Optimum unambiguous discrimination between subsets of nonorthogonal quantum states

Yuqing Sun,[1] János A. Bergou,[1,2] and Mark Hillery[1]
[1]*Department of Physics, Hunter College, City University of New York, 695 Park Avenue, New York, New York 10021*
[2]*Institute of Physics, Janus Pannonius University, H-7624 Pécs, Ifjúság útja 6, Hungary*
(Received 4 December 2001; published 23 September 2002)

It is known that unambiguous discrimination among nonorthogonal but linearly independent quantum states is possible with a certain probability of success. Here, we consider a variant of that problem. Instead of discriminating among all of the different states, we shall only discriminate between two subsets of them. In particular, for the case of three nonorthogonal states, $\{|\psi_1\rangle,|\psi_2\rangle,|\psi_3\rangle\}$, we show that the optimal strategy to distinguish $|\psi_1\rangle$ from the set $\{|\psi_2\rangle,|\psi_3\rangle\}$ has a higher success rate than if we wish to discriminate among all three states. Somewhat surprisingly, for unambiguous discrimination the subsets need not be linearly independent. A fully analytical solution is presented, and we also show how to construct generalized interferometers (multiport) which provide an optical implementation of the optimal strategy.

DOI: 10.1103/PhysRevA.66.032315 PACS number(s): 03.67.−a, 03.65.Ta, 42.50.−p

I. INTRODUCTION

According to the quantum theory of measurement, it is impossible to unambiguously discriminate between nonorthogonal quantum states with unit success probability. If, however, we settle for less and do not require that we succeed every time, then unambiguous discrimination becomes possible. This procedure uses a nonunitary operation that maps the nonorthogonal states onto orthogonal ones, and these can then be discriminated without error using a standard von Neumann measurement. Although such an operation will always have a certain probability of failure, we can always tell whether or not the desired transformation has succeeded. This allows us to achieve unambiguous discrimination. When the attempt fails, we obtain an inconclusive answer. The optimal strategy for accomplishing this is the one that minimizes the average probability of failure.

The problem of unambiguously distinguishing between two nonorthogonal states was first considered by Ivanovic [1], and then subsequently by Dieks [2] and Peres [3]. These authors found the optimal solution when the two states are being selected from an ensemble in which they are equally likely. The optimal solution for the situation in which the states have different weights was found by Jaeger and Shimony [4]. We proposed an optical implementation of the optimal procedure along with a more compact rederivation of the general results and also showed that the method is useful in other areas of quantum information processing [5] such as, for example, entanglement enhancement [6]. State discrimination measurements have been performed in laboratory, first by Huttner *et al.* [7] and, more recently, by Clarke *et al.* [8]. Both used the polarization states of photons to represent qubits. The case of three states was examined by Peres and Terno [9]. It was subsequently extended to the general problem of discriminating among N states. Chefles [10] found that N nonorthogonal states can be probabilistically discriminated without error if and only if they are linearly independent. Chefles and Barnett [11] solved the case in which the probability of the procedure succeeding is the same for each of the states. Duan and Guo [12] considered general unitary transformations and measurements on a Hilbert space containing the states to be distinguished and an ancilla, which would allow one to discriminate among N states, and derived matrix inequalities which must be satisfied for the desired transformations to exist. In our previous paper [13], we presented the necessary conditions for optimal unambiguous discrimination and used them to derive a method for implementing the optimal solution. For the case of three states, we presented optical networks that accomplish this. One can also consider what happens if the discrimination is not completely unambiguous, i.e., if it is possible for errors to occur, and this was done by Chefles and Barnett [14]. For an overview of the state of the art on state discrimination see the excellent recent review by Chefles [15].

