_i\langle\psi_i|Z^{-1}|\psi_i\rangle = p_iC^{-1}\langle\psi_i|\sqrt{\tilde{\rho}^{-1}}|\psi_i\rangle = 1 \tag{35}$$

concluding the proof that the PGM associated with $\{\tilde{p}_i\}$ is optimal.

Finally, note that

$$\tilde{p}_i\langle\psi_i|\sqrt{\tilde{\rho}^{-1}}|\psi_i\rangle = \langle i|\tilde{M}^\dagger\sqrt{\tilde{\rho}^{-1}}\tilde{M}|i\rangle = \langle i|\sqrt{\tilde{G}}|i\rangle, \tag{36}$$

which proves that $\{\tilde{p}_i\}$ and $\{p_i\}$ are also related by the expression $p_i = C\tilde{p}_i/\langle i|\sqrt{\tilde{G}}|i\rangle$. ■

A quick corollary of the above theorem is that the regular PGM (i.e., when $\tilde{p}_i = p_i$ for all i) is optimal when $\langle i|\sqrt{\tilde{G}}|i\rangle$ is a constant for all i, reproducing the result obtained by Sasaki *et al.* [11].

We have so far shown that for every set of positive numbers $\{\tilde{p}_i\}$ defining a measurement, there is a set of *a priori* probabilities $\{p_i\}$ for which they are optimal. We now intend to show that the converse is true as well.

Theorem 6. For every set of linearly independent pure states $|\psi_1\rangle,\ldots,|\psi_n\rangle$ and nonzero *a priori* probabilities $p_1,\ldots,p_n$, the optimal measurement for the minimum-error pure-state discrimination problem is the PGM associated with $\{\tilde{p}_i\}$ for some positive numbers $\tilde{p}_1,\ldots,\tilde{p}_n$. Furthermore, the mapping is one-to-one if we require the numbers $\{\tilde{p}_i\}$ to sum to 1, and the states to be irreducible.

Proof. Let $\mathbb{R}_+^n$ be the subset of $\mathbb{R}^n$ where all coordinates are positive. Define the map $f:\mathbb{R}_+^n \to \mathbb{R}_+^n$ by

$$f_j(\tilde{p}_1,\ldots,\tilde{p}_n) = \frac{1}{\langle\psi_j|\sqrt{\tilde{\rho}^{-1}}|\psi_j\rangle}, \tag{37}$$

where $f_j(\{\tilde{p}_i\})$ is the jth component of the n-dimensional vector $f(\{\tilde{p}_i\})$, and as usual $\tilde{\rho} = \Sigma_i\tilde{p}_i|\psi_i\rangle\langle\psi_i|$. It is a well defined function because the states are linearly independent and the variables $\{\tilde{p}_i\}$ are positive, guaranteeing that $\tilde{\rho}$ is positive definite.

To prove the first part of the theorem (i.e., the existence of $\{\tilde{p}_i\}$ given $\{p_i\}$), it is sufficient to show that the range of f is the whole set $\mathbb{R}_+^n$. Because f is continuous, its range is a connected region. The boundary of the range of f is the set of points such that every neighborhood containing one of these points contains both points in the range of f and points outside the range of f. We intend to show that all boundary points lie outside $\mathbb{R}_+^n$, which will prove the first part of the theorem.

There are two kinds of boundary points, those that are in the range of f and those that are outside its range. The latter points arise as the image of sequences such that either $\tilde{p}_i \to 0$ or $\tilde{p}_i \to \infty$ for at least one i. We will show that the image of such sequences has either $p_j \to \infty$ or $p_j \to 0$ for at least one p_j.

If $\tilde{p}_i \to \infty$, then consider

$$\begin{aligned} n = \mathrm{Tr}[I] = \mathrm{Tr}[\sqrt{\tilde{\rho}^{-1}}\tilde{\rho}\sqrt{\tilde{\rho}^{-1}}] &\geqslant \langle\psi_i|\sqrt{\tilde{\rho}^{-1}}\tilde{\rho}\sqrt{\tilde{\rho}^{-1}}|\psi_i\rangle \\ &\geqslant \tilde{p}_i(\langle\psi_i|\sqrt{\tilde{\rho}^{-1}}|\psi_i\rangle)^2, \end{aligned} \tag{38}$$

where in the last step we used $\tilde{\rho} \geqslant \tilde{p}_i|\psi_i\rangle\langle\psi_i|$. The above expression shows that as $\tilde{p}_i$ goes to infinity, $\langle\psi_i|\sqrt{\tilde{\rho}^{-1}}|\psi_i\rangle$ goes to zero and therefore p_i goes to infinity.

