On Quantum Detection and the Square-Root Measurement

Yonina C. Eldar, *Student Member, IEEE,* and G. David Forney, Jr., *Fellow, IEEE*

***Abstract*—In this paper, we consider the problem of constructing measurements optimized to distinguish between a collection of possibly nonorthogonal quantum states. We consider a collection of pure states and seek a positive operator-valued measure (POVM) consisting of rank-one operators with measurement vectors closest in squared norm to the given states. We compare our results to previous measurements suggested by Peres and Wootters [11] and Hausladen *et al.* [10], where we refer to the latter as the square-root measurement (SRM). We obtain a new characterization of the SRM, and prove that it is optimal in a least-squares sense. In addition, we show that for a geometrically uniform state set the SRM minimizes the probability of a detection error. This generalizes a similar result of Ban *et al.* [7].**

***Index Terms*—Geometrically uniform quantum states, least-squares measurement, quantum detection, singular value decomposition, square-root measurement (SRM).**

I. Introduction

SUPPOSE that a transmitter, Alice, wants to convey classical information to a receiver, Bob, using a quantum-mechanical channel. Alice represents messages by preparing the quantum channel in a pure quantum state drawn from a collection of known states. Bob detects the information by subjecting the channel to a measurement in order to determine the state prepared. If the quantum states are mutually orthogonal, then Bob can perform an optimal orthogonal (von Neumann) measurement that will determine the state correctly with probability one [1]. The optimal measurement consists of projections onto the given states. However, if the given states are not orthogonal, then no measurement will allow Bob to distinguish perfectly between them. Bob's problem is therefore to construct a measurement optimized to distinguish between nonorthogonal pure quantum states.

We may formulate this problem as a quantum detection problem, and seek a measurement that minimizes the probability of a detection error, or more generally, minimizes the Bayes cost. Necessary and sufficient conditions for an optimum measurement minimizing the Bayes cost have been derived [2]–[4]. However, except in some particular cases [4]–[7], obtaining a closed-form analytical expression for the optimal measurement directly from these conditions is a difficult and unsolved problem. Thus, in practice, iterative procedures minimizing the Bayes cost [8] or *ad hoc* suboptimal measurements are used.

In this paper, we take an alternative approach of choosing a different optimality criterion, namely, a squared-error criterion, and seeking a measurement that minimizes this criterion. It turns out that the optimal measurement for this criterion is the "square-root measurement" (SRM), which has previously been proposed as a "pretty good" *ad hoc* measurement [9], [10].

This work was originally motivated by the problems studied by Peres and Wootters in [11] and by Hausladen *et al.* in [10]. Peres and Wootters [11] consider a source that emits three two-qubit states with equal probability. In order to distinguish between these states, they propose an orthogonal measurement consisting of projections onto measurement vectors "close" to the given states. Their choice of measurement results in a high probability of correctly determining the state emitted by the source, and a large mutual information between the state and the measurement outcome. However, they do not explain how they construct their measurement, and do not prove that it is optimal in any sense. Moreover, the measurement they propose is specific for the problem that they pose; they do not describe a general procedure for constructing an orthogonal measurement with measurement vectors close to given states. They also remark that improved probabilities might be obtained by considering a general positive operator-valued measure (POVM) [12] consisting of positive Hermitian operators Π_i satisfying $\sum_i \Pi_i = I$, where the operators Π_i are not required to be orthogonal projection operators as in an orthogonal measurement.

Hausladen *et al.* [10] consider the general problem of distinguishing between an arbitrary set of pure states, where the number of states is no larger than the dimension of the space $\mathcal{U}$ they span. They describe a procedure for constructing a general "decoding observable," corresponding to a POVM consisting of rank-one operators that distinguishes between the states "pretty well"; this measurement has subsequently been called the *square-root measurement (SRM)* (see e.g., [13]–[15]). However, they make no assertion of (nonasymptotic) optimality. Although they mention the problem studied by Peres and Wootters in [11], they make no connection between their measurement and the Peres–Wootters measurement.

The SRM [7], [9], [10], [13]–[15] has many desirable properties. Its construction is relatively simple; it can be determined directly from the given collection of states; it minimizes the proba-

Manuscript received May 30, 2000; revised September 12, 2000. This work was supported in part through collaborative participation in the Advanced Sensors Consortium sponsored by the U.S. Army Research Laboratory under Cooperative Agreement DAAL01-96-2-0001 and supported in part by the Texas Instruments Leadership University Program. Y. C. Eldar is currently supported by an IBM Research Fellowship.

Y. C. Eldar is with the Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, MA 02139 USA (e-mail: yonina@mit.edu).

G. D. Forney, Jr. is with the Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, Cambridge, MA 02139 USA (e-mail: forneyd@mediaone.net).

Communicated by P. W. Shor, Associate Editor for Quantum Information Theory.

Publisher Item Identifier S 0018-9448(01)01334-7.

bility of a detection error when the states exhibit certain symmetries [7]; it is "pretty good" when the states to be distinguished are equally likely and almost orthogonal [9]; and it is asymptotically optimal [10]. Because of these properties, the SRM has been employed as a detection measurement in many applications (see, e.g., [13]–[15]). However, apart from some particular cases mentioned above [7], no assertion of (nonasymptotic) optimality is known for the SRM.

In this paper, we systematically construct detection measurements optimized to distinguish between a collection of quantum states. Motivated by the example studied by Peres and Wootters [11], we consider pure-state ensembles and seek a POVM consisting of rank-one positive operators with measurement vectors that minimize the sum of the squared norms of the error vectors, where the ith error vector is defined as the difference between the ith state vector and the ith measurement vector. We refer to the optimizing measurement as the least-squares measurement (LSM). We then generalize this approach to allow for unequal weighting of the squared norms of the error vectors. This weighted criterion may be of interest when the given states have unequal prior probabilities. We refer to the resulting measurement as the weighted LSM (WLSM). We show that the SRM coincides with the LSM when the prior probabilities are equal, and with the WLSM otherwise (if the weights are proportional to the square roots of the prior probabilities).

We then consider the case in which the collection of states has a strong symmetry property called geometric uniformity [16]. We show that for such a state set the SRM minimizes the probability of a detection error. This generalizes a similar result of Ban *et al.* [7].

The organization of this paper is as follows. In Section II, we formulate our problem and present our main results. In Section III, we construct a measurement consisting of rank-one operators with measurement vectors closest to a given collection of states in the least-squares sense. In Section IV, we construct the optimal orthogonal LSM. Section V generalizes these results to allow for weighting of the squared norms of the error vectors. In Section VII, we discuss the relationships between our results and the previous results of Peres and Wootters [11] and Hausladen *et al.* [10]. We obtain a new characterization of the SRM, and summarize the properties of the SRM that follow from this characterization. In Section VIII, we discuss connections between the SRM and the measurement minimizing the probability of a detection error (minimum probability-of-error measurement (MPEM)). We show that for a geometrically uniform state set, the SRM is equivalent to the MPEM. We will consistently use [10] as our principal reference on the SRM.

II. Problem Statement and Main Results

In this section, we formulate our problem and describe our main results.

A. Problem Formulation

Assume that Alice conveys classical information to Bob by preparing a quantum channel in a pure quantum state drawn from a collection of given states $\{|\phi_i\rangle\}$. Bob's problem is to construct a measurement that will correctly determine the state of the channel with high probability.

Therefore, let $\{|\phi_i\rangle\}$ be a collection of $m \leq n$ normalized vectors $|\phi_i\rangle$ in an n-dimensional complex Hilbert space $\mathcal{H}$. Concretely, we may always identify $\mathcal{H}$ with $\mathbb{C}^n$ by choosisng appropriate coordinates. In general, these vectors are nonorthogonal and span an r-dimensional subspace $\mathcal{U} \subseteq \mathcal{H}$. The vectors are linearly independent if $r = m$.

For our measurement, we restrict our attention to POVMs consisting of m rank-one operators of the form $\Pi_i = |\mu_i\rangle\langle\mu_i|$ with measurement vectors $|\mu_i\rangle \in \mathcal{U}$. We do not require the vectors $|\mu_i\rangle$ to be orthogonal or normalized. However, to constitute a POVM the measurement vectors must satisfy

$$\sum_{i=1}^{m} \Pi_i = \sum_{i=1}^{m} |\mu_i\rangle\langle\mu_i| = P_{\mathcal{U}} \tag{1}$$

where $P_{\mathcal{U}}$ is the projection operator onto $\mathcal{U}$; i.e., the operators Π_i must be a resolution of the identity on $\mathcal{U}$.[1]

We seek the measurement vectors $|\mu_i\rangle$ such that one of the following quantities is minimized.

1) Squared error

$$E = \sum_{i=1}^{m} \langle e_i | e_i \rangle$$

where $|e_i\rangle = |\phi_i\rangle - |\mu_i\rangle$.

2) Weighted squared error

$$E_w = \sum_{i=1}^{m} w_i \langle e_i | e_i \rangle$$

for a given set of positive weights w_i.

B. Main Results

If the states $|\phi_i\rangle$ are linearly independent (i.e., if $r = m$), then the optimal solutions to problems (1) and (2) are of the same general form. We express this optimal solution in different ways. In particular, we find that the optimal solution is an orthogonal measurement and not a general POVM.

If $r < m$, then the solution to problem (1) still has the same general form. We show how it can be realized as an orthogonal measurement in an m-dimensional space. This orthogonal measurement is just a realization of the optimal POVM in a larger space than $\mathcal{U}$, along the lines suggested by Neumark's theorem [12], and it furnishes a physical interpretation of the optimal POVM.

We define a geometrically uniform (GU) state set as a collection of vectors $\mathcal{S} = \{|\phi_i\rangle = U_i|\phi\rangle,\ U_i \in \mathcal{G}\}$, where $\mathcal{G}$ is a finite abelian (commutative) group of m unitary matrices U_i, and $|\phi\rangle$

[1] Often these operators are supplemented by a projection

$$\Pi_0 = P_{\mathcal{U}^\perp} = I_{\mathcal{H}} - P_{\mathcal{U}}$$

onto the orthogonal subspace $\mathcal{U}^\perp \subseteq \mathcal{H}$, so that

$$\sum_{i=0}^{m} \Pi_i = I_{\mathcal{H}}$$

i.e., the augmented POVM is a resolution of the identity on $\mathcal{H}$. However, if the state vectors are confined to $\mathcal{U}$, then the probability of this additional outcome is 0, so we omit it.

is an arbitrary state. We show that for such a state set, the SRM minimizes the probability of a detection error.

Using these results, we can make the following remarks about [11] and the SRM [10].

1) The Peres–Wootters measurement is optimal in the least-squares sense and is equal to the SRM (strangely, this was not noticed in [10]); it also minimizes the probability of a detection error.

2) The SRM proposed by Hausladen *et al.* [10] minimizes the squared error. It may always be chosen as an orthogonal measurement equivalent to the optimal measurement in the linearly independent case. Further properties of the SRM are summarized in Theorem 3 (Section VII).

III. Least-Squares Measurement

Our objective is to construct a POVM with measurement vectors $|\mu_i\rangle$, optimized to distinguish between a collection of m pure states $|\phi_i\rangle$ that span a space $\mathcal{U} \subseteq \mathcal{H}$. A reasonable approach is to find a set of vectors $|\mu_i\rangle \in \mathcal{U}$ that are "closest" to the states $|\phi_i\rangle$ in the least-squares sense. Thus, our measurement consists of m rank-one positive operators of the form $\Pi_i = |\mu_i\rangle\langle\mu_i|$, $1 \le i \le m$. The measurement vectors $|\mu_i\rangle$ are chosen to minimize the squared error E, defined by

$$E = \sum_{i=1}^{m} \langle e_i | e_i \rangle \tag{2}$$

where $|e_i\rangle$ denotes the ith error vector

$$|e_i\rangle = |\phi_i\rangle - |\mu_i\rangle \tag{3}$$

subject to the constraint (1); i.e., the operators Π_i must be a resolution of the identity on $\mathcal{U}$.

If the vectors $|\phi_i\rangle$ are mutually orthonormal, then the solution to (2) satisfying the constraint (1) is simply $|\phi_i\rangle = |\mu_i\rangle$, $1 \le i \le m$, which yields $E = 0$.

To derive the solution in the general case where the vectors $|\phi_i\rangle$ are not orthonormal, denote by M and Φ the $n \times m$ matrices whose columns are the vectors $|\mu_i\rangle$ and $|\phi_i\rangle$, respectively. The squared error E of (2), (3) may then be expressed in terms of these matrices as

$$E = \mathrm{Tr}((\Phi - M)^*(\Phi - M)) = \mathrm{Tr}((\Phi - M)(\Phi - M)^*) \tag{4}$$

where $\mathrm{Tr}(\,\cdot\,)$ and $(\,\cdot\,)^*$ denote the trace and the Hermitian conjugate, respectively, and the second equality follows from the identity $\mathrm{Tr}(AB) = \mathrm{Tr}(BA)$ for all matrices A, B. The constraint (1) may then be restated as

$$MM^* = P_{\mathcal{U}}. \tag{5}$$

A. The Singular Value Decomposition (SVD)

The least-squares problem of (4) seeks a measurement matrix M that is "close" to the matrix Φ. If the two matrices are close, then we expect that the underlying linear transformations they represent will share similar properties. We therefore begin by decomposing the matrix Φ into elementary matrices that reveal these properties via the *singular value decomposition* (SVD) [17].