In these works discrimination among all of the states was considered. In the present paper, we consider a variant of that problem. Instead of discriminating among all states, we ask what happens if we just want to discriminate between subsets of them. A motivation to consider this variant comes from its application to comparing strings of qubits in order to find out if they are identical or not, which is certainly one of the basic tasks in quantum information processing. In particular, if there are three nonorthogonal states, $\{|\psi_1\rangle,|\psi_2\rangle,|\psi_3\rangle\}$, we wish to find the optimal strategy to unambiguously distinguish $|\psi_1\rangle$ from the set $\{|\psi_2\rangle,|\psi_3\rangle\}$. We refer to this problem as unambiguous quantum state filtering. In this context we should note that recently an analytical solution has been found to the following closely related problem. Instead of unambiguously distinguishing between two complementary subsets of an arbitary number N of nonorthogonal quantum states, occupying a two-dimensional Hilbert space, errors are allowed but the probability of erroneously assigning the state to one of the substes is minimized [16]. The term "quantum state filtering" has been introduced there for the case when one of the subsets contains one state and the other contains all of the remaining $N-1$ states. Here, we shall present the analytical solution for the case of the other possible discrimnation strategy, namely, that of unambiguous quantum state filtering.

The paper is divided into six sections. In Sec. II, based on simple but rigorous arguments, we present the optimal analytical solution to the problem. In Sec. III, we compare these optimal failure probabilities for two different procedures: discrimination between $|\psi_1\rangle$ and $\{|\psi_2\rangle,|\psi_3\rangle\}$ and discrimination among all three states. We find that the failure probability for the first procedure is smaller than that for the second. In Sec. IV, we propose a possible experimental implementation using the method proposed in our previous paper [13], which uses a single-photon representation of the quantum states and an optical multiport together with photon detection at the output ports to implement the procedure. A brief discussion and conclusions are given in Sec. V. Finally, in the Appendix, we present an alternative derivation, based on the method of Lagrange multipliers, to obtain the results of Sec. II. The method closely parallels the techniques used for unambiguous discrimination between all states.

II. DERIVATION OF THE OPTIMAL SOLUTION

Suppose we are given a quantum system prepared in the state $|\psi\rangle$, which is guaranteed to be a member of the set of three nonorthogonal states $\{|\psi_1\rangle,|\psi_2\rangle,|\psi_3\rangle\}$, but we do not know which one. We want to find a procedure which will tell us that $|\psi\rangle$ was prepared in $|\psi_1\rangle$, or will tell us that $|\psi\rangle$ was prepared in one of $\{|\psi_2\rangle,|\psi_3\rangle\}$. That is, the procedure can distinguish $|\psi_1\rangle$ from $\{|\psi_2\rangle,|\psi_3\rangle\}$. We also want this procedure to be error free, i.e., the procedure may fail to give us any information about the state, and if it fails, it must let us know that it has, but if it succeeds, it should never give us a wrong answer. We shall refer to such a procedure as quantum state filtering without error. We find that, in contrast to the unambiguous state discrimination problem, this will be possible even if $|\psi_1\rangle$ is not linearly independent from the set $\{|\psi_2\rangle,|\psi_3\rangle\}$.

If the states are not orthogonal then, according to the quantum theory of measurement, they cannot be discriminated perfectly. In other words, if we are given $|\psi_i\rangle$, we will have some probability p_i to determine what it is successfully and, correspondingly, some failure probability, $q_i=1-p_i$, to obtain an inconclusive answer. If we denote by η_i the *a priori* probability that the system was prepared in the state $|\psi_i\rangle$, the average probabilities of success and of failure to distinguish the states $|\psi_i\rangle$ are

$$P=\sum_i \eta_i p_i,$$

$$Q=\sum_i \eta_i q_i, \tag{2.1}$$

respectively. Our objective is to find the set of $\{p_i\}$ that maximizes the probability of success, P.

The procedure we shall use is a "generalized measurement," which can be described as follows. Let $\mathcal{K}$ denote a total Hilbert space, which is the direct sum of two subspaces, $\mathcal{K}=\mathcal{H}\oplus\mathcal{A}$. The space $\mathcal{H}$ is a three-dimensional space that contains the vectors $|\psi_i\rangle$, and $\mathcal{A}$ is an auxiliary space. The input state of the system is one of the vectors $|\psi_i\rangle$, which is now a vector in the subspace $\mathcal{H}$ of the total space $\mathcal{K}$, so that

$$|\psi_i^{\mathcal{K}}\rangle_{in}=|\psi_i^{\mathcal{H}}\rangle. \tag{2.2}$$

A unitary transformation, U, which acts in the entire space $\mathcal{K}$ is now applied to the input vector, resulting in the state $|\psi_i^{\mathcal{K}}\rangle_{out}$, which is given by