If $\tilde{p}_i \to 0$, then let $|\psi_i^d\rangle$ be a vector such that $\langle\psi_i^d|\psi_j\rangle = 0$ for all j not equal to i, which exists because of the linear independence of the original states. As $\tilde{p}_i \to 0$, then $\langle\psi_i^d|\tilde{\rho}|\psi_i^d\rangle$ goes to zero as well, and since $\langle\psi_i^d|\psi_i^d\rangle\langle\psi_i^d|\tilde{\rho}|\psi_i^d\rangle \geqslant \langle\psi_i^d|\sqrt{\tilde{\rho}}|\psi_i^d\rangle^2$, the positive matrix $\sqrt{\tilde{\rho}}$ must have an eigenvalue that becomes arbitrarily small. Therefore, $\sqrt{\tilde{\rho}^{-1}}$ must have an eigenvalue that becomes arbitrarily large and at least one of the p_i must become arbitrarily small.

The only possibility that remains is that a point in the range of f is a boundary point. We will exclude this possibility by proving that f is locally one-to-one (i.e., that the partial derivative matrix $\partial f_i/\partial \tilde{p}_j$ is positive definite).

For $\epsilon > 0$, R positive definite, and X a Hermitian matrix, we have

$$(R^2 + \epsilon X)^{-1/2} = R^{-1} + \epsilon Y + O(\epsilon^2) \tag{39}$$

with Y defined by

$$Y_{a,b} = -\frac{X_{a,b}}{r_a r_b (r_a + r_b)}, \tag{40}$$

where the subscripts indicate the components of X and Y in the eigenbasis of R, and $\{r_a\}$ denote the corresponding eigenvalues. The above can easily be verifying by computing

$$(R^2 + \epsilon X)[R^{-1} + \epsilon Y + O(\epsilon^2)]^2 = I + \epsilon R(RY + YR + R^{-1}XR^{-1})R^{-1} + O(\epsilon^2) \tag{41}$$

and verifying that the expression in parentheses vanishes.

Using $R = \sqrt{\tilde{\rho}}$ and $X = |\psi_j\rangle\langle\psi_j|$, we find that

$$\frac{\partial f_i}{\partial \tilde{p}_j} = \sum_{a,b} \frac{(|\psi_i\rangle\langle\psi_i|)_{a,b}(|\psi_j\rangle\langle\psi_j|)_{a,b}}{\sqrt{\tilde{\rho}}_a \sqrt{\tilde{\rho}}_b (\sqrt{\tilde{\rho}}_a + \sqrt{\tilde{\rho}}_b)}, \tag{42}$$

where the matrices are expressed in the eigenbasis of $\tilde{\rho}$. The partial derivative matrix is positive definite because for any vector $(\alpha_1, \ldots, \alpha_n)$ we have

$$\sum_{i,j} \alpha_i \frac{\partial f_i}{\partial \tilde{p}_j} \alpha_j = \sum_{a,b} \frac{\left(\sum_i \alpha_i |\psi_i\rangle\langle\psi_i|\right)^2_{a,b}}{\sqrt{\tilde{\rho}}_a \sqrt{\tilde{\rho}}_b (\sqrt{\tilde{\rho}}_a + \sqrt{\tilde{\rho}}_b)} > 0. \tag{43}$$

To finish proving the theorem, we note that if all the $\tilde{p}_i$ are scaled by a constant a, then f gets scaled by $\sqrt{a}$. Therefore, we can require that the $\{\tilde{p}_i\}$ sum to 1 if we reintroduce the normalization constant C.

Furthermore, to prove that the mapping between normalized $\{p_i\}$ and normalized $\{p_i\}$ is one-to-one, it is sufficient to show that the function f is globally one-to-one (i.e., if two different normalized sets of $\{\tilde{p}_i\}$ mapped to the same normalized $\{p_i\}$, we can find two unnormalized $\{\tilde{p}_i\}$ that map to the same point).