The SVD is known in quantum mechanics, but possibly not very well known. It has sometimes been presented as a corollary of the polar decomposition (e.g., in [18, Appendix A]). We present here a brief derivation based on the properties of eigendecompositions, since the SVD can be interpreted as a sort of "square root" of an eigendecomposition.

Let Φ be an arbitrary $n \times m$ complex matrix of rank r. Theorem 1 below asserts that Φ has an SVD of the form $\Phi = U\Sigma V^*$, with U and V unitary matrices and Σ diagonal. The elements of the SVD may be found from the eigenvalues and eigenvectors of the $m \times m$ nonnegative definite Hermitian matrix $S = \Phi^*\Phi$ and the $n \times n$ nonnegative definite Hermitian matrix $T = \Phi\Phi^*$. Notice that S is the Gram matrix of inner products $\langle\phi_i|\phi_j\rangle$, which completely determines the relative geometry of the vectors $\{|\phi_i\rangle\}$. It is elementary that both S and T have the same rank r as Φ, and that their nonzero eigenvalues are the same set of r positive numbers $\{\sigma_i^2,\, 1 \le i \le r\}$.

Theorem 1 (SVD): Let $\{|\phi_i\rangle\}$ be a set of m vectors in an n-dimensional complex Hilbert space $\mathcal{H}$, let $\mathcal{U} \subseteq \mathcal{H}$ be the subspace spanned by these vectors, and let $r = \dim\mathcal{U}$. Let Φ be the rank-r $n \times m$ matrix whose columns are the vectors $\{|\phi_i\rangle\}$. Then

$$\Phi = U\Sigma V^* = \sum_{i=1}^{r} \sigma_i |u_i\rangle\langle v_i|$$

where

1) $\Phi^*\Phi = V(\Sigma^*\Sigma)V^* = \sum_{i=1}^{r} \sigma_i^2 |v_i\rangle\langle v_i|$ is an eigendecomposition of the rank-r $m \times m$ matrix $S = \Phi^*\Phi$, in which

   a) the r positive real numbers $\{\sigma_i^2,\, 1 \le i \le r\}$ are the nonzero eigenvalues of S, and σ_i is the positive square root of σ_i^2;

   b) the r vectors $\{|v_i\rangle \in \mathbb{C}^m,\, 1 \le i \le r\}$ are the corresponding eigenvectors in the m-dimensional complex Hilbert space $\mathbb{C}^m$, normalized so that $\langle v_i|v_i\rangle = 1$;

   c) Σ is a diagonal $n \times m$ matrix whose first r diagonal elements are σ_i, and whose remaining $m - r$ diagonal elements are 0, so $\Sigma^*\Sigma$ is a diagonal $m \times m$ matrix with diagonal elements σ_i^2 for $1 \le i \le r$ and 0 otherwise;

   d) V is an $m \times m$ unitary matrix whose first r columns are the eigenvectors $|v_i\rangle$, which span a subspace $\mathcal{V} \subseteq \mathbb{C}^m$, and whose remaining $m - r$ columns $|v_i\rangle$ span the orthogonal complement $\mathcal{V}^\perp \subseteq \mathbb{C}^m$;

and

2) $\Phi\Phi^* = U(\Sigma\Sigma^*)U^* = \sum_{i=1}^{r} \sigma_i^2 |u_i\rangle\langle u_i|$ is an eigendecomposition of the rank-r $n \times n$ matrix $T = \Phi\Phi^*$, in which

   a) the r positive real numbers $\{\sigma_i^2,\, 1 \le i \le r\}$ are as before, but are now identified as the nonzero eigenvalues of T;

   b) the r vectors $\{|u_i\rangle \in \mathcal{H},\, 1 \le i \le r\}$ are the corresponding eigenvectors, normalized so that $\langle u_i|u_i\rangle = 1$;

c) Σ is as before, so $\Sigma\Sigma^*$ is a diagonal $n \times n$ matrix with diagonal elements σ_i^2 for $1 \leq i \leq r$ and 0 otherwise;

d) U is an $n \times n$ unitary matrix whose first r columns are the eigenvectors $|u_i\rangle$, which span the subspace $\mathcal{U} \subseteq \mathcal{H}$, and whose remaining $n - r$ columns $|u_i\rangle$ span the orthogonal complement $\mathcal{U}^\perp \subseteq \mathcal{H}$.

Since U is unitary, we have not only $U^*U = I_n$, which implies that the vectors $|u_k\rangle \in \mathcal{H}$ are orthonormal, $\langle u_k|u_j\rangle = \delta_{kj}$, but also that $UU^* = I_\mathcal{H}$, which implies that the rank-one projection operators $|u_k\rangle\langle u_k|$ are a resolution of the identity,

$$\sum_k |u_k\rangle\langle u_k| = I_\mathcal{H}.$$

Similarly, the vectors $|v_k\rangle \in \mathbb{C}^m$ are orthonormal and

$$\sum_k |v_k\rangle\langle v_k| = I_m.$$

These orthonormal bases for $\mathcal{H}$ and $\mathbb{C}^m$ will be called the U-basis and the V-basis, respectively. The first r vectors of the U-basis and the V-basis span the subspaces $\mathcal{U}$ and $\mathcal{V}$, respectively. Thus we refer to the set of vectors $\{|u_k\rangle, 1 \leq k \leq r\}$ as the $\mathcal{U}$-basis, and to the set $\{|v_k\rangle, 1 \leq k \leq r\}$ as the $\mathcal{V}$-basis.

The matrix Φ may be viewed as defining a linear transformation Φ: $\mathbb{C}^m \to \mathcal{H}$ according to $|v\rangle \mapsto \Phi|v\rangle$. The SVD allows us to interpret this map as follows. A vector $|v\rangle \in \mathbb{C}^m$ is first decomposed into its V-basis components via $|v\rangle = \sum_i |v_i\rangle\langle v_i|v\rangle$. Since Φ maps $|v_i\rangle$ to $\sigma_i|u_i\rangle$, Φ maps the ith component $|v_i\rangle\langle v_i|v\rangle$ to $\sigma_i|u_i\rangle\langle v_i|v\rangle$. Therefore, by superposition, Φ maps $|v\rangle$ to $\sum_i \sigma_i|u_i\rangle\langle v_i|v\rangle$. The kernel of the map Φ is thus $\mathcal{V}^\perp \subseteq \mathbb{C}^m$, and its image is $\mathcal{U} \subseteq \mathcal{H}$.

Similarly, the conjugate Hermitian matrix Φ^* defines the adjoint linear transformation Φ^*: $\mathcal{H} \to \mathbb{C}^m$ as follows: Φ^* maps $|u\rangle \in \mathcal{H}$ to $\sum_i \sigma_i|v_i\rangle\langle u_i|u\rangle \in \mathbb{C}^m$. The kernel of the adjoint map Φ^* is thus $\mathcal{U}^\perp \subseteq \mathcal{H}$, and its image is $\mathcal{V} \subseteq \mathbb{C}^m$.

The key element in these maps is the "transjector" (partial isometry) $|u_i\rangle\langle v_i|$, which maps the rank-one eigenspace of S generated by $|v_i\rangle$ into the corresponding eigenspace of T generated by $|u_i\rangle$, and the adjoint transjector $|v_i\rangle\langle u_i|$, which performs the inverse map.

B. The Least-Squares POVM

The SVD of Φ specifies orthonormal bases for $\mathcal{V}$ and $\mathcal{U}$ such that the linear transformations Φ and Φ^* map one basis to the other with appropriate scale factors. Thus, to find an M close to Φ we need to find a linear transformation M that performs a map similar to Φ.

Employing the SVD $\Phi = U\Sigma V^*$, we rewrite the squared error E of (4) as

$$\begin{aligned} E &= \mathrm{Tr}((\Phi - M)(\Phi - M)^*) \\ &= \mathrm{Tr}(U^*(\Phi - M)(\Phi - M)^*U) = \sum_{i=1}^n \langle d_i|d_i\rangle \end{aligned} \tag{6}$$

where

$$|d_i\rangle = (\Phi - M)^*|u_i\rangle. \tag{7}$$

The vectors $\{|u_i\rangle, 1 \leq i \leq r\}$ form an orthonormal basis for $\mathcal{U}$. Therefore, the projection operator onto $\mathcal{U}$ is given by

$$P_\mathcal{U} = \sum_{i=1}^r |u_i\rangle\langle u_i|. \tag{8}$$

Essentially, we want to construct a map M^* such that the images of the maps defined by Φ^* and M^* are as close as possible in the squared norm sense, subject to the constraint

$$MM^* = \sum_{i=1}^r |u_i\rangle\langle u_i|. \tag{9}$$

The SVD of Φ^* is given by $\Phi^* = V\Sigma^*U^*$. Consequently,

$$\Phi^*|u_i\rangle = \begin{cases} \sigma_i|v_i\rangle, & 1 \leq i \leq r \\ |0\rangle, & r+1 \leq i \leq n \end{cases} \tag{10}$$

where $|0\rangle$ denotes the zero vector. Denoting the image of $|u_i\rangle$ under M^* by $|a_i\rangle = M^*|u_i\rangle$, for any choice of M satisfying the constraint (9) we have

$$\langle a_i|a_i\rangle = \langle u_i|MM^*|u_i\rangle = \begin{cases} 1, & 1 \leq i \leq r \\ 0, & r+1 \leq i \leq n \end{cases} \tag{11}$$

and

$$\langle a_i|a_j\rangle = \langle u_i|MM^*|u_j\rangle = 0, \qquad i \neq j. \tag{12}$$

Thus, the vectors $|a_i\rangle$, $1 \leq i \leq r$, are mutually orthonormal and $|a_i\rangle = |0\rangle$, $r+1 \leq i \leq n$. Combining (10) and (11), we may express $|d_i\rangle$ as

$$|d_i\rangle = \begin{cases} \sigma_i|v_i\rangle - |a_i\rangle, & 1 \leq i \leq r \\ |0\rangle, & r+1 \leq i \leq n. \end{cases} \tag{13}$$

Our problem therefore reduces to finding a set of r orthonormal vectors $|a_i\rangle$ that minimize $E = \sum_{i=1}^r \langle d_i|d_i\rangle$, where $|d_i\rangle = \sigma_i|v_i\rangle - |a_i\rangle$. Since the vectors $|v_i\rangle$ are orthonormal, the minimizing vectors must be $|a_i\rangle = |v_i\rangle$, $1 \leq i \leq r$.

Thus, the optimal measurement matrix M, denoted by $\hat{M}$, satisfies

$$\hat{M}^*|u_i\rangle = \begin{cases} |v_i\rangle, & 1 \leq i \leq r \\ |0\rangle, & r+1 \leq i \leq n. \end{cases} \tag{14}$$

Consequently

$$\hat{M} = \sum_{i=1}^r |u_i\rangle\langle v_i|. \tag{15}$$

In other words, the optimal $\hat{M}$ is just the sum of the r transjectors of the map Φ.

We may express $\hat{M}$ in matrix form as

$$\hat{M} = UZ_rV^* \tag{16}$$

where Z_r, $1 \leq r \leq m$ is an $n \times m$ matrix defined by

$$Z_r = \left[\begin{array}{c|c} I_r & 0 \\ \hline 0 & 0 \end{array}\right]. \tag{17}$$

The residual squared error is then

$$E_{\min} = \sum_{i=1}^r (1 - \sigma_i)^2\langle v_i|v_i\rangle = \sum_{i=1}^r (1 - \sigma_i)^2. \tag{18}$$

Recall that $S = \Phi^*\Phi = V\Sigma^*\Sigma V^*$; thus $\mathrm{Tr}(S) = \sum_{i=1}^r \sigma_i^2$. Also, if the vectors $|\phi_i\rangle$ are normalized, then the diagonal elements of S are all equal to 1, so $\mathrm{Tr}(S) = m$. Therefore,

$$E_{\min} = \sum_{i=1}^{r}(1-\sigma_i)^2 = r + m - 2\sum_{i=1}^{r}\sigma_i. \qquad (19)$$

Note that if the singular values σ_i are distinct, then the vectors $|u_i\rangle$, $1 \le i \le r$ are unique (up to a phase factor $e^{j\theta_i}$). Given the vectors $|u_i\rangle$, the vectors $|v_i\rangle$ are uniquely determined, so the optimal measurement vectors corresponding to $\hat{M}$ are unique.