$$|\psi_i^{\mathcal{K}}\rangle_{out}=|\psi_i'^{\mathcal{H}}\rangle+|\phi_i^{\mathcal{A}}\rangle=U|\psi_i^{\mathcal{K}}\rangle_{in}, \tag{2.3}$$

where, in our case, $|\psi_1'\rangle$ can always be unambiguously distinguished from the set $\{|\psi_2'\rangle,|\psi_3'\rangle\}$. Then a measurement is performed on $|\psi_i^{\mathcal{K}}\rangle_{out}$ that projects $|\psi_i^{\mathcal{K}}\rangle_{out}$ either onto $|\psi_i'\rangle$ or $|\phi_i\rangle$ (by construction, they are in orthogonal subspaces). If it projects $|\psi_i^{\mathcal{K}}\rangle_{out}$ onto $|\psi_i'\rangle$, the procedure succeeds, because $|\psi_1'\rangle$ can always be distinguished from $\{|\psi_2'\rangle,|\psi_3'\rangle\}$. The probability to get this outcome, if the input state is $|\psi_i\rangle$, is

$$p_i=\langle\psi_i'|\psi_i'\rangle. \tag{2.4}$$

If the measurement projects $|\psi_i^{\mathcal{K}}\rangle_{out}$ onto $|\phi_i\rangle$, the procedure fails. The probability of this outcome is

$$q_i=1-p_i=\langle\phi_i|\phi_i\rangle. \tag{2.5}$$

The nature of the problem we are trying to solve imposes a number of requirements on the output vectors. The condition that $|\psi_1'\rangle$ be distinguishable from $|\psi_2'\rangle$ and $|\psi_3'\rangle$ requires that

$$\langle\psi_1'|\psi_2'\rangle=\langle\psi_1'|\psi_3'\rangle=0. \tag{2.6}$$

These lead to conditions on the failure vectors, $|\phi_i\rangle$. Taking the scalar product between $|\psi_1^{\mathcal{K}}\rangle_{out}$ and the other two output states and using Eq. (2.6) and the fact that U is unitary leads to the conditions

$$\langle\phi_1|\phi_2\rangle=\langle\psi_1|\psi_2\rangle,$$

$$\langle\phi_1|\phi_3\rangle=\langle\psi_1|\psi_3\rangle. \tag{2.7}$$

Our objective is to find the optimal $|\psi_i'\rangle$ and $|\phi_i\rangle$ which satisfy Eqs. (2.4)–(2.7) and also give the maximum success probability P.

Let us now consider the failure vectors. If they were linearly independent, we could apply a state discrimination procedure to them [10]. That means that if our original procedure fails, and we end up in the failure space, $\mathcal{A}$, then we still have some chance of determining what our input state was. This clearly implies that our original procedure, which led to the vectors $|\psi'\rangle$, was not optimal, because that process followed by another on the failure vectors would lead to a higher probability of distinguishing $|\psi_1\rangle$ from $|\psi_2\rangle$ and $|\psi_3\rangle$. Therefore, the optimal procedure should lead to failure vectors to which we cannot successfully apply a state discrimination procedure, implying that they are linearly dependent. In fact, we will now prove that for optimal discrimination they must be collinear, by demonstrating that the contrary leads to contradiction. To this end, we assume that we have achieved optimal unambiguous discrimination of $|\psi_1\rangle$ from $|\psi_2\rangle$ and $|\psi_3\rangle$ but the failure vectors are *not* collinear. Then at least one of the two failure vectors, $|\phi_2\rangle,|\phi_3\rangle$, will have a component in the direction that is perpendicular to $|\phi_1\rangle$. We can set up a detector projecting onto this direction and a positive outcome of the measurement (a click of the detector) will tell us that our input state was not $|\psi_1\rangle$ but one of the other two states. Thus, contrary to our assumption that our procedure has been optimal, further distinction is possible. Hence, the failure vectors must be collinear for optimal discrimination.