Assume that two distinct points map to the same point under f. Then every path connecting the two points must map to a closed loop under f. Because f is locally one-to-one, we can deform the loop and make it arbitrarily small. This would imply that there is a path between the two original points such that every point along the path maps under f to the same point. This contradicts the fact that f is locally one-to-one. ■

III. LINEARLY DEPENDENT STATES

Linearly dependent states introduce new complexities into the problem. Consider the four states $|0\rangle$, $|1\rangle$, and $|\pm\rangle = (|0\rangle \pm |1\rangle)/\sqrt{2}$ each occurring with equal probability. This problem falls under the case of states that sum to the identity [1]. It follows that both

$$E_0 = |0\rangle\langle 0|, \quad E_1 = |1\rangle\langle 1|, \quad E_+ = 0, \quad E_- = 0 \tag{44}$$

and

$$E_0' = 0, \quad E_1' = 0, \quad E_+' = |+\rangle\langle +|, \quad E_-' = |-\rangle\langle -| \tag{45}$$

are optimal measurements. Linear combinations of the above are also optimal. This example illustrates two important facts about optimal measurements for linearly dependent states: the optimal measurement is in general not unique and the POVM often includes zero elements.

Furthermore, the first of the above measurements is optimal for the same states with *a priori* probabilities a, a, b, and b, respectively (with $2a+2b=1$), so long as $a \geqslant b$. That is an example where the same measurement is optimal for many different *a priori* probabilities.

The generalized PGM for linearly dependent states is defined as in Sec. II C, and is well defined for nonzero $\{\tilde{p}_i\}$ as long as $\tilde{\rho}$ is positive definite (and otherwise the POVM can be completed by adding an extra projector, though this will not be used below). Of course, in general, the PGM no longer consists of projectors (i.e., $E_i^2 \neq E_i$). Furthermore, different assignments for the variables $\{\tilde{p}_i\}$ can lead to the same measurement.

The following theorem shows us how we can associate to a generalized PGM a state discrimination problem for which it is optimal. In fact, when some of the $\{\tilde{p}_i\}$ are zero, there are many different problems for which the associated PGM is optimal.

Theorem 7. Given a set of pure states $|\psi_1\rangle, \ldots, |\psi_n\rangle$ and non-negative numbers $\tilde{p}_1, \ldots, \tilde{p}_n$ such that $\tilde{\rho} = \Sigma_i \tilde{p}_i |\psi_i\rangle\langle\psi_i|$ is a positive definite operator, the associated PGM is optimal for any minimum-error pure-state discrimination problem with the same states and *a priori* probabilities satisfying

$$p_i \leqslant \frac{C}{\langle\psi_i|\sqrt{\tilde{\rho}^{-1}}|\psi_i\rangle} \quad \text{for all } i \text{ such that } \tilde{p}_i = 0, \tag{46}$$

$$p_i = \frac{C}{\langle\psi_i|\sqrt{\tilde{\rho}^{-1}}|\psi_i\rangle} \quad \text{for all } i \text{ such that } \tilde{p}_i \neq 0 \tag{47}$$

for some normalization constant C.

Proof. The proof follows the same lines as Theorem 5. The operator Z defined by

$$Z = \sum_i p_i |\psi_i\rangle\langle\psi_i| E_i = \sum_i p_i \tilde{p}_i |\psi_i\rangle\langle\psi_i|\sqrt{\tilde{\rho}^{-1}}|\psi_i\rangle\langle\psi_i|\sqrt{\tilde{\rho}^{-1}} = C\left(\sum_i \tilde{p}_i |\psi_i\rangle\langle\psi_i|\right)\sqrt{\tilde{\rho}^{-1}} = C\sqrt{\tilde{\rho}} \tag{48}$$

has the same form as before because only the terms with $\tilde{p}_i \neq 0$ contribute. It is clearly positive definite and what remains to be checked is that $Z \geqslant p_i |\psi_i\rangle\langle\psi_i|$ for all i, which as