If, on the other hand, there are repeated singular values, then the corresponding vectors are not unique. Nonetheless, the choice of singular vectors does not affect $\hat{M}$. Indeed, if the vectors corresponding to a repeated singular value σ are $\{|u_j\rangle\}$, then $\sum_j |u_j\rangle\langle u_j|$ is a projection onto the corresponding eigenspace, and therefore is the same regardless of the choice of the vectors $\{|u_j\rangle\}$. Thus

$$\sum_j |u_j\rangle\langle v_j| = \sum_j \frac{1}{\sigma}|u_j\rangle\langle u_j|\Phi$$

independent of the choice of $\{|u_j\rangle\}$, and the optimal measurement is unique.

We may express $\hat{M}$ directly in terms of Φ as

$$\hat{M} = \Phi((\Phi^*\Phi)^{1/2})^\dagger \qquad (20)$$

where $(\cdot)^\dagger$ denotes the *Moore–Penrose pseudo-inverse* [17]; the inverse is taken on the subspace spanned by the columns of the matrix. Thus

$$((\Phi^*\Phi)^{1/2})^\dagger = V((\Sigma^*\Sigma)^{1/2})^\dagger V^*$$

where $((\Sigma^*\Sigma)^{1/2})^\dagger$ is a diagonal matrix with diagonal elements $1/\sigma_i$ for $1 \le i \le r$ and 0 otherwise; consequently, $\Phi((\Phi^*\Phi)^{1/2})^\dagger = UZ_rV^*$.

Alternatively, $\hat{M}$ may be expressed as

$$\hat{M} = ((\Phi\Phi^*)^{1/2})^\dagger\Phi \qquad (21)$$

where

$$((\Phi\Phi^*)^{1/2})^\dagger = U((\Sigma\Sigma^*)^{1/2})^\dagger U^*.$$

In Section VII, we will show that (21) is equivalent to the SRM proposed by Hausladen *et al.* [10].

In Appendix A we discuss some of the properties of the residual squared error $E_{\min}$.

IV. Orthogonal Least-Squares Measurement

In the previous section we sought the POVM consisting of rank-one operators that minimizes the least-squares error. We may similarly seek the optimal orthogonal measurement of the same form. We will explore the connection between the resulting optimal measurements both in the case of linearly independent states $|\phi_i\rangle$ $(r = m)$, and in the case of linearly dependent states $(r < m)$.

Linearly Independent States: If the states $|\phi_i\rangle$ are linearly independent and consequently Φ has full column rank (i.e., $r = m$), then (20) reduces to

$$\hat{M} = \Phi(\Phi^*\Phi)^{-1/2}. \qquad (22)$$

The optimal measurement vectors $|\hat{\mu}_i\rangle$ are mutually orthonormal, since their Gram matrix is

$$\hat{M}^*\hat{M} = (\Phi^*\Phi)^{-1/2}\Phi^*\Phi(\Phi^*\Phi)^{-1/2} = I_m. \qquad (23)$$

Thus, the optimal POVM is in fact an orthogonal measurement corresponding to projections onto a set of mutually orthonormal measurement vectors, which must of course be the optimal orthogonal measurement as well.

Linearly Dependent States: If the vectors $|\phi_i\rangle$ are linearly dependent, so that the matrix Φ does not have full column rank (i.e., $r < m$), then the m measurement vectors $|\hat{\mu}_i\rangle$ cannot be mutually orthonormal since they span an r-dimensional subspace. We therefore seek the orthogonal measurement M that minimizes the squared error E given by (4), subject to the orthonormality constraint $M^*M = I_m$.

In the previous section the constraint was on MM^*. Here the constraint is on M^*M, so we now write the squared error E as

$$\begin{aligned} E &= \mathrm{Tr}((\Phi - M)^*(\Phi - M)) \\ &= \mathrm{Tr}(V^*(\Phi - M)^*(\Phi - M)V) = \sum_{i=1}^{m}\langle\tilde{d}_i|\tilde{d}_i\rangle \end{aligned} \qquad (24)$$

where

$$|\tilde{d}_i\rangle = (\Phi - M)|v_i\rangle \qquad (25)$$

and where the columns $|v_i\rangle$ of V form the V-basis in the SVD of Φ. Essentially, we now want the images of the maps defined by Φ and M to be as close as possible in the squared norm sense.

The SVD of Φ is given by $\Phi = U\Sigma V^*$. Thus

$$\Phi|v_i\rangle = \begin{cases} \sigma_i|u_i\rangle, & 1 \le i \le r \\ |0\rangle, & r+1 \le i \le m. \end{cases} \qquad (26)$$

Denoting the images of $|v_i\rangle$ under M by $|b_i\rangle = M|v_i\rangle$, it follows from the constraint $M^*M = I$ that the vectors $|b_i\rangle$, $1 \le i \le m$, are orthonormal.

Our problem therefore reduces to finding a set of r orthonormal vectors $|b_i\rangle$ that minimize $\sum_{i=1}^r\langle\tilde{d}_i|\tilde{d}_i\rangle$, where $|\tilde{d}_i\rangle = \sigma_i|u_i\rangle - |b_i\rangle$ (since

$$\sum_{i=r+1}^{m}\langle\tilde{d}_i|\tilde{d}_i\rangle = \sum_{i=r+1}^{m}\langle b_i|b_i\rangle = m - r$$

independent of the choice of $|b_i\rangle$, $r+1 \le i \le m$). Since the vectors $|u_i\rangle$ are orthonormal, the minimizing vectors must be $|b_i\rangle = |u_i\rangle$, $1 \le i \le r$.

We may choose the remaining vectors $|b_i\rangle$, $r+1 \le i \le m$, arbitrarily, as long as the resulting m vectors $|b_i\rangle$ are mutually orthonormal. This choice will not affect the residual squared error. A convenient choice is $|b_i\rangle = |u_i\rangle$, $r+1 \le i \le m$. This results in an optimal measurement matrix denoted by $\tilde{M}$, namely

$$\tilde{M} = \sum_{i=1}^{m}|u_i\rangle\langle v_i|. \qquad (27)$$

We may express $\tilde{M}$ in matrix form as

$$\tilde{M} = UZ_mV^* \qquad (28)$$

where Z_m is given by (17) with $r = m$.

The residual squared error is then

$$\tilde{E}_{\min} = \sum_{i=1}^{r}(1-\sigma_i)^2\langle u_i|u_i\rangle + \sum_{i=r+1}^{m}\langle u_i|u_i\rangle = \sum_{i=1}^{r}(1-\sigma_i)^2 + m - r = E_{\min} + m - r \quad (29)$$

where $E_{\min}$ is given by (18).

Evidently, the optimal orthogonal measurement is not strictly unique. However, its action in the subspace $\mathcal{U}$ spanned by the vectors $|\phi_i\rangle$ and the resulting $\tilde{E}_{\min}$ are unique.

A. The Optimal Orthogonal Measurement and Neumark's Theorem

We now try to gain some insight into the orthogonal measurement. Our problem is to find a set of measurement vectors that are as close as possible to the states $|\phi_i\rangle$, where the states lie in an r-dimensional subspace $\mathcal{U}$. When $r = m$ we showed that the optimal measurement vectors $|\hat{\mu}_i\rangle$ are mutually orthonormal. However, when $r < m$, there are at most r orthonormal vectors in $\mathcal{U}$. Therefore, imposing an orthogonality constraint forces the optimal orthonormal measurement vectors $|\tilde{\mu}_i\rangle$ to lie partly in the orthogonal complement $\mathcal{U}^{\perp}$. The corresponding measurement consists of projections onto m orthonormal measurement vectors, where each vector has a component in $\mathcal{U}$, $|\tilde{\mu}_i^{\mathcal{U}}\rangle$, and a component in $\mathcal{U}^{\perp}$, $|\tilde{\mu}_i^{\mathcal{U}^{\perp}}\rangle$. We may express $\tilde{M}$ in terms of these components as

$$\tilde{M} = \tilde{M}^{\mathcal{U}} + \tilde{M}^{\mathcal{U}^{\perp}} \quad (30)$$

where $|\tilde{\mu}_i^{\mathcal{U}}\rangle$ and $|\tilde{\mu}_i^{\mathcal{U}^{\perp}}\rangle$ are the columns of $\tilde{M}^{\mathcal{U}}$ and $\tilde{M}^{\mathcal{U}^{\perp}}$, respectively. From (27) it then follows that

$$\tilde{M}^{\mathcal{U}} = \sum_{i=1}^{r}|u_i\rangle\langle v_i| \quad (31)$$

and

$$\tilde{M}^{\mathcal{U}^{\perp}} = \sum_{i=r+1}^{m}|u_i\rangle\langle v_i|. \quad (32)$$

Comparing (31) with (15), we conclude that $\tilde{M}^{\mathcal{U}} = \hat{M}$ and therefore $|\tilde{\mu}_i^{\mathcal{U}}\rangle = |\hat{\mu}_i\rangle$. Thus, although $|\tilde{\mu}_i\rangle \neq |\hat{\mu}_i\rangle$, their components in $\mathcal{U}$ are equal, i.e., $P_{\mathcal{U}}|\tilde{\mu}_i\rangle = |\hat{\mu}_i\rangle$.

Essentially, the optimal orthogonal measurement seeks m orthonormal measurement vectors $|\tilde{\mu}_i\rangle$ whose projections onto $\mathcal{U}$ are as close as possible to the m states $|\phi_i\rangle$. We now see that these projections are the measurement vectors $|\hat{\mu}_i\rangle$ of the optimal POVM. If we consider only the components of the measurement vectors that lie in $\mathcal{U}$, then

$$\tilde{E}_{\min}\sum_{i=1}^{r} = (1-\sigma_i)^2\langle u_i|u_i\rangle = E_{\min}.$$

Indeed, Neumark's theorem [12] shows that our optimal orthogonal measurement is just a realization of the optimal POVM. This theorem guarantees that any POVM with measurement operators of the form $\Pi_i = |\mu_i\rangle\langle\mu_i|$ may be realized by a set of orthogonal projection operators $\tilde{\Pi}_i$ in an extended space such that $\Pi_i = P\tilde{\Pi}_i P$, where P is the projection operator onto the original smaller space. Denoting by $\hat{\Pi}_i$ and $\tilde{\Pi}_i$ the optimal rank-one operators $|\hat{\mu}_i\rangle\langle\hat{\mu}_i|$ and $|\tilde{\mu}_i\rangle\langle\tilde{\mu}_i|$, respectively, (31) asserts that

$$\hat{\Pi}_i = P_{\mathcal{U}}\tilde{\Pi}_i P_{\mathcal{U}}. \quad (33)$$

Thus the optimal orthogonal measurement is a set of m projection operators in $\mathcal{H}$ that realizes the optimal POVM in the r-dimensional space $\mathcal{U} \subseteq \mathcal{H}$. This furnishes a physical interpretation of the optimal POVM. The two measurements are equivalent on the subspace $\mathcal{U}$.

We summarize our results regarding the LSM in the following theorem.

Theorem 2 (LSM): Let $\{|\phi_i\rangle\}$ be a set of m vectors in an n-dimensional complex Hilbert space $\mathcal{H}$ that span an r-dimensional subspace $\mathcal{U} \subseteq \mathcal{H}$. Let $\{|\hat{\mu}_i\rangle\}$ denote the optimal m measurement vectors that minimize the least-squares error defined by (2), (3), subject to the constraint (1). Let $\Phi = U\Sigma V^*$ be the rank-r $n \times m$ matrix whose columns are the vectors $|\phi_i\rangle$, and let $\hat{M}$ be the $n \times m$ measurement matrix whose columns are the vectors $|\hat{\mu}_i\rangle$. Then the unique optimal $\hat{M}$ is given by

$$\hat{M} = \sum_{i=1}^{r}|u_i\rangle\langle v_i| = UZ_rV^* = \Phi((\Phi^*\Phi)^{1/2})^{\dagger} = ((\Phi\Phi^*)^{1/2})^{\dagger}\Phi$$

where $|u_i\rangle$ and $|v_i\rangle$ denote the columns of U and V, respectively, and Z_r is defined in (17).

The residual squared error is given by

$$E_{\min} = \sum_{i=1}^{r}(1-\sigma_i)^2 = r + m - 2\sum_{i=1}^{r}\sigma_i$$

where $\{\sigma_i,\ 1 \le i \le r\}$ are the nonzero singular values of Φ. In addition

1) if $r = m$,
   a) $\hat{M} = \Phi(\Phi^*\Phi)^{-1/2}$;
   b) $\hat{M}^*\hat{M} = I_m$ and the corresponding measurement is an orthogonal measurement;
2) if $r < m$,
   a) $\hat{M}$ may be realized by the optimal orthogonal measurement
   $$\tilde{M} = \sum_{i=1}^{m}|u_i\rangle\langle v_i| = UZ_mV^*;$$
   b) the action of the two optimal measurements in the subspace $\mathcal{U}$ is the same.

V. Weighted LSM

In the previous section we sought a set of vectors $|\mu_i\rangle$ to minimize the sum of the squared errors $E = \sum_{i=1}^{m}\langle e_i|e_i\rangle$, where $|e_i\rangle = |\phi_i\rangle - |\mu_i\rangle$ is the ith error vector. Essentially, we are assigning equal weights to the different errors. However, in many cases we might choose to weight these errors according to some prior knowledge regarding the states $|\phi_i\rangle$. For example, if the state $|\phi_j\rangle$ is prepared with high probability, then we might wish to assign a large weight to $\langle e_j|e_j\rangle$. It may therefore be of interest to seek the vectors $|\mu_i\rangle$ that minimize a weighted squared error.