We shall now explore the consequences of this conclusion. Since $|\phi_i\rangle(i=1,\dots,n)$ are collinear, the failure space, $\mathcal{A}$, is one dimensional. If $|u\rangle$ is the basis vector spanning this Hilbert space we can write the failure vectors as $|\phi_i\rangle=\sqrt{q_i}e^{i\chi_i}|u\rangle$. Substituting this representation of the failure vectors into Eq. (2.7), we find that

$$q_1q_2=|\langle\psi_1|\psi_2\rangle|^2,$$
$$q_1q_3=|\langle\psi_1|\psi_3\rangle|^2. \tag{2.8}$$

These two conditions are a consequence of unitarity and imply that only one of the three failure probabilities can be chosen independently. If we chose q_1 as the independent one we can express the other two as $q_2=|\langle\psi_1|\psi_2\rangle|^2/q_1$ and $q_3=|\langle\psi_1|\psi_3\rangle|^2/q_1$. If we introduce the notation $O_{ij}=\langle\psi_i|\psi_j\rangle$ then, with the help of these two equations, the average failure probability can be written explicitly as

$$Q=\sum_i \eta_i q_i$$
$$=\eta_1q_1+\frac{\eta_2|O_{12}|^2+\eta_3|O_{13}|^2}{q_1}. \tag{2.9}$$

If we further introduce the notation $A=\eta_2|O_{12}|^2+\eta_3|O_{13}|^2$ for the frequently occuring average overlap then, from the condition

$$\frac{dQ}{dq_1}=0, \tag{2.10}$$

we find the optimal value of q_1 to be

$$q_1=\sqrt{A/\eta_1}. \tag{2.11}$$

This value, however, cannot always be realized. For it to be true, there must be a unitary transformation, from Eq. (2.3), that takes $|\psi_j\rangle$ to $|\psi_j\rangle_{out}$ which, together with the one dimensionality of the failure space yields

$$|\psi_j\rangle_{out}=|\psi_j'\rangle+\sqrt{q_j}|e^{i\chi_j}|u\rangle. \tag{2.12}$$

Here we have that $\langle\psi_j'|u\rangle=0$, $\langle\psi_1'|\psi_j'\rangle=0$ for $j=2,3$, and the phase factors are fixed by the requirement [cf. Eq. (2.7)] that

$$\langle\psi_1|\psi_j\rangle=\sqrt{q_1q_j}e^{i(\chi_j-\chi_1)} \tag{2.13}$$

for $j=2,3$. These equations imply that

$$\langle\psi_j'|\psi_k'\rangle=\langle\psi_j|\psi_k\rangle-\sqrt{q_jq_k}e^{i(\chi_k-\chi_j)}. \tag{2.14}$$

This set of equations can only be true if the matrix M, where

$$M_{jk}=\langle\psi_j|\psi_k\rangle-\sqrt{q_jq_k}e^{i(\chi_k-\chi_j)}, \tag{2.15}$$

is positive semidefinite, as discussed in detail in Ref. [13].

Using again $O_{jk}=\langle\psi_j|\psi_k\rangle$, M can be expressed as

$$M=\begin{pmatrix} 1-q_1 & 0 & 0 \\ 0 & 1-\frac{|O_{12}|^2}{q_1} & O_{23}-\frac{O_{21}O_{13}}{q_1} \\ 0 & O_{32}-\frac{O_{31}O_{12}}{q_1} & 1-\frac{|O_{13}|^2}{q_1} \end{pmatrix}. \tag{2.16}$$

Clearly, this matrix will be positive semidefinite if $0\leqslant q_1\leqslant 1$, and if the 2×2 submatrix is also positive semidefinite. This will be true if both the trace and determinant of the submatrix are greater than or equal to zero. Positivity requires that the diagonal matrix elements of the submatrix be non-negative, so that it must be true that $q_1\geqslant|O_{12}|^2$ and $q_1\geqslant|O_{13}|^2$. Without loss of generality, we can assume that $|O_{12}|\geqslant|O_{13}|$ by simply arranging the states in set 2 in the order of decreasing overlaps with $|\psi_1\rangle$. Doing so and imposing the condition that $q_1\geqslant|O_{12}|^2$ guarantees that the condition $q_1\geqslant|O_{13}|^2$ is also satisfied, and together they imply that the trace is greater than or equal to zero.