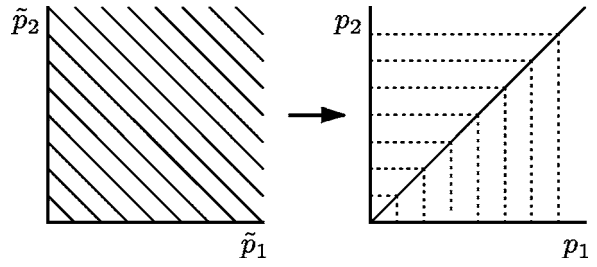

FIG. 1. The mapping from $(\tilde{p}_1,\tilde{p}_2)$ to (p_1,p_2), given by Eqs. (46) and (47) with C fixed to 1, for two identical states.

before is equivalent to verifying that $1 \geqslant p_i\langle\psi_i|Z^{-1}|\psi_i\rangle = p_i C^{-1}\langle\psi_i|\sqrt{\tilde{\rho}^{-1}}|\psi_i\rangle$ for all i, which follows directly from the definition of p_i. □

We conclude by exemplifying the above theorem with the simplest possible case of state discrimination: two identical states. For $|\psi_1\rangle=|\psi_2\rangle$ (contained in a one-dimensional Hilbert space), the PGM associated to $\tilde{p}_1$, $\tilde{p}_2$ is $E_i=\tilde{p}_i I/(\tilde{p}_1+\tilde{p}_2)$. The map given by Eqs. (46) and (47) works as follows: setting $C=1$ and requiring equality to hold, the function maps the $(\tilde{p}_1,\tilde{p}_2)$ plane to the line $p_1=p_2=\tilde{p}_1+\tilde{p}_2$. The rest of the (p_1,p_2) plane is filled up by the inequality cases (i.e., when either $\tilde{p}_1=0$ or $\tilde{p}_2=0$), as depicted in Fig. 1. Of course, once we use C to normalize the *a priori* probabilities, the second plane is collapsed to the line $p_1+p_2=1$. In the end we see that if $p_1=p_2$, then the full range of measurements is allowed, but once one of them is larger, the optimal POVM is unique, which should be intuitively expected.

ACKNOWLEDGMENTS

The author would like to thank Andrew Childs for his help in proofreading this manuscript, and Jon Tyson for his help with some of the bibliography. This work was supported in part by the National Science Foundation under Grant No. EIA-0086038 and by the U.S. Department of Energy under Grant No. DE-FG03-92-ER40701.

APPENDIX A: ALTERNATE DERIVATION

In this appendix, we shall give an alternate derivation of the relationship between the *a priori* probabilities $\{p_i\}$ and the numbers $\{\tilde{p}_i\}$ used to define the optimal PGM for the case of linearly independent states. The proof will not directly use the optimality conditions and provides some motivation to the construction. However, given that the main result has already been formally proven, the following argument will be fairly informal.

We can describe the problem of state discrimination as follows: Alice prepares the mixed state

$$\sum_i p_i|i\rangle\langle i|_{\mathcal{A}} \otimes |\psi_i\rangle\langle\psi_i|_{\mathcal{B}} \tag{A1}$$

and sends $\mathcal{B}$ to Bob, who is then allowed to apply any quantum operation to it (together with any private qubits he may hold) and then returns his half to Alice, who measures with the projector

$$\Pi = \sum_i |i\rangle\langle i|_{\mathcal{A}} \otimes |i\rangle\langle i|_{\mathcal{B}}. \tag{A2}$$

Outcome Π implies Bob guesses the state correctly and outcome $I-\Pi$ indicates a failure. Note that because that states $\{|\psi_i\rangle\}$ are linearly independent and span the space, $\mathcal{B}$ is n-dimensional and the same space can be used for Bob to tell Alice his guess.

It is not hard to show that because Π is block diagonal, Alice can start by preparing the pure state

$$|\Phi\rangle = (I_{\mathcal{A}} \otimes M_{\mathcal{B}})\sum_i |i\rangle_{\mathcal{A}} \otimes |i\rangle_{\mathcal{B}} = \sum_i \sqrt{p_i}|i\rangle_{\mathcal{A}} \otimes |\psi_i\rangle_{\mathcal{B}} \tag{A3}$$

instead of the mixed state from Eq. (A1) without changing Bob's probability of success. The above state has the nice property that $\mathrm{Tr}_{\mathcal{B}}|\Phi\rangle\langle\Phi|=G$ and $\mathrm{Tr}_{\mathcal{A}}|\Phi\rangle\langle\Phi|=\rho$.