Thus, we consider the more general problem of minimizing the weighted squared error E_w given by

$$E_w = \sum_{i=1}^{m} w_i \langle e_i | e_i \rangle = \sum_{i=1}^{m} w_i (\langle \phi_i | - \langle \mu_i |)(|\phi_i\rangle - |\mu_i\rangle) \quad (34)$$

subject to the constraint

$$\sum_{i=1}^{m} |\mu_i\rangle\langle\mu_i| = P_{\mathcal{U}} \quad (35)$$

where $w_i > 0$ is the weight given to the ith squared norm error. Throughout this section we will assume that the vectors $|\phi_i\rangle$ are linearly independent and normalized.

The derivation of the solution to this minimization problem is analogous to the derivation of the LSM with a slight modification. In addition to the matrices M and Φ, we define an $m \times m$ diagonal matrix W with diagonal elements w_i. We further define $\Phi_w = \Phi W$. We then express E_w in terms of M, Φ_w, and W as

$$\begin{aligned} E_w &= \mathrm{Tr}((\Phi - M)^*(\Phi - M)W) \\ &= \mathrm{Tr}((\Phi_w - M)(\Phi_w - M)^*) + \mathrm{Tr}((W - I_m)M^*M) \\ &\quad + \mathrm{Tr}(W(I_m - W)\Phi^*\Phi). \end{aligned} \quad (36)$$

From (8) and (9), M must satisfy

$$MM^* = \sum_{i=1}^{m} |u_i\rangle\langle u_i| = P_{\mathcal{U}}$$

where $|u_i\rangle$ are the columns of U, the U-basis in the SVD of Φ. Consequently, M must be of the form $M = \sum_{i=1}^{m} |u_i\rangle\langle q_i|$, where the $|q_i\rangle$ are orthonormal vectors in $\mathbb{C}^m$, from which it follows that $M^*M = I_m$. Thus

$$\mathrm{Tr}(W(I_m - W)M^*M) = \mathrm{Tr}(W(I_m - W)).$$

Moreover, since $W(I_m - W)$ is diagonal and the vectors $|\phi_i\rangle$ are normalized, we have

$$\mathrm{Tr}(W(I_m - W)\Phi^*\Phi) = \mathrm{Tr}(W(I_m - W)).$$

Thus, we may express the squared error E_w as

$$\begin{aligned} E_w &= \mathrm{Tr}((\Phi_w - M)(\Phi_w - M)^*) - \mathrm{Tr}((I_m - W)(I_m - W)) \\ &= E'_w - \sum_{i=1}^{m} (1 - w_i)^2 \end{aligned} \quad (37)$$

where E'_w is defined as

$$E'_w = \mathrm{Tr}((\Phi_w - M)(\Phi_w - M)^*). \quad (38)$$

Thus minimization of E_w is equivalent to minimization of E'_w. Furthermore, this minimization problem is equivalent to the least-squares minimization given by (4), if we substitute Φ_w for Φ.

Therefore, we now employ the SVD of Φ_w, namely, $\Phi_w = U_w \Sigma_w V_w^*$. Since W is assumed to be invertible, the space spanned by the columns of $\Phi_w = \Phi W$ is equivalent to the space spanned by the columns of Φ, namely $\mathcal{U}$. Thus, the first m columns of U_w, denoted by $|u_i^w\rangle$, constitute an orthonormal basis for $\mathcal{U}$, and $MM^* = P_{\mathcal{U}}$, where

$$P_{\mathcal{U}} = \sum_{i=1}^{m} |u_i^w\rangle\langle u_i^w|. \quad (39)$$

We now follow the derivation of the previous section, where we substitute Φ_w for Φ and U_w, V_w and σ_i^w for U, V, and σ_i, respectively. The minimizing $\hat{M}_w$ follows from Theorem 2

$$\begin{aligned} \hat{M}_w &= \sum_{i=1}^{m} |u_i^w\rangle\langle v_i^w| = U_w Z_m V_w^* = \Phi_w(\Phi_w^*\Phi_w)^{-1/2} \\ &= \Phi W (W^*\Phi^*\Phi W)^{-1/2} \end{aligned} \quad (40)$$

where the $|v_i^w\rangle$ are the columns of V_w. The resulting error $E'_{\min}$ is given by

$$E'_{\min} = \sum_{i=1}^{m} (1 - \sigma_i^w)^2. \quad (41)$$

Defining $S_w = \Phi_w^*\Phi_w = V_w \Sigma_w^* \Sigma_w V_w^*$, we have

$$\mathrm{Tr}(S_w) = \sum_{i=1}^{m} (\sigma_i^w)^2.$$

In addition, $S_w = W\Phi^*\Phi W = WSW$. Assuming the vectors $|\phi_i\rangle$ are normalized, the diagonal elements of S are all equal to 1, so $\mathrm{Tr}(S_w) = \sum_{i=1}^{m} w_i^2$ and

$$E'_{\min} = m + \sum_{i=1}^{m} (w_i^2 - 2\sigma_i^w). \quad (42)$$

From (37), the residual squared error $E_{\min}^w$ is therefore given by

$$E_{\min}^w = 2\sum_{i=1}^{m} (w_i - \sigma_i^w). \quad (43)$$

Note that if $W = aI_m$ where $a > 0$ is an arbitrary constant, then $U_w = U$ and $V_w = V$, where U and V are the unitary matrices in the SVD of Φ. Thus in this case, as we expect, $\hat{M}_w = \hat{M}$, where $\hat{M}$ is the LSM given by (22).

It is interesting to compare the minimal residual squared error $E_{\min}^w$ of (43) with the $E_{\min}$ of (19) derived in the previous section for the nonweighted case, which for the case $r = m$ reduces to $E_{\min} = 2\sum_{i=1}^{m}(1 - \sigma_i)$. In the nonweighted case, $w_i = 1$ for all i, resulting in $W = I$ and $\mathrm{Tr}(W) = m$. Therefore, in order to compare the two cases, the weights should be chosen such that $\mathrm{Tr}(W) = \sum_{i=1}^{m} w_i = m$. (Note that only the ratios of the w_i's affect the WLSM. The normalization $\mathrm{Tr}(W) = m$ is chosen for comparison only.) In this case

$$E_{\min}^w - E_{\min} = 2\sum_{i=1}^{m} (\sigma_i - \sigma_i^w). \quad (44)$$

Recall that $(\sigma_i^w)^2$ and σ_i^2 are the eigenvalues of $S_w = WSW$ and S, respectively. We may therefore use Ostrowski's theorem (see Appendix A) to obtain the following bounds:

$$\begin{aligned} 2\left(1 - \max_i w_i\right)\sum_{i=1}^{m} \sigma_i &\le E_{\min}^w - E_{\min} \\ &\le 2\left(1 - \min_i w_i\right)\sum_{i=1}^{m} \sigma_i. \end{aligned} \quad (45)$$

Since $\max_i w_i \ge 1$ and $\min_i w_i \le 1$, $E_{\min}^w$ can be greater or smaller then $E_{\min}$, depending on the weights w_i.

VI. Example of the LSM and the WLSM

We now give an example illustrating the LSM and the WLSM. Consider the two states

$$|\phi_1\rangle = [1 \quad 0]^* \qquad |\phi_2\rangle = \tfrac{1}{2}[-1 \quad \sqrt{3}]^*. \tag{46}$$

We wish to construct the optimal LSM for distinguishing between these two states. We begin by forming the matrix Φ

$$\Phi = \frac{1}{2}\begin{bmatrix} 2 & -1 \\ 0 & \sqrt{3} \end{bmatrix}. \tag{47}$$

The vectors $|\phi_1\rangle$ and $|\phi_2\rangle$ are linearly independent, so Φ is a full-rank matrix ($r = 2$). Using Theorem 1 we may determine the SVD $\Phi = U\Sigma V^*$, which yields

$$\begin{aligned} U &= \frac{1}{2}\begin{bmatrix} \sqrt{3} & -1 \\ -1 & -\sqrt{3} \end{bmatrix} \\ \Sigma &= \frac{1}{\sqrt{2}}\begin{bmatrix} \sqrt{3} & 0 \\ 0 & 1 \end{bmatrix} \\ V &= \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & -1 \\ -1 & -1 \end{bmatrix}. \end{aligned} \tag{48}$$

From (16) and (17), we now have

$$\hat{M} = UV^* = \begin{bmatrix} 0.97 & -0.26 \\ 0.26 & 0.97 \end{bmatrix} \tag{49}$$

and

$$|\hat{\mu}_1\rangle = [0.97 \quad 0.26]^* \qquad |\hat{\mu}_2\rangle = [-0.26 \quad 0.97]^* \tag{50}$$

where $|\hat{\mu}_1\rangle$ and $|\hat{\mu}_2\rangle$ are the optimal measurement vectors that minimize the least-squares error defined by (2), (3). Using (22) we may express the optimal measurement vectors directly in terms of the vectors $|\phi_1\rangle$ and $|\phi_2\rangle$

$$\hat{M} = \Phi(\Phi^*\Phi)^{-1/2} = \Phi\begin{bmatrix} 1.12 & 0.30 \\ 0.30 & 1.12 \end{bmatrix} \tag{51}$$

thus

$$\begin{aligned} |\hat{\mu}_1\rangle &= 1.12|\phi_1\rangle + 0.30|\phi_2\rangle \\ |\hat{\mu}_2\rangle &= 0.30|\phi_1\rangle + 1.12|\phi_2\rangle. \end{aligned} \tag{52}$$

As expected from Theorem 2, $\langle\hat{\mu}_1|\hat{\mu}_2\rangle = 0$; the vectors $|\phi_1\rangle$ and $|\phi_2\rangle$ are linearly independent, so the optimal measurement vectors must be orthonormal. The LSM then consists of the orthogonal projection operators $\Pi_1 = |\hat{\mu}_1\rangle\langle\hat{\mu}_1|$ and $\Pi_2 = |\hat{\mu}_2\rangle\langle\hat{\mu}_2|$.

Fig. 1 depicts the vectors $|\phi_1\rangle$ and $|\phi_2\rangle$ together with the optimal measurement vectors $|\hat{\mu}_1\rangle$ and $|\hat{\mu}_2\rangle$. As is evident from (52) and from Fig. 1, the optimal measurement vectors are as close as possible to the corresponding states, given that they must be orthogonal.

Suppose now we are given the additional information $p_1 = p$ and $p_2 = 1-p$, where p_1 and p_2 denote the prior probabilities of $|\phi_1\rangle$ and $|\phi_2\rangle$, respectively, and $p \in (0, 1)$. We may still employ the LSM to distinguish between the two states. However, we expect that a smaller residual squared error may be achieved by employing a WLSM. In Fig. 2, we plot the residual squared error $E_{\min}^w$ given by (43) as a function of p, when using a WLSM with weights $w_1 = \sqrt{p}$ and $w_2 = \sqrt{1-p}$ (we will justify this choice of weights in Section VII). When $p = 1/2$, $w_1 = w_2$, and the resulting WLSM is equivalent to the LSM. For $p \neq 1/2$, the WLSM does indeed yield a smaller residual squared error than the LSM (for which the residual squared error is approximately 0.095).

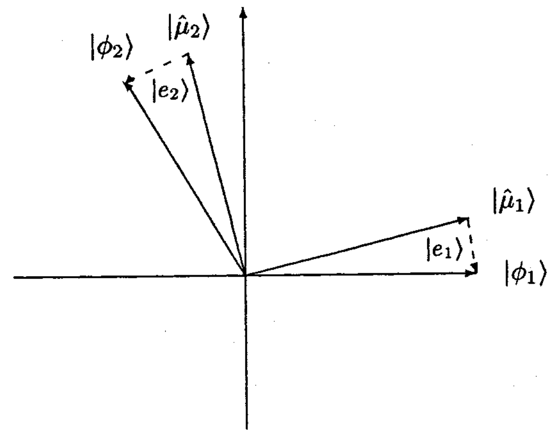

Fig. 1. Two-dimensional example of the LSM. The state vectors $|\phi_1\rangle$ and $|\phi_2\rangle$ are given by (46), the optimal measurement vectors $|\hat{\mu}_1\rangle$ and $|\hat{\mu}_2\rangle$ are given by (50) and are orthonormal, and $|e_1\rangle$ and $|e_2\rangle$ denote the error vectors defined in (3).