The condition that the determinant be non-negative gives us a lower bound on q_1,

$$q_1\geqslant\frac{|O_{12}|^2+|O_{13}|^2-(O_{12}O_{23}O_{31}+O_{13}O_{32}O_{21})}{1-|O_{23}|^2}. \tag{2.17}$$

We want to interpret this inequality, in particular, we want to find what the right-hand side (rhs) is equal to. In order to do so, we shall find the projection operator, P_{23}, that projects onto the subspace spanned by $|\psi_2\rangle$ and $|\psi_3\rangle$. One of the basis vectors in this subspace can be chosen to be $|\psi_2\rangle$ and, using the Gram-Schmidt orthogonalization method, the other is defined as the (normalized) orthogonal component of $|\psi_3\rangle$,

$$|\tilde{\psi}_3\rangle=\frac{1}{\sqrt{1-|O_{23}|^2}}(|\psi_3\rangle-O_{23}|\psi_2\rangle). \tag{2.18}$$

This leads to

$$P_{23}=|\psi_2\rangle\langle\psi_2|+|\tilde{\psi}_3\rangle\langle\tilde{\psi}_3|. \tag{2.19}$$

Let us represent the input state, $|\psi_1\rangle$, as $|\psi_1\rangle=|\psi_1^{\perp}\rangle+|\psi_1^{\parallel}\rangle$, where $|\psi_1^{\perp}\rangle=(1-P_{23})|\psi_1\rangle$ is the component of the input vector that is perpendicular to the subspace spanned by $|\psi_2\rangle$ and $|\psi_3\rangle$ and $|\psi_1^{\parallel}\rangle=P_{23}|\psi_1\rangle$ is the component in that subspace. Then, using Eqs. (2.18) and (2.19), the explicit expression for the parallel component is given by

$$|\psi_1^{\parallel}\rangle=\frac{O_{21}-O_{23}O_{31}}{1-|O_{23}|^2}|\psi_2\rangle+\frac{O_{31}-O_{32}O_{21}}{1-|O_{23}|^2}|\psi_3\rangle]. \tag{2.20}$$

Calculating the norm of this expression yields

$$\langle \psi_1^{\parallel} | \psi_1^{\parallel} \rangle = \frac{|O_{12}|^2 + |O_{13}|^2 - (O_{12}O_{23}O_{31} + O_{13}O_{32}O_{21})}{1 - |O_{23}|^2}, \tag{2.21}$$

which is identical to the right-hand side of Eq. (2.17).

Thus, Eq. (2.17) tells us that the failure probability, q_1, has a lower bound which is given by the weight of $|\psi_1\rangle$ in the other subspace, $\|P_{23}\psi_1\|^2 = \langle \psi_1 | P_{23} | \psi_1 \rangle = \langle \psi_1^{\parallel} | \psi_1^{\parallel} \rangle$, a result that is intuitively obvious. Clearly, this expression is larger than (or at most equal to) $|O_{12}|^2$. This implies that, because $q_2 = |O_{12}|^2 / q_1$, we have

$$q_2 \leqslant \frac{|O_{12}|^2}{\langle \psi_1^{\parallel} | \psi_1^{\parallel} \rangle} = \frac{|O_{12}|^2}{|O_{12}|^2 + |\langle \tilde{\psi}_3 | \psi_1 \rangle|^2} \leqslant 1, \tag{2.22}$$

and similarly for q_3.

We can then distinguish three different regimes of the parameters. If the rhs of Eq. (2.11) is greater than 1 then $q_1 = 1$, if it is less than $\langle \psi_1^{\parallel} | \psi_1^{\parallel} \rangle$ then $q_1 = \langle \psi_1^{\parallel} | \psi_1^{\parallel} \rangle$, and in the intermediate range the optimum given by Eq. (2.11) is realized. This can be summarized as follows.