The problem can now be written as the SDP,

$$\text{Maximize: } \mathrm{Tr}[\Pi\sigma], \tag{A4}$$

$$\text{Subject to: } \mathrm{Tr}_{\mathcal{B}}\sigma = \mathrm{Tr}_{\mathcal{B}}|\Phi\rangle\langle\Phi| = G, \tag{A5}$$

$$\sigma \geqslant 0, \tag{A6}$$

where σ is a density operator on $\mathcal{A}\otimes\mathcal{B}$. The first constraint simply says that Bob can perform any quantum operation on $\mathcal{B}$ but that he must leave $\mathcal{A}$ intact.

For the linearly independent case, we know that the optimal strategy is simply to apply a unitary, say W, to the space $\mathcal{B}$ so that $\sigma=(I_{\mathcal{A}}\otimes W_{\mathcal{B}})|\Phi\rangle\langle\Phi|(I_{\mathcal{A}}\otimes W_{\mathcal{B}}^{\dagger})$. In terms of this optimal unitary, define the diagonal matrix Γ with entries given by

$$\Gamma_i = \langle i|_{\mathcal{A}} \otimes \langle i|_{\mathcal{B}}(I_{\mathcal{A}} \otimes W_{\mathcal{B}})|\Phi\rangle. \tag{A7}$$

We assume that Γ has real non-negative entries (otherwise W could be modified to correct for the extra phase without changing the success probability). This matrix encodes the success probability because $p_{\mathrm{succ}}=\mathrm{Tr}[\Pi\sigma]=\mathrm{Tr}[\Gamma^2]$.

However, we can now modify the Alice-Bob game so that Alice replaces the projector Π with the projector onto the normalized state

$$|\Upsilon\rangle = \frac{1}{\sqrt{\mathrm{Tr}[\Gamma^2]}}(I_{\mathcal{A}} \otimes \Gamma_{\mathcal{B}})\sum_i |i\rangle_{\mathcal{A}} \otimes |i\rangle_{\mathcal{B}}. \tag{A8}$$

Clearly Bob can achieve the same success rate as before, and yet because $\Pi \geqslant |\Upsilon\rangle\langle\Upsilon|$ he cannot increase his success probability in this modified game.

The game now has the following form: Alice prepares the bipartite state $|\Phi\rangle$, and Bob must apply a unitary (or nonunitary superoperator, though we know these are not optimal) to maximize the state's overlap with $|\Upsilon\rangle$. The probability of success is the square of the fidelity between $\mathrm{Tr}_{\mathcal{B}}|\Phi\rangle\langle\Phi|=G$ and $\mathrm{Tr}_{\mathcal{B}}|\Upsilon\rangle\langle\Upsilon|=\Gamma^2/\mathrm{Tr}[\Gamma^2]$.

The above description may seem strange as it is written in terms of Γ, which was defined in terms of the optimal unitary which remains unknown. However, it leads to some interesting descriptions, for instance

$$p_{\text{succ}} = \max_{\sigma \in \text{diag}} F^2(G,\sigma), \tag{A9}$$

where the maximum is taken over density operators that are diagonal in the basis that was used to define the Gram matrix. The above is true because varying over σ is equivalent to varying the final projection state over $(I_{\mathcal{A}} \otimes \sqrt{\sigma_{\mathcal{B}}})\Sigma_i |i\rangle_{\mathcal{A}} \otimes |i\rangle_{\mathcal{B}}$. All these states are in the image of Π so they give Bob no extra advantage.

Alternatively, fix Γ and $|\Upsilon\rangle$ as above, but allow Bob's unitary, W, to vary. We have

$$p_{\text{succ}} = |\langle \Upsilon | (I_{\mathcal{A}} \otimes W_{\mathcal{B}}) | \Phi \rangle|^2 = \frac{1}{\text{Tr}[\Gamma^2]} (\text{Tr}[\Gamma W M])^2$$

$$= \frac{1}{\text{Tr}[\Gamma^2]} (\text{Tr}[W M \Gamma])^2. \tag{A10}$$

It is clear that W should be chosen so that $WM\Gamma^\dagger$ is positive semidefinite. This is generally difficult. However, we can introduce an alternate set of probabilities $\tilde{p}_1, \ldots, \tilde{p}_n$ and use $M = \tilde{M}\sqrt{P\tilde{P}^{-1}}$ [from Eq. (21)] and then choose the alternate probabilities so that $\tilde{P} = P\Gamma^2$ (equivalently $\tilde{p}_i = p_i \Gamma_i^2$). We now need to choose W in order to make $W\tilde{M}$ positive semidefinite, but this is clearly given by $W = \tilde{V}\tilde{U}^{-1}$, which is the unitary that is used for the PGM associated with $\tilde{p}_1, \ldots, \tilde{p}_n$.