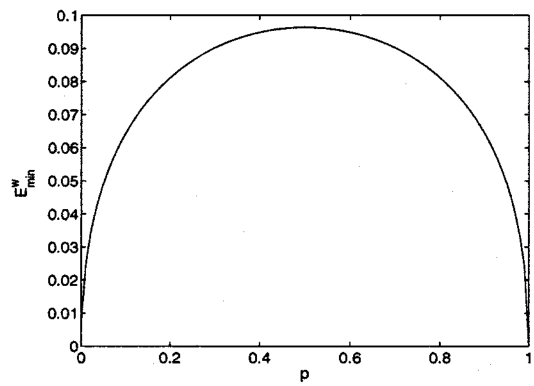

Fig. 2. Residual squared error $E_{\min}^w$ (43) as a function of p, the prior probability of $|\phi_1\rangle$, when using a WLSM. The weights are chosen as $w_1 = \sqrt{p}$ and $w_2 = \sqrt{1-p}$. For $p = 1/2$, the WLSM and the LSM coincide.

VII. Comparison with Other Proposed Measurements

We now compare our results with the SRM proposed by Hausladen *et al.* in [10], and with the measurement proposed by Peres and Wootters in [11].

Hausladen *et al.* construct a POVM consisting of rank-one operators $\Pi_i = |\mu_i\rangle\langle\mu_i|$ to distinguish between an arbitrary set of vectors $|\phi_i\rangle$. We refer to this POVM as the SRM. They give two alternative definitions of their measurement: Explicitly,

$$\overline{M} = ((\Phi\Phi^*)^{1/2})^\dagger\Phi \tag{53}$$

where $\overline{M}$ denotes the matrix of columns $|\overline{\mu}_i\rangle$. Implicitly, the optimal measurement vectors $|\overline{\mu_i}\rangle$ are those that satisfy

$$S^{1/2} = \{\langle\overline{\mu}_j|\phi_k\rangle\} \tag{54}$$

i.e., $\langle\overline{\mu}_j|\phi_k\rangle$ is equal to the jkth element of $S^{1/2}$, where $S = \Phi^*\Phi$.

Comparing (53) with (21), it is evident that the SRM coincides with the optimal LSM. Furthermore, following the discussion in Section IV, if the states are linearly independent then this measurement is a simple orthogonal measurement and not a more general POVM. (This observation was made in [13] as well.)

The implicit definition of (54) does not have a unique solution when the vectors $|\phi_i\rangle$ are linearly dependent. The columns of $\overline{M}$ are one solution of this equation. Since the definition depends only on the product $M^*\Phi$, any measurement vectors that are columns of M such that $M^*\Phi = \overline{M}^*\Phi$ constitutes a solution as well. In particular, the optimal orthogonal LSM $\tilde{M}$ for the linearly dependent case, given by (27), satisfies $\tilde{M}^*\Phi = \overline{M}^*\Phi$, rendering the optimal orthogonal LSM a solution to (54). Consequently, even in the case of linearly dependent states, the SRM proposed by Hausladen *et al.* and used to achieve the classical capacity of a quantum channel may always be chosen as an orthogonal measurement. In addition, this measurement is optimal in the least-squares sense.

We summarize our results regarding the SRM in the following theorem.

Theorem 3 (SRM): Let $\{|\phi_i\rangle\}$ be a set of m vectors in an n-dimensional complex Hilbert space $\mathcal{H}$ that span an r-dimensional subspace $\mathcal{U} \subseteq \mathcal{H}$. Let $\Phi = U\Sigma V^*$ be the rank-r $n \times m$ matrix whose columns are the vectors $|\phi_i\rangle$. Let $|u_i\rangle$ and $|v_i\rangle$ denote the columns of the unitary matrices U and V, respectively, and let Z_r be defined as in (17). Let $\{|\overline{\mu}_i\rangle\}$ be m vectors satisfying

$$S^{1/2} = \{\langle\overline{\mu}_j|\phi_k\rangle\}$$

where $S = \Phi^*\Phi$; a POVM consisting of the operators $\overline{\Pi}_i = |\overline{\mu}_i\rangle\langle\overline{\mu}_i|, 1 \le i \le m$, is referred to as SRM. Let $\overline{M}$ be the $n \times m$ measurement matrix whose columns are the vectors $|\overline{\mu}_i\rangle$; $\overline{M}$ is referred to as SRM matrix. Then

1) if $r = m$,
   a) $$\overline{M} = \sum_{i=1}^{m} |u_i\rangle\langle v_i| = UZ_mV^* = \Phi(\Phi^*\Phi)^{-1/2} = ((\Phi\Phi^*)^{1/2})^\dagger\Phi$$
   is unique;
   b) $\overline{M}^*\overline{M} = I_m$ and the corresponding SRM is an orthogonal measurement;
   c) the SRM is equal to the optimal LSM;

2) if $r < m$,
   a) the SRM is not unique;
   b) $$\overline{M} = \sum_{i=1}^{m} |u_i\rangle\langle v_i| = UZ_mV^*$$
   is SRM matrix; the corresponding SRM is equal to the optimal orthogonal LSM;
   c) define $\overline{M}_\mathcal{U} = P_\mathcal{U}\overline{M}$, where $P_\mathcal{U}$ is a projection onto $\mathcal{U}$ and $\overline{M}$ is any SRM matrix; then
      i) $\overline{M}_\mathcal{U}$ is unique, and is given by
      $$\overline{M}_\mathcal{U} = \sum_{i=1}^{r} |u_i\rangle\langle v_i| = UZ_rV^* = \Phi((\Phi^*\Phi)^{1/2})^\dagger = ((\Phi\Phi^*)^{1/2})^\dagger\Phi;$$
      ii) $\overline{M}_\mathcal{U}$ is a SRM matrix; the corresponding SRM is equal to the optimal LSM;
      iii) $\overline{M}_\mathcal{U}$ may be realized by the optimal orthogonal LSM
      $$\tilde{M} = \sum_{i=1}^{m} |u_i\rangle\langle v_i| = UZ_mV^* = \overline{M}.$$

The SRM defined in [10] does not take the prior probabilities of the states $|\phi_i\rangle$ into account. In [9], a more general definition of the SRM that accounts for the prior probabilities is given by defining new vectors $|\phi_i^w\rangle = \sqrt{p_i}|\phi_i\rangle$. The weighted SRM (WSRM) is then defined as the SRM corresponding to the vectors $|\phi_i^w\rangle$. Similarly, the WLSM is equal to the LSM corresponding to the vectors $w_i|\phi_i\rangle$. Thus, if we choose the weights w_i proportional to $\sqrt{p_i}$, then the WLSM coincides with the WSRM. A theorem similar to Theorem 3 may then be formulated where the WSRM and the WLSM are substituted for the SRM and the LSM.

We next apply our results to a problem considered by Peres and Wootters in [11]. The problem is to distinguish between three two-qubit states

$$|\phi_1\rangle = |aa\rangle \qquad |\phi_2\rangle = |bb\rangle \qquad |\phi_3\rangle = |cc\rangle \tag{55}$$

where $|a\rangle$, $|b\rangle$, and $|c\rangle$ correspond to polarizations of a photon at 0°, 60°, and 120°, and the states have equal prior probabilities. Since the vectors $|\phi_i\rangle$ are linearly independent, the optimal measurement vectors are the columns of $\hat{M}$ given by (20)

$$\hat{M} = \Phi(\Phi^*\Phi)^{-1/2}. \tag{56}$$

Substituting (55) in (56) results in the same measurement vectors $|\hat{\mu}_i\rangle$ as those proposed by Peres and Wootters. Thus, their measurement is optimal in the least-squares sense. Furthermore, the measurement that they propose coincides with the SRM for this case. In the next section, we will show that this measurement also minimizes the probability of a detection error.

VIII. The SRM for Geometrically Uniform State Sets

In this section, we will consider the case in which the collection of states has a strong symmetry property, called geometric uniformity [16]. Under these conditions, we show that the SRM is equivalent to the measurement minimizing the probability of a detection error, which we refer to as the MPEM. This result generalizes a similar result of Ban *et al.* [7].

A. Geometrically Uniform State Sets

Let $\mathcal{G}$ be a finite abelian (commutative) group of m unitary matrices U_i. That is, $\mathcal{G}$ contains the identity matrix I; if $\mathcal{G}$ contains U_i, then it also contains its inverse $U_i^{-1} = U_i^*$; the product U_iU_j of any two elements of $\mathcal{G}$ is in $\mathcal{G}$; and $U_iU_j = U_jU_i$ for any two elements in $\mathcal{G}$ [19].

A state set generated by $\mathcal{G}$ is a set

$$\mathcal{S} = \{|\phi_i\rangle = U_i|\phi\rangle, U_i \in \mathcal{G}\}$$

where $|\phi\rangle$ is an arbitrary state. The group $\mathcal{G}$ will be called the *generating group* of $\mathcal{S}$. Such a state set has strong symmetry properties, and will be called *geometrically uniform* (GU). For consistency with the symmetry of $\mathcal{S}$, we will assume equiprobable prior probabilities on $\mathcal{S}$.

If the group $\mathcal{G}$ contains a rotation R such that $R^k = I$ for some integer $k > 1$, then the GU state set $\mathcal{S}$ is linearly dependent, because $\sum_{j=1}^{k} R^j|\phi\rangle$ is a fixed point under R, and the only fixed point of a rotation is the zero vector $|0\rangle$.

Since $U_i^* = U_i^{-1}$, the inner product of two vectors in $\mathcal{S}$ is

$$\langle \phi_i|\phi_j\rangle = \langle \phi|U_i^{-1}U_j|\phi\rangle = s(U_i^{-1}U_j) \tag{57}$$

where s is the function on $\mathcal{G}$ defined by

$$s(U_i) = \langle \phi|U_i|\phi\rangle. \tag{58}$$

For fixed i, the set

$$U_i^{-1}\mathcal{G} = \{U_i^{-1}U_j,\, U_j \in \mathcal{G}\}$$

is just a permutation of $\mathcal{G}$ since $U_i^{-1}U_j \in \mathcal{G}$ for all i, j [19]. Therefore, the m numbers $\{s(U_i^{-1}U_j),\, 1 \le j \le m\}$ are a permutation of the numbers $\{s(U_i),\, 1 \le i \le m\}$. The same is true for fixed j. Consequently, every row and column of the $m \times m$ Gram matrix $S = \{\langle \phi_i|\phi_j\rangle\}$ is a permutation of the numbers $\{s(U_i),\, 1 \le i \le m\}$.

It will be convenient to replace the multiplicative group $\mathcal{G}$ by an additive group G to which $\mathcal{G}$ is isomorphic.[2] Every finite abelian group $\mathcal{G}$ is isomorphic to a direct product G of a finite number of cyclic groups: $\mathcal{G} \cong G = \mathbb{Z}_{m_1} \times \cdots \times \mathbb{Z}_{m_p}$, where $\mathbb{Z}_{m_k}$ is the cyclic additive group of integers modulo m_k, and $m = \prod_k m_k$ [19]. Thus, every element $U_i \in \mathcal{G}$ can be associated with an element $g \in G$ of the form $g = (g_1, g_2, \ldots, g_p)$, where $g_k \in \mathbb{Z}_{m_k}$. We denote this one-to-one correspondence by $U_i \leftrightarrow g$. Because the correspondence is an isomorphism, it follows that if $U_i \leftrightarrow g$, $U_k \leftrightarrow g'$, $U_l \leftrightarrow g''$, and $U_i = U_k U_l$, then $g = g' + g''$, where the addition of $g' = (g_1', g_2', \ldots, g_p')$ and $g'' = (g_1'', g_2'', \ldots, g_p'')$ is performed by componentwise addition modulo the corresponding m_k.

Each state vector $|\phi_i\rangle = U_i|\phi\rangle$ will henceforth be denoted as $|\phi(g)\rangle$, where $g \in G$ is the group element corresponding to $U_i \in \mathcal{G}$. The zero element $0 = (0, 0, \ldots, 0) \in G$ corresponds to the identity matrix $I \in \mathcal{G}$, and an additive inverse $-g \in G$ corresponds to a multiplicative inverse $U_i^{-1} = U_i^* \in \mathcal{G}$. The Gram matrix is then the $m \times m$ matrix

$$S = \{\langle \phi(g')|\phi(g)\rangle,\, g', g \in G\} = \{s(g - g'),\, g', g \in G\} \tag{59}$$

with row and column indices $g', g \in G$, where s is now the function on G defined by

$$s(g) = \langle \phi(0)|\phi(g)\rangle. \tag{60}$$

B. The SRM

We now obtain the SRM for a GU state set. We begin by determining the SVD of Φ. To this end we introduce the following definition. The Fourier transform (FT) of a complex-valued function φ: $G \to \mathbb{C}$ defined on $G = \mathbb{Z}_{m_1} \times \cdots \times \mathbb{Z}_{m_p}$ is the complex-valued function $\hat{\varphi}$: $G \to \mathbb{C}$ defined by

$$\hat{\varphi}(h) = \frac{1}{\sqrt{m}} \sum_{g \in G} \langle h, g\rangle \varphi(g) \tag{61}$$

where the Fourier kernel $\langle h, g\rangle$ is

$$\langle h, g\rangle = \prod_{k=1}^{p} e^{-2\pi i h_k g_k / m_k}. \tag{62}$$

[2]Two groups $\mathcal{G}$ and $\mathcal{G}'$ are *isomorphic*, denoted by $\mathcal{G} \cong \mathcal{G}'$, if there is a bijection (one-to-one and onto map) φ: $\mathcal{G} \to \mathcal{G}'$ which satisfies $\varphi(xy) = \varphi(x)\varphi(y)$ for all $x, y \in \mathcal{G}$ [19].