(i) If $\eta_1 |\langle \psi_1^{\parallel} | \psi_1^{\parallel} \rangle|^2 \leqslant A \leqslant \eta_1$, then

$$q_1 = \sqrt{A/\eta_1},$$
$$q_2 = \sqrt{\eta_1 / A}|O_{12}|^2,$$
$$q_3 = \sqrt{\eta_1 / A}|O_{13}|^2, \tag{2.23}$$

yielding the average failure probability

$$Q = 2\sqrt{\eta_1 A}. \tag{2.24}$$

(ii) If $A \geqslant \eta_1$, then

$$q_1 = 1,$$
$$q_2 = |O_{12}|^2,$$
$$q_3 = |O_{13}|^2. \tag{2.25}$$

yielding the average failure probability

$$Q = \eta_1 + A. \tag{2.26}$$

(iii) If $A \leqslant \eta_1 |\langle \psi_1^{\parallel} | \psi_1^{\parallel} \rangle|^2$, then

$$q_1 = \langle \psi_1^{\parallel} | \psi_1^{\parallel} \rangle,$$
$$q_2 = \frac{|O_{12}|^2}{\langle \psi_1^{\parallel} | \psi_1^{\parallel} \rangle},$$
$$q_3 = \frac{|O_{13}|^2}{\langle \psi_1^{\parallel} | \psi_1^{\parallel} \rangle}, \tag{2.27}$$

yielding the average failure probability

$$Q = \eta_1 \langle \psi_1^{\parallel} | \psi_1^{\parallel} \rangle + \frac{A}{\langle \psi_1^{\parallel} | \psi_1^{\parallel} \rangle}. \tag{2.28}$$

Equations (2.23)–(2.28) summarize our main results. In the intermediate range of the average overlap, A, the optimal failure probability, Eq. (2.24), is achieved by a generalized measurement or positive operator valued measurement (POVM). Outside this region, for very large average overlap, $A \geqslant \eta_1$, or very small average overlap, $A \leqslant \eta_1 |\langle \psi_1^{\parallel} | \psi_1^{\parallel} \rangle|^2$, the optimal failure probabilities, Eqs. (2.26) and (2.28), are realized by standard von Neumann measurements. For very large A the optimal von Neumann measurement consists of projections onto $|\psi_1\rangle$ and two orthogonal directions whose directionality needs not be specified further. A click along $|\psi_1\rangle$ corresponds to failure because it can have its origin in any of the two subsets and a click in the orthogonal directions uniquely assigns the input state to the set $\{|\psi_2\rangle, |\psi_3\rangle\}$. For very small A the optimal von Neumann measurement consists of projections onto $|\psi_1^{\parallel}\rangle$ and two orthogonal directions that are uniquely determined by the requirement that they correspond to two mutually exclusive alternatives. One of them is onto $|\psi_1^{\perp}\rangle$ and the other onto the remaining orthogonal direction in the subspace of $\{|\psi_2\rangle, |\psi_3\rangle\}$. A click along $|\psi_1^{\parallel}\rangle$ corresponds to failure because it can originate from any of the input states while a click in any of the alternative directions unambiguously assigns the input to one or the other of the two mutually exclusive subsets. It is interesting to observe that the failure space is one dimensional for each of the three different optimal measurements in the three different regions. At the boundaries of their respective regions of validity, the optimal measurements transform into one another continuously. Furthermore, each of the two von Neumann expressions can be written as the arithmetic mean of two terms and the POVM result as the geometric mean of the same two terms. Therefore, in its range of validity the POVM performs better than any von Neumann measurement.

In closing this section we want to point out an interesting feature of the solution. The results hold true even when there is no perpendicular component of the first input state, $|\psi_1^{\perp}\rangle = 0$, i.e., it lies entirely in the Hilbert space spanned by the other two vectors or, in other words, the two sets are linearly dependent. In this case the two von Neumann measurements coincide and the range of validity of the POVM solution shrinks to zero. A click in the detector along the first input vector corresponds to failure—it might originate from either of the two subsets—and a click in the detector along the single direction orthogonal to it unambiguously identifies the set of the other two vectors.

An alternative derivation of the above results, which is based on the method of Langrange multipliers, is given in the Appendix.