Starting from Eq. (A7), and using our choice of $W = \tilde{V}\tilde{U}^{-1}$, we can find a second relationship between Γ_i, p_i, and $\tilde{p}_i$,

$$\Gamma_i = \langle i|_{\mathcal{A}} \otimes \langle i|_{\mathcal{B}} (I_{\mathcal{A}} \otimes W_{\mathcal{B}}) |\Phi\rangle = \langle i | WM | i \rangle = \langle i | W\tilde{M}\sqrt{P\tilde{P}^{-1}} | i \rangle$$

$$= \sqrt{p_i \tilde{p}_i^{-1}} \langle i | \sqrt{\tilde{G}} | i \rangle. \tag{A11}$$

Combined with $\tilde{p}_i = p_i \Gamma_i^2$, which we used to define the alternate probabilities, we can now eliminate Γ_i to find

$$p_i = \frac{\tilde{p}_i}{\langle i | \sqrt{\tilde{G}} | i \rangle}. \tag{A12}$$

The left hand side depends only on the *a priori* probabilities $\{p_i\}$ and the right hand side only on the positive numbers $\{\tilde{p}_i\}$ with are used to define the PGM that is optimal. Note that no constant of proportionality is needed if we use the above equation to fix the normalization of the $\{\tilde{p}_i\}$, though such a choice will in general not result in the $\{\tilde{p}_i\}$ summing to 1.

[1] H. P. Yuen, R. S. Kennedy, and M. Lax, IEEE Trans. Inf. Theory **IT-21**, 125 (1975).
[2] L. Ip (unpublished).
[3] D. Bacon, A. M. Childs, and W. van Dam, e-print quant-ph/0501044.
[4] P. Hausladen and W. K. Wootters, J. Mod. Opt. **41**, 2385 (1994).
[5] A. S. Holevo, J. Multivariate Anal. **3**, 337 (1973).
[6] C. W. Helstrom, *Quantum Detection and Estimation Theory* (Academic Press, New York, 1976), pp. 106–113.
[7] E. Andersson, S. M. Barnett, C. R. Gilson, and K. Hunter, Phys. Rev. A **65**, 052308(R) (2002).
[8] M. Ban, K. Kurokawa, R. Momose, and O. Hirota, Int. J. Theor. Phys. **36**, 1269 (1997).
[9] Y. C. Eldar, A. Megretski, and G. C. Verghese, IEEE Trans. Inf. Theory **50**, 1198 (2004).
[10] S. M. Barnett, Phys. Rev. A **64**, 030303(R) (2001).
[11] M. Sasaki, K. Kato, M. Izutsu, and O. Hirota, Phys. Rev. A **58**, 146 (1998).
[12] Y. C. Eldar and G. D. Forney, Jr., IEEE Trans. Inf. Theory **47**, 858 (2001).
[13] A. S. Kholevo, Theor. Probab. Appl. **23**, 414 (1978).
[14] R. S. Kennedy, Quarterly Progress Report, MIT Research Laboratory of Electronics Vol. 110 (1973), p. 142.
[15] Y. C. Eldar, Phys. Rev. A **68**, 052303 (2003).
[16] K. Hunter, Phys. Rev. A **68**, 012306(R) (2003).
[17] C. W. Helstrom, IEEE Trans. Inf. Theory **IT-28**, 359 (1982).
[18] J. A. Bergou, U. Herzog, and M. Hillery, in *Quantum State Estimation*, edited by M. Paris and J. Rehacek (Springer, Berlin, 2004), Vol. 3, pp. 417–465.
[19] Y. C. Eldar, A. Megretski, and G. C. Verghese, IEEE Trans. Inf. Theory **49**, 1007 (2003).