Here, h_k and g_k are the kth components of h and g, respectively, and the product $h_k g_k$ is taken as an ordinary integer modulo m_k. The Fourier kernel evidently satisfies

$$\langle h, g\rangle = \langle g, h\rangle \tag{63}$$

$$\langle h, g\rangle^* = \langle -h, g\rangle = \langle h, -g\rangle \tag{64}$$

$$\langle h + h', g\rangle = \langle h, g\rangle \langle h', g\rangle \tag{65}$$

$$\langle h, g + g'\rangle = \langle h, g\rangle \langle h, g'\rangle. \tag{66}$$

We define the FT matrix over G as the $m \times m$ matrix

$$\mathcal{F} = \left\{ \frac{1}{\sqrt{m}} \langle h, g\rangle,\, h, g \in G \right\}.$$

The FT of a column vector $|\varphi\rangle = \{\varphi(g),\, g \in G\}$ is then the column vector $|\hat{\varphi}\rangle = \{\hat{\varphi}(h),\, h \in G\}$ given by $|\hat{\varphi}\rangle = \mathcal{F}|\varphi\rangle$. It is easy to show that the rows and columns of $\mathcal{F}$ are orthonormal, i.e., $\mathcal{F}$ is unitary

$$\mathcal{F}^*\mathcal{F} = \mathcal{F}\mathcal{F}^* = I_m. \tag{67}$$

Consequently, we obtain the inverse FT formula

$$|\varphi\rangle = \mathcal{F}^*|\hat{\varphi}\rangle = \left\{ \frac{1}{\sqrt{m}} \sum_{h \in G} \langle h, g\rangle^* \hat{\varphi}(h),\, g \in G \right\}. \tag{68}$$

We now show that the eigenvectors of the Gram matrix S of (59) are the column vectors

$$|\mathcal{F}(h)\rangle = \left\{ \frac{1}{\sqrt{m}} \langle h, g\rangle,\, g \in G \right\}$$

of $\mathcal{F}$. Let $\langle S(g')| = \{s(g - g'),\, g \in G\}$ be the g'th row of S. Then

$$\begin{aligned}\langle S(g')|\mathcal{F}(h)\rangle &= \frac{1}{\sqrt{m}} \sum_{g \in G} \langle h, g\rangle s(g - g') \\ &= \frac{1}{\sqrt{m}} \sum_{g'' \in G} \langle h, g' + g''\rangle s(g'') = \langle h, g'\rangle \hat{s}(h)\end{aligned} \tag{69}$$

where the last equality follows from (66), and $\{\hat{s}(h),\, h \in G\}$ is the FT of $\{s(g),\, g \in G\}$. Thus, S has the eigendecomposition

$$S = \mathcal{F}\overline{\Sigma}^2\mathcal{F}^* \tag{70}$$

where $\overline{\Sigma}$ is an $m \times m$ diagonal matrix with diagonal elements

$$\{\sigma(h) = m^{1/4}\sqrt{\hat{s}(h)},\, h \in G\}$$

(the eigenvalues $\sigma^2(h)$ are real and nonnegative because S is Hermitian). Consequently, the V-basis of the SVD of Φ is $V = \mathcal{F}$, and the singular values of Φ are $\sigma(h)$.

We now write the SVD of Φ in the following form:

$$\Phi = \Upsilon\overline{\Sigma}\mathcal{F}^* = \sum_{h \in G} \sigma(h)|u(h)\rangle\langle \mathcal{F}^*(h)| \tag{71}$$

where Υ is the $n \times m$ matrix whose columns $|u(h)\rangle$ are the columns of the U-basis of the SVD of Φ for values of $h \in G$ such that $\sigma(h) \ne 0$ and are zero columns otherwise, and

$$\mathcal{F}^* = \left\{ \frac{1}{\sqrt{m}} \langle h, g\rangle^*,\, h, g \in G \right\}$$

has rows

$$\langle \mathcal{F}^*(h)| = \left\{ \frac{1}{\sqrt{m}} \langle h, g\rangle^*,\, g \in G \right\}.$$

It then follows that

$$|u(h)\rangle = \begin{cases} \Phi|\mathcal{F}(h)\rangle/\sigma(h) = |\hat{\phi}(h)\rangle/\sigma(h), & \text{if } \sigma(h) \ne 0 \\ |0\rangle, & \text{otherwise} \end{cases} \tag{72}$$

where

$$|\hat{\phi}(h)\rangle = \frac{1}{\sqrt{m}} \sum_{g \in G} \langle h, g\rangle |\phi(g)\rangle \tag{73}$$

is the hth element of the FT of Φ regarded as a row vector of column vectors $\Phi = \{|\phi(g)\rangle, g \in G\}$.

Finally, the SRM is given by the measurement matrix

$$M = \Upsilon \mathcal{F}^* = \sum_{h \in G} |u(h)\rangle\langle \mathcal{F}^*(h)|. \tag{74}$$

The measurement vectors $|\mu(g)\rangle$ (the columns of M) are thus the inverse FT of the columns of Υ

$$|\mu(g)\rangle = \frac{1}{\sqrt{m}} \sum_{h \in G} \langle g, h\rangle^* |u(h)\rangle. \tag{75}$$

Note that if $|\phi(g)\rangle = U_i|\phi\rangle$ where $U_i \leftrightarrow g$, and $U_j \leftrightarrow g'$, then

$$U_j|\phi(g)\rangle = U_j U_i |\phi\rangle = |\phi(g + g')\rangle.$$

Therefore, left multiplication of the state vectors

$$\Phi = \{|\phi(g)\rangle, g \in G\}$$

by U_j permutes the state vectors to

$$U_j \Phi = \{|\phi(g + g')\rangle, g \in G\}.$$

We now show that under this transformation the measurement vectors are similarly permuted, i.e.,

$$U_j M = \{|\mu(g + g')\rangle, g \in G\}.$$

The FT of the permuted vectors $\{|\phi(g + g')\rangle, g \in G\}$ is

$$\begin{aligned} |\hat{\phi}'(h)\rangle &= \frac{1}{\sqrt{m}} \sum_{g \in G} \langle h, g\rangle |\phi(g + g')\rangle \\ &= \frac{1}{\sqrt{m}} \sum_{g'' \in G} \langle h, g'' - g'\rangle |\phi(g'')\rangle \\ &= \langle h, g'\rangle^* |\hat{\phi}(h)\rangle. \end{aligned} \tag{76}$$

Normalization by $\sigma(h)^{-1}$ when $\sigma(h) \neq 0$ yields $|u'(h)\rangle = \langle h, g'\rangle^* |u(h)\rangle$. Finally, the inverse FT yields the measurement vectors

$$\begin{aligned} |\mu'(g)\rangle &= \frac{1}{\sqrt{m}} \sum_{h \in G} \langle g, h\rangle^* |u'(h)\rangle \\ &= \frac{1}{\sqrt{m}} \sum_{h \in G} \langle g + g', h\rangle^* |u(h)\rangle = |\mu(g + g')\rangle \end{aligned} \tag{77}$$

where we have used (63) and (65).

This shows that the measurement vectors $|\mu(g)\rangle$ have the same symmetries as the state vectors, i.e., they also form a GU set with generating group $\mathcal{G}$. Explicitly, if $U_i \leftrightarrow g$, then $|\mu(g)\rangle = U_i|\mu\rangle$, where $|\mu\rangle$ denotes $|\mu(0)\rangle$.

C. *The SRM and the MPEM*

We now show that for GU state sets, the SRM is equivalent to the MPEM. In the process, we derive a sufficient condition for the SRM to minimize the probability of a detection error for a general state set (not necessarily GU) comprised of linearly independent states.

Holevo [2], [4] and Yuen *et al.* [3] showed that a set of measurement operators Π_i comprises the MPEM for a set of weighted density operators $W_i = p_i \rho_i$ if they satisfy

$$\Pi_i (W_j - W_i) \Pi_j = \mathbf{0} \qquad \forall g, g' \tag{78}$$

$$\Gamma - W_i \geq \mathbf{0} \qquad \forall g \tag{79}$$

where

$$\Gamma = \sum_{j=1}^{m} \Pi_j W_j \tag{80}$$

and is required to be Hermitian. Note that if (78) is satisfied, then Γ is Hermitian.

In our case, the measurement operators Π_i are the operators $|\mu(g)\rangle\langle\mu(g)|$, and the weighted density operators may be taken simply as the projectors $|\phi(g)\rangle\langle\phi(g)|$, since their prior probabilities are equal. The conditions (78), (79) then become

$$\begin{aligned} &|\mu(g)\rangle\langle\mu(g)|\phi(g')\rangle\langle\phi(g')|\mu(g')\rangle\langle\mu(g')| \\ &\quad = |\mu(g)\rangle\langle\mu(g)|\phi(g)\rangle\langle\phi(g)|\mu(g')\rangle\langle\mu(g')| \qquad \forall g, g' \end{aligned} \tag{81}$$

$$\begin{aligned} &\sum_{g'} |\mu(g')\rangle\langle\mu(g')|\phi(g')\rangle\langle\phi(g')| \\ &\quad - |\phi(g)\rangle\langle\phi(g)| \geq \mathbf{0} \qquad \forall g. \end{aligned} \tag{82}$$

We first verify that the conditions (78) (or equivalently (81)) are satisfied. Since the matrix $M^*\Phi = \mathcal{F}\overline{\Sigma}\mathcal{F}^*$ is symmetric

$$\langle\mu(g')|\phi(g)\rangle = \langle\mu|U_j^{-1}U_i|\phi\rangle = w(g - g')$$

where $w(g) = \langle\mu|\phi(g)\rangle$ is a complex-valued function that satisfies $w(-g) = w^*(g)$. Therefore

$$\langle\mu(g)|\phi(g')\rangle = w(g' - g) = w^*(g - g') = \langle\phi(g)|\mu(g')\rangle \tag{83}$$

$$\langle\phi(g')|\mu(g')\rangle = w^*(0) = w(0) = \langle\mu(g)|\phi(g)\rangle. \tag{84}$$

Substituting these relations back into (81), we obtain

$$\begin{aligned} &w(0)w(g' - g)|\mu(g)\rangle\langle\mu(g')| \\ &\quad = w(0)w(g' - g)|\mu(g)\rangle\langle\mu(g')| \qquad \forall g, g' \end{aligned} \tag{85}$$

which verifies that the conditions (78) are satisfied.

Next, we show that conditions (79) are satisfied. Our proof is similar to that given in [7]. Since $M^*\Phi = \mathcal{F}\overline{\Sigma}\mathcal{F}^*$

$$w(0) = \langle\mu(g)|\phi(g)\rangle = \langle\mathcal{F}(g)|\overline{\Sigma}|\mathcal{F}(g)\rangle \tag{86}$$

where $\langle\mathcal{F}(g)|$ denotes the row of $\mathcal{F}$ corresponding to g. Then

$$\begin{aligned} \Gamma &= \sum_{g'} |\mu(g')\rangle\langle\mu(g')|\phi(g')\rangle\langle\phi(g')| \\ &= w(0) \sum_{g'} |\mu(g')\rangle\langle\phi(g')|. \end{aligned} \tag{87}$$

From (71) and (74) we have

$$\sum_{g'} |\mu(g')\rangle\langle\phi(g')| = \Upsilon\overline{\Sigma}\Upsilon^* \tag{88}$$

and

$$|\phi(g)\rangle\langle\phi(g)| = \Upsilon\overline{\Sigma}|\mathcal{F}(g)\rangle\langle\mathcal{F}(g)|\overline{\Sigma}\Upsilon^*. \tag{89}$$

Substituting (87)–(89) back into (82), the conditions of (82) reduce to

$$\Upsilon\left(w(0)\overline{\Sigma} - \overline{\Sigma}|\mathcal{F}(g)\rangle\langle\mathcal{F}(g)|\overline{\Sigma}\right)\Upsilon^* \geq \mathbf{0} \tag{90}$$

where $w(0)$ is given by (86). It is therefore sufficient to show that

$$T = w(0)\overline{\Sigma} - \overline{\Sigma}|\mathcal{F}(g)\rangle\langle\mathcal{F}(g)|\overline{\Sigma} \geq \mathbf{0} \tag{91}$$

or equivalently, that $\langle u|T|u\rangle \geq 0$ for any $|u\rangle \in \mathbb{C}^m$. Using the Cauchy–Schwartz inequality, we have

$$\begin{aligned}\langle u|T|u\rangle &= \langle\mathcal{F}(g)|\overline{\Sigma}|\mathcal{F}(g)\rangle\langle u|\overline{\Sigma}|u\rangle - \langle u|\overline{\Sigma}|\mathcal{F}(g)\rangle\langle\mathcal{F}(g)|\overline{\Sigma}|u\rangle\\ &\geq \langle\mathcal{F}(g)|\overline{\Sigma}|\mathcal{F}(g)\rangle\langle u|\overline{\Sigma}|u\rangle - \langle\mathcal{F}(g)|\overline{\Sigma}|\mathcal{F}(g)\rangle\langle u|\overline{\Sigma}|u\rangle\\ &= 0\end{aligned} \tag{92}$$

which verifies that the conditions (79) are satisfied.