III. COMPARISON TO THE CASE WHEN ALL STATES ARE DISCRIMINATED

In this section we want to compare the average probability of failure Q of the filtering problem to that of distinguishing all three states. Let Q' denote the average probability of failure for distinguishing all the states $\{|\psi_1\rangle,|\psi_2\rangle,|\psi_3\rangle\}$. We can see immediately, that the probability of failure to distinguish $|\psi_1\rangle$ from $\{|\psi_2\rangle,|\psi_3\rangle\}$, Q, should be no larger than Q'. For the latter problem, the necessary condition for achieving optimal discrimination is

$$\begin{vmatrix} q_1 & O_{12} & O_{13} \\ O_{12}^* & q_2 & O_{23} \\ O_{13}^* & O_{23}^* & q_3 \end{vmatrix} = 0. \tag{3.1}$$

When comparing this equation to Eq. (A1), we see that, instead of a given constant O_{23} that appears in Eq. (3.1), there are the variables r and θ in Eq. (A1). These variables are chosen to minimize the average probability of failure Q. Therefore, Q should be no larger than Q', $Q \leqslant Q'$.

To illustrate this point, we use a simple symmetric case, where all of the overlaps between the states are real and equal,

$$\langle\psi_1|\psi_2\rangle = \langle\psi_1|\psi_3\rangle = \langle\psi_2|\psi_3\rangle = s, \tag{3.2}$$

with $0<s<1$. We shall also assume that the *a priori* probabilities are equal for all the examples in this paper. From previous work we know that in this case, the optimal values of the failure probabilities when we wish to distinguish among all of the states $\{|\psi_1\rangle,|\psi_2\rangle,|\psi_3\rangle\}$ are $q_i = s$, which implies that $Q' = s$ [13].

For the problem of distinguishing $|\psi_1\rangle$ from $\{|\psi_2\rangle,|\psi_3\rangle\}$, from the results of Eqs. (2.23) and (2.25), we have (i) if $0 < s \leqslant \sqrt{2}/2$, then

$$q_1 = \sqrt{2}s,$$

$$q_2 = q_3 = \frac{\sqrt{2}}{2}s,$$

$$Q = \frac{2\sqrt{2}}{3}s. \tag{3.3}$$

So the average probability of failure Q is less than $Q' = s$; (ii) if $\sqrt{2}/2 < s < 1$, then

$$q_1 = 1,$$

$$q_2 = q_3 = s^2,$$

$$Q = \frac{1}{3} + \frac{2}{3}s^2. \tag{3.4}$$

These solutions are illustrated and compared to Q' in Fig. 1. Note that in both cases we have that $Q < s = Q'$.

Now we shall compare filtering to the problem of distinguishing two states $\{|\psi_1\rangle,|\psi_2\rangle\}$, when all the *a priori* probabilities are equal. If we denote by Q'' the average probability of failure when distinguishing between the two states $\{|\psi_1\rangle$ and $|\psi_2\rangle\}$, we know that $Q'' = |O_{12}|$ (Refs. [1–4]). For the case we are considering, $|O_{12}| = |O_{13}| = s$, and we see that $Q < Q''$.

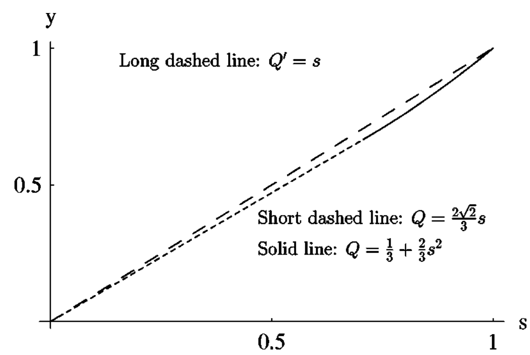

FIG. 1. We compare Q and Q'. For $0<s\leqslant\sqrt{2}/2$ we have that $Q'=s$ and $Q=(2\sqrt{2}/3)s$. For $\sqrt{2}/2<s\leqslant 1$, we still have that $Q'=s$, but $Q=\frac{1}{3}+\frac{2}{3}s^2$. Note that Q is always smaller than Q'.