We conclude that when the state set $\mathcal{S}$ is GU, the SRM is also the MPEM.

An alternative way of deriving this result for the case of linearly independent states $|\phi_i\rangle$ is by use of the following criterion of Sasaki *et al.* [13]. Denote by Φ_w the matrix whose columns are the vectors $|\phi_i^w\rangle = \sqrt{p_i}|\phi_i\rangle$ where p_i is the prior probability of state i. If the states are linearly independent and $S^{1/2} = (\Phi_w^*\Phi_w)^{1/2}$ has constant diagonal elements, then the SRM corresponding to the vectors $|\phi_i^w\rangle$ (i.e., a WSRM), is equivalent to the MPEM.

This condition is hard to verify directly from the vectors $|\phi_i^w\rangle$. The difficulty arises from the fact that generally there is no simple relation between the diagonal elements of $S^{1/2}$ and the elements of S. Thus, given an ensemble of pure states $|\phi_i\rangle$ with prior probabilities p_i, we typically need to calculate $S^{1/2}$ (which in itself is not simple to do analytically) in order to verify the condition above. However, as we now show, in some cases this condition may be verified directly from the elements of S using the SVD.

Employing the SVD $\Phi_w = U\Sigma V^*$ we may express $S^{1/2}$ as

$$S^{1/2} = (\Phi_w^*\Phi_w)^{1/2} = V(\Sigma^*\Sigma)^{1/2}V^* = V\overline{\Sigma}V^* \tag{93}$$

where $\overline{\Sigma}$ is a diagonal matrix with the first r diagonal elements equal to σ_i, and the remaining elements all equal to zero, where the σ_i are the singular values of Φ_w. Thus, the WSRM is equal to the MPEM if $\langle\overline{v}_i|\overline{\Sigma}|\overline{v}_i\rangle = c$, $1 \leq i \leq m$, where the vectors $|\overline{v}_i\rangle$ denote the columns of V^*, and c is a constant. In particular, if the elements of V all have equal magnitude, then $\langle\overline{v}_i|\overline{\Sigma}|\overline{v}_i\rangle$ is constant, and the SRM minimizes the probability of a detection error.

If the state set $\mathcal{S}$ is GU, then the matrix V is the FT matrix $\mathcal{F}$, whose elements all have equal magnitude. Thus, if the states are linearly independent and GU, then the SRM is equivalent to the MPEM.

We summarize our results regarding GU state sets in the following theorem.

Theorem 4 (SRM for GU State Sets): Let

$$\mathcal{S} = \{|\phi_i\rangle = U_i|\phi\rangle,\, U_i \in \mathcal{G}\}$$

be a geometrically uniform state set generated by a finite abelian group $\mathcal{G}$ of unitary matrices, where $|\phi\rangle$ is an arbitrary state. Let $\mathcal{G} \cong G$, and let Φ be the matrix of columns $|\phi_i\rangle$. Then the SRM is given by the measurement matrix

$$M = \Phi\mathcal{F}\overline{\Sigma}^{\dagger}\mathcal{F}^* = \sum_{h\in G} |u(h)\rangle\langle\mathcal{F}^*(h)|$$

where $\mathcal{F}$ is the FT matrix over G, $\overline{\Sigma}^{\dagger}$ is the diagonal matrix whose diagonal elements are $\sigma(h)^{-1}$ when $\sigma(h) \neq 0$ and 0 otherwise, where $\{\sigma(h),\, h \in G\}$ are the singular values of Φ

$$|u(h)\rangle = |\hat{\phi}(h)\rangle/\sigma(h)$$

when $\sigma(h) \neq 0$ and $|0\rangle$ otherwise, where $\{|\hat{\phi}(h)\rangle,\, h \in G\}$ is the FT of $\{|\phi(g)\rangle,\, g \in G\}$, and $\langle\mathcal{F}^*(h)|$ is the hth row of $\mathcal{F}^*$.

The SRM has the following properties:

1) the measurement matrix M has the same symmetries as Φ;
2) the SRM is the LSM;
3) the SRM is the MPEM.

D. Example of a GU State Set

We now consider an example demonstrating the ideas of the previous section. Consider the group $\mathcal{G}$ of $m = 4$ unitary matrices U_i, where

$$U_1 = I_4 \qquad U_2 = \begin{bmatrix} -1 & 0 & 0 & 0\\ 0 & 1 & 0 & 0\\ 0 & 0 & -1 & 0\\ 0 & 0 & 0 & -1\end{bmatrix}$$
$$U_3 = \begin{bmatrix} -1 & 0 & 0 & 0\\ 0 & -1 & 0 & 0\\ 0 & 0 & 1 & 0\\ 0 & 0 & 0 & -1\end{bmatrix} \qquad U_4 = U_2U_3. \tag{94}$$

Let the state set be

$$\mathcal{S} = \{|\phi_i\rangle = U_i|\phi\rangle,\, 1 \leq i \leq 4\}$$

where $|\phi\rangle = \frac{1}{2}[1\ 1\ 1\ 1]^*$. Then Φ is

$$\Phi = \frac{1}{2}\begin{bmatrix} 1 & -1 & -1 & 1\\ 1 & 1 & -1 & -1\\ 1 & -1 & 1 & -1\\ 1 & -1 & -1 & 1\end{bmatrix} \tag{95}$$

and the Gram matrix S is given by

$$S = \frac{1}{2}\begin{bmatrix} 2 & -1 & -1 & 0\\ -1 & 2 & 0 & -1\\ -1 & 0 & 2 & -1\\ 0 & -1 & -1 & 2\end{bmatrix}. \tag{96}$$

Note that the sum of the states $|\phi_i\rangle$ is $|0\rangle$, so the state set is linearly dependent.

In this case, $\mathcal{G}$ is isomorphic to $G = \mathbb{Z}_2 \times \mathbb{Z}_2$, i.e.,

$$G = \{(0,0),\, (0,1),\, (1,0),\, (1,1)\}.$$

The multiplication table of the group $\mathcal{G}$ is

$$\begin{array}{c|cccc} & U_1 & U_2 & U_3 & U_4 \\ \hline U_1 & U_1 & U_2 & U_3 & U_4 \\ U_2 & U_2 & U_1 & U_4 & U_3 \\ U_3 & U_3 & U_4 & U_1 & U_2 \\ U_4 & U_4 & U_3 & U_2 & U_1. \end{array} \tag{97}$$

If we define the correspondence

$$U_1 \leftrightarrow (0, 0) \quad U_2 \leftrightarrow (0, 1) \quad U_3 \leftrightarrow (1, 0) \quad U_4 \leftrightarrow (1, 1) \tag{98}$$

then this table becomes the addition table of $G = \mathbb{Z}_2 \times \mathbb{Z}_2$:

$$\begin{array}{c|cccc} & (0,0) & (0,1) & (1,0) & (1,1) \\ \hline (0,0) & (0,0) & (0,1) & (1,0) & (1,1) \\ (0,1) & (0,1) & (0,0) & (1,1) & (1,0) \\ (1,0) & (1,0) & (1,1) & (0,0) & (0,1) \\ (1,1) & (1,1) & (1,0) & (0,1) & (0,0). \end{array} \tag{99}$$

Only the way in which the elements are labeled distinguishes the table of (99) from the table of (97); thus $\mathcal{G} \cong G$. Comparing (97) and (99) with (96), we see that the tables and the matrix S have the same symmetries.

Over $G = \mathbb{Z}_2 \times \mathbb{Z}_2$, the Fourier matrix $\mathcal{F}$ is the Hadamard matrix

$$\mathcal{F} = \frac{1}{2}\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix}. \tag{100}$$

Using (72) and (74), we may find the measurement matrix of the SRM

$$M = \frac{1}{2\sqrt{2}}\begin{bmatrix} 1 & -1 & -1 & 1 \\ \sqrt{2} & \sqrt{2} & -\sqrt{2} & -\sqrt{2} \\ \sqrt{2} & -\sqrt{2} & \sqrt{2} & -\sqrt{2} \\ 1 & -1 & -1 & 1 \end{bmatrix}. \tag{101}$$

We verify that the columns $|\mu_i\rangle$ of M may be expressed as $|\mu_i\rangle = U_i|\mu_1\rangle$, $1 \le i \le 4$, where $|\mu_1\rangle = \frac{1}{2\sqrt{2}}[1\ \sqrt{2}\ \sqrt{2}\ 1]^*$. Thus, the measurement vectors $|\mu_i\rangle$ also form a GU set generated by $\mathcal{G}$.

E. Applications of GU State Sets

We now discuss some applications of Theorem 4.

1) Binary State Set: Any binary state set $\mathcal{S} = \{|\phi_1\rangle, |\phi_2\rangle\}$ is GU, because it can be generated by the binary group $\mathcal{G} = \{I, R\}$, where I is the identity and R is the reflection about the hyperplane halfway between the two states. Specifically, if the two states $|\phi_1\rangle$ and $|\phi_2\rangle$ are real, then

$$R = I - 2\frac{|w\rangle\langle w|}{\langle w|w\rangle} \tag{102}$$

where $|w\rangle = |\phi_2\rangle - |\phi_1\rangle$. We may immediately verify that $R^2 = I$, so that $R^{-1} = R$, and that $|\phi_2\rangle = R|\phi_1\rangle$.

If the states are complex with $\langle\phi_1|\phi_2\rangle = ae^{j\theta}$, then define $|\phi_2'\rangle = e^{-j\theta}|\phi_2\rangle$. The states $|\phi_2\rangle$ and $|\phi_2'\rangle$ differ by a phase factor and therefore correspond to the same physical state. We may therefore replace our state set $\mathcal{S} = \{|\phi_1\rangle, |\phi_2\rangle\}$ by the equivalent state set $\mathcal{S} = \{|\phi_1\rangle, |\phi_2'\rangle\}$. Now the generating group is $\mathcal{G} = \{I, R\}$, where R is defined by (102), with $|w\rangle = |\phi_2'\rangle - |\phi_1\rangle$.

The generating group $\mathcal{G} = \{I, R\}$ is isomorphic to $G = \mathbb{Z}_2$. The Fourier matrix $\mathcal{F}$ therefore reduces to the 2×2 discrete FT (DFT) matrix

$$\mathcal{F} = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}. \tag{103}$$

The squares of the singular values of Φ are therefore

$$\{\sigma^2(h) = \sqrt{2}\hat{s}(h),\ h \in G\}$$

where $\{\hat{s}(h),\ h \in G\}$ are the DFT values of $\{s(g),\ g \in G\}$, with $s(0) = 1$ and $s(1) = a$. Thus

$$\begin{aligned} \sigma^2(0) &= 1 + a \\ \sigma^2(1) &= 1 - a. \end{aligned} \tag{104}$$

From Theorem 4 we then have

$$\begin{aligned} M &= \Phi\mathcal{F}\overline{\Sigma}^{\dagger}\mathcal{F}^* \\ &= \frac{1}{2}\Phi\begin{bmatrix} \frac{1}{\sigma(0)} + \frac{1}{\sigma(1)} & \frac{1}{\sigma(0)} - \frac{1}{\sigma(1)} \\ \frac{1}{\sigma(0)} - \frac{1}{\sigma(1)} & \frac{1}{\sigma(0)} + \frac{1}{\sigma(1)} \end{bmatrix}. \end{aligned} \tag{105}$$

We may now apply (105) to the example of Section VI. In that example $a = \langle\phi_1|\phi_2\rangle = -1/2$. From (104) it then follows that $\sigma(0) = 1/\sqrt{2}$ and $\sigma(1) = \sqrt{3/2}$. Substituting these values in (105) yields

$$M = \Phi\begin{bmatrix} 1.12 & 0.30 \\ 0.30 & 1.12 \end{bmatrix} \tag{106}$$

which is equivalent to the optimal measurement matrix obtained in Section VI.

We could have obtained the measurement vectors directly from the symmetry property of Theorem 4.1. The state set $\mathcal{S} = \{|\phi_1\rangle, |\phi_2\rangle\}$ is invariant under a reflection about the line halfway between the two states, as illustrated in Fig. 3. The measurement vectors must also be invariant under the same reflection. In addition, since the states are linearly independent, the measurement vectors must be orthonormal. This completely determines the measurement vectors shown in Fig. 3. (The only other possibility, namely, the negatives of these two vectors, is physically equivalent.)

2) Cyclic State Set: A cyclic generating group $\mathcal{G}$ has elements $U_i = Q^{i-1}$, $1 \le i \le m$, where Q is a unitary matrix with $Q^m = I$. A cyclic group generates a cyclic state set

$$\mathcal{S} = \{|\phi_i\rangle = Q^{i-1}|\phi\rangle,\ 1 \le i \le m\}$$

where $|\phi\rangle$ is arbitrary. Ban *et al.* [7] refer to such a cyclic state set as a symmetrical state set, and show that in that case the SRM is equivalent to the MPEM. This result is a special case of Theorem 4.

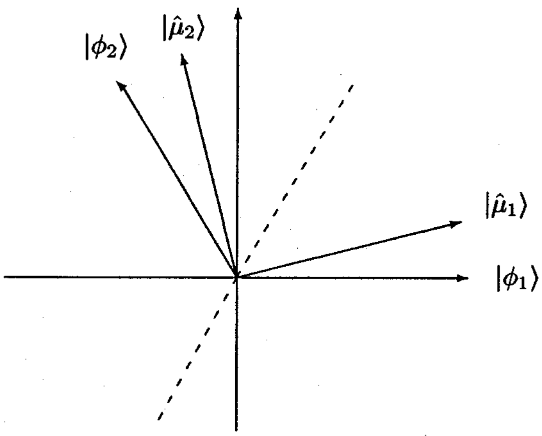

Fig. 3. Symmetry property of the state set $\mathcal{S} = \{|\phi_1\rangle, |\phi_2\rangle\}$ and the optimum measurement vectors $\{|\hat{\mu}_1\rangle, |\hat{\mu}_2\rangle\}$. $|\phi_1\rangle$ and $|\phi_2\rangle$ are given by (46), and $|\hat{\mu}_1\rangle$ and $|\hat{\mu}_2\rangle$ are given by (50). Because the state vectors are invariant under a reflection about the dashed line, the optimum measurement vectors must also have this property. In addition, the measurement vectors must be orthonormal. The symmetry and orthonormality properties completely determine the optimum measurement vectors $\{|\hat{\mu}_1\rangle, |\hat{\mu}_2\rangle\}$ (up to sign reversal).

Using Theorem 4 we may obtain the measurement matrix M as follows. If $\mathcal{G}$ is cyclic, then S is a circulant matrix,[3] and G is the cyclic group $\mathbb{Z}_m$. The FT kernel is then $\langle h, g\rangle = e^{-2\pi i hg/m}$ for $h, g \in \mathbb{Z}_m$, and the Fourier matrix $\mathcal{F}$ reduces to the $m \times m$ DFT matrix. The singular values of Φ are $m^{1/4}$ times the square roots of the DFT values of the inner products

$$\{\langle\phi_1|\phi_j\rangle, 1 \le j \le m\}.$$

We then calculate $M = \Phi\mathcal{F}\overline{\Sigma}^{\dagger}\mathcal{F}^*$.

3) Peres–Wootters Measurement: We may apply these results to the Peres–Wootters problem considered at the end of Section VII. In this problem, the states to be distinguished are given by $|\phi_1\rangle = |aa\rangle$, $|\phi_2\rangle = |bb\rangle$, and $|\phi_3\rangle = |cc\rangle$, where $|a\rangle$, $|b\rangle$, and $|c\rangle$ correspond to polarizations of a photon at 0°, 60°, and 120°, and the states have equal prior probabilities. The state set $\mathcal{S} = \{|\phi_1\rangle, |\phi_2\rangle, |\phi_3\rangle\}$ is thus a cyclic state set with $|\phi_i\rangle = U_i|\phi_1\rangle$, $1 \le i \le 3$, where $U_i = (Q \otimes Q)^{i-1}$ and Q is a rotation by 60°.

In Section VII, we concluded that the Peres–Wootters measurement is equivalent to the SRM and consequently minimizes the squared error. From Theorem 4 we now conclude that the Peres–Wootters measurement minimizes the probability of a detection error as well.

IX. Conclusion

In this paper, we constructed optimal measurements in the least-squares sense for distinguishing between a collection of quantum states. We considered POVMs consisting of rank-one operators, where the vectors were chosen to minimize a possibly weighted sum of squared errors. We saw that for linearly independent states, the optimal LSM is an orthogonal measurement, which coincides with the SRM proposed by Hausladen *et al.* [10]. If the states are linearly dependent, then the optimal POVM still has the same general form. We showed that it may be realized by an orthogonal measurement of the same form as in the linearly independent case. We also noted that the SRM, which was constructed by Hausladen *et al.* [10] and used to achieve the classical channel capacity of a quantum channel, may always be chosen as an orthogonal measurement.

We showed that for a GU state set the SRM minimizes the probability of a detection error. We also derived a sufficient condition for the SRM to minimize the probability of a detection error in the case of linearly independent states based on the properties of the SVD.

Appendix A
Properties of the Residual Squared Error

We noted at the beginning of Section III that if the vectors $|\phi_i\rangle$ are mutually orthonormal, then the optimal measurement is a set of projections onto the states $|\phi_i\rangle$, and the resulting squared error is zero. In this case, $S = \Phi^*\Phi = I_m$ and $\sigma_i = 1$, $1 \le i \le m$.

If the vectors $|\phi_i\rangle$ are normalized but not orthogonal, then we may decompose S as $S = I_m + D$, where D is the matrix of inner products $\langle\phi_i|\phi_j\rangle$ for $i \ne j$ and has diagonal elements all equal to 0. We expect that if the inner products are relatively small, i.e., if the states $|\phi_i\rangle$ are nearly orthonormal, then we will be able to distinguish between them pretty well; equivalently, we would expect the singular values to be close to 1. Indeed, from [20] we have the following bound on the singular values of $S = I + D$:

$$|\sigma_i^2 - 1|^2 \le \mathrm{Tr}(D^*D), \qquad 1 \le i \le m. \tag{107}$$

We now point out some properties of the minimal achievable squared error $E_{\min}$ given by (19). For a given m, $E_{\min}$ depends only on the singular values of the matrix Φ. Consequently, any linear operation on the vectors $|\phi_i\rangle$ that does not affect the singular values of Φ will not affect $E_{\min}$.

For example, if we obtain a new set of states $|\phi_i'\rangle$ by unitary mixing of the states $|\phi_i\rangle$, i.e., $\Phi' = \Phi Q^*$ where Q is an $m \times m$ unitary matrix, then the new optimal measurement vectors $|\mu_i'\rangle$ will typically differ from the measurement vectors $|\hat{\mu}_i\rangle$; however, the minimal achievable squared error is the same. Indeed, defining $S' = \Phi'^*\Phi' = QSQ^*$, where $S = \Phi^*\Phi$, we see that the matrices S' and S are related through a similarity transformation and consequently have equal eigenvalues [20].

Next, suppose we obtain a new set of states $|\phi_i'\rangle$ by a general nonsingular linear mixing of the states $|\phi_i\rangle$, i.e., $\Phi' = \Phi A^*$, where A is an arbitrary $m \times m$ nonsingular matrix. In this case, the eigenvalues of $S' = ASA^*$ will in general differ from the eigenvalues of S. Nevertheless, we have the following theorem:

[3] A circulant matrix is a matrix where every row (or column) is obtained by a right circular shift (by one position) of the previous row (or column). An example is

$$\begin{bmatrix} a_0 & a_2 & a_1 \\ a_1 & a_0 & a_2 \\ a_2 & a_1 & a_0 \end{bmatrix}.$$

Theorem 5: Let $E_{\min}$ and $E'_{\min}$ denote the minimal achievable squared error when distinguishing between the pure state ensembles $\{|\phi_i\rangle\}$ and $\{|\phi'_i\rangle\}$ respectively, where

$$|\phi'_i\rangle = \sum_{j=1}^{m} a^*_{ij} |\phi_j\rangle.$$

Let A denote the matrix whose ijth element is a_{ij}. Let $\lambda_1(AA^*)$ and $\lambda_m(AA^*)$ denote the largest and smallest eigenvalues of AA^*, respectively, and let $\{\sigma_i,\ 1 \leq i \leq r\}$ denote the singular values of the matrix Φ of columns $|\phi_i\rangle$. Then

$$2\left(1-\sqrt{\lambda_1(AA^*)}\right)\sum_{i=1}^{r}\sigma_i \leq E'_{\min} - E_{\min} \leq 2\left(1-\sqrt{\lambda_m(AA^*)}\right)\sum_{i=1}^{r}\sigma_i.$$

Thus

$$E'_{\min} \leq E_{\min}, \qquad \text{if } \lambda_m(AA^*) \geq 1$$

and

$$E'_{\min} \geq E_{\min}, \qquad \text{if } \lambda_1(AA^*) \leq 1.$$

In particular, if A is unitary then $E_{\min} = E'_{\min}$.

Proof: We rely on the following theorem due to Ostrowski (see, e.g., [20, p. 224]).

Ostrowski Theorem: Let A and S denote $m \times m$ matrices with S Hermitian and A nonsingular, and let $S' = ASA^*$. Let $\lambda_k(\cdot)$ denote the kth eigenvalue of the corresponding matrix, where the eigenvalues are arranged in decreasing order. For every $1 \leq i \leq m$, there exists a positive real number a_i such that $\lambda_m(AA^*) \leq a_i \leq \lambda_1(AA^*)$ and $\lambda_i(S') = a_i \lambda_i(S)$.

Combining this theorem with the expression (19) for the residual squared error results in

$$E'_{\min} - E_{\min} = 2\sum_{i=1}^{r}(1-\sqrt{a_i})\sigma_i.$$

Substituting $\lambda_m(AA^*) \leq a_i \leq \lambda_1(AA^*)$ results in Theorem 5. If A is unitary, then $AA^* = I$, and $\lambda_i(AA^*) = 1$ for all i.

Acknowledgment

The authors wish to thank A. S. Holevo and H. P. Yuen for helpful comments. The first author wishes to thank A. V. Oppenheim for his encouragement and support.

References

[1] A. Peres, *Quantum Theory: Concepts and Methods*. Boston, MA: Kluwer, 1995.

[2] A. S. Holevo, "Statistical decisions in quantum theory," *J. Multivar. Anal.*, vol. 3, pp. 337–394, Dec. 1973.

[3] H. P. Yuen, R. S. Kennedy, and M. Lax, "Optimum testing of multiple hypotheses in quantum detection theory," *IEEE Trans. Inform. Theory*, vol. IT-21, pp. 125–134, Mar. 1975.

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

[5] M. Charbit, C. Bendjaballah, and C. W. Helstrom, "Cutoff rate for the M-ary PSK modulation channel with optimal quantum detection," *IEEE Trans. Inform. Theory*, vol. 35, pp. 1131–1133, Sept. 1989.

[6] M. Osaki, M. Ban, and O. Hirota, "Derivation and physical interpretation of the optimum detection operators for coherent-state signals," *Phys. Rev. A*, vol. 54, pp. 1691–1701, Aug. 1996.

[7] M. Ban, K. Kurukowa, R. Momose, and O. Hirota, "Optimum measurements for discrimination among symmetric quantum states and parameter estimation," *Int. J. Theor. Phys.*, vol. 36, pp. 1269–1288, 1997.

[8] C. W. Helstrom, "Bayes-cost reduction algorithm in quantum hypothesis testing," *IEEE Trans. Inform. Theory*, vol. IT-28, pp. 359–366, Mar. 1982.

[9] P. Hausladen and W. K. Wootters, "A 'pretty good' measurement for distinguishing quantum states," *J. Mod. Opt.*, vol. 41, pp. 2385–2390, 1994.

[10] P. Hausladen, R. Josza, B. Schumacher, M. Westmoreland, and W. K. Wootters, "Classical information capacity of a quantum channel," *Phys. Rev. A*, vol. 54, pp. 1869–1876, Sept. 1996.

[11] A. Peres and W. K. Wootters, "Optimal detection of quantum information," *Phys. Rev. Lett.*, vol. 66, pp. 1119–1122, Mar. 1991.

[12] A. Peres, "Neumark's theorem and quantum inseparability," *Found. Phys.*, vol. 20, pp. 1441–1453, 1990.

[13] M. Sasaki, K. Kato, M. Izutsu, and O. Hirota, "Quantum channels showing superadditivity in classical capacity," *Phys. Rev. A*, vol. 58, pp. 146–158, July 1998.

[14] M. Sasaki, T. Sasaki-Usuda, M. Izutsu, and O. Hirota, "Realization of a collective decoding of code-word states," *Phys. Rev. A*, vol. 58, pp. 159–164, July 1998.

[15] K. Kato, M. Osaki, M. Sasaki, and O. Hirota, "Quantum detection and mutual information for QAM and PSK signals," *IEEE Trans. Commun.*, vol. 47, pp. 248–254, Feb. 1999.

[16] G. D. Forney, Jr., "Geometrically uniform codes," *IEEE Trans. Inform. Theory*, vol. 37, pp. 1241–1260, Sept. 1991.

[17] G. H. Golub and C. F. Van Loan, *Matrix Computations*. Baltimore, MD: Johns Hopkins Univ. Press, 1983.

[18] C. King and M. B. Ruskai, "Minimal entropy of states emerging from noisy quantum channels," *IEEE Trans. Inform. Theory*, to be published.

[19] M. A. Armstrong, *Groups and Symmetry*. New York: Springer-Verlag, 1988.

[20] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1985.