A second example is more illuminating. The overlaps are now given by

$$\langle\psi_1|\psi_2\rangle = \langle\psi_1|\psi_3\rangle = s_1,$$

$$\langle\psi_2|\psi_3\rangle = s_2, \tag{3.5}$$

where, for simplicity, s_1 and s_2 are real, $0<s_1,s_2<1$, and

$$0 < s_1 < \frac{\sqrt{2}}{2},\ s_1^2 < s_2,\ \text{and}\ s_1 < 2s_2. \tag{3.6}$$

The probabilities of failure for discriminating $|\psi_1\rangle$ from $\{|\psi_2\rangle,|\psi_3\rangle\}$ are

$$q_1 = \sqrt{2}s_1,$$

$$q_2 = q_3 = \frac{\sqrt{2}}{2}s_1, \tag{3.7}$$

and the average failure probability is

$$Q = \frac{2\sqrt{2}}{3}s_1. \tag{3.8}$$

The optimal probabilities of failure for discriminating among all three states $\{|\psi_1\rangle,|\psi_2\rangle,|\psi_3\rangle\}$ are given by [13]

$$q_1' = \frac{s_1^2}{s_2},$$

$$q_2' = q_3' = s_2,$$

$$Q' = \frac{1}{3}[(s_1^2/s_2) + 2s_2]. \tag{3.9}$$

Q can be compared to Q' by examining the ratio

$$\frac{Q}{Q'} = \frac{2\sqrt{2}s_1 s_2}{s_1^2 + 2s_2^2} \leqslant 1. \tag{3.10}$$

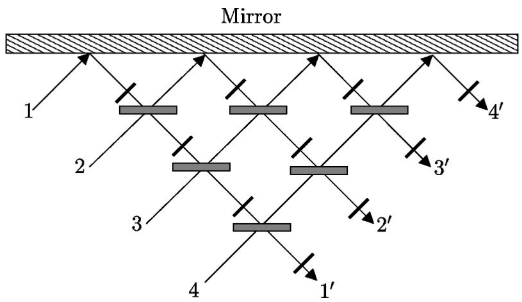

FIG. 2. An optical eight-port interferometer. The beams are straight lines, a suitable beam splitter is placed at each point where two beams intersect, phase shifters are at one input of each beam splitter and at each output.

From the above equation, we see that when s_1 is much smaller than s_2, Q is much smaller than Q'. For example, when $s_1=\sqrt{2}/5$, $s_2=\frac{4}{5}$, $Q/Q'=0.47$.

IV. OPTICAL REALIZATION

Now we shall present a scheme for a possible experimental realization of the optimal discrimination between $|\psi_1\rangle$ and $\{|\psi_2\rangle,|\psi_3\rangle\}$. The method is similar to the one we proposed in a previous publication [13]. We shall use single-photon states to represent the input and output states, and an optical eight-port interferometer together with photon detectors placed at the output ports to realize the unitary transformation and subsequent measurements.

Our states will be a single-photon split among several modes. Each mode will serve as an input to an optical eight-port interferometer. Recall that the dimension of the total Hilbert space is four, so we shall require four modes, and the input states $|\psi_i\rangle$ will be represented by single-photon states as

$$|\psi_i\rangle=\sum_{j=1}^{4} d_{ij}\hat{a}_j^{\dagger}|0\rangle, \tag{4.1}$$

where $\Sigma_{j=1}^4|d_{ij}|^2=1$, and $\hat{a}_j^{\dagger}$ is the creation operator for the jth mode. We shall require $d_{i4}=0$ for $i=1,2,3$, that is, the initial single-photon state is sent to the first three input ports, and the vacuum into the fourth input port. The first three modes correspond to the space, $\mathcal{H}$, containing the states to be distinguished and the fourth mode to the failure space, $\mathcal{A}$.

In general, an optical $2N$-port interferometer is a lossless linear device with N input ports and N output ports. Its action on the input states can be described by a unitary operator, U_{2N}, and physically it consists of an arrangement of beam splitters, phase shifters, and mirrors. Since the dimension of the input and output states is four, here we shall use an eight-port interferometer (see Fig. 2). If we denote the annihilation operators corresponding to the input modes of the eight-port interferometer by a_j, $j=1,\ldots,4$, then the output operators are given by

$$a_{j\,out}=U^{-1}a_jU=\sum_{k=1}^{4} M_{jk}a_k, \tag{4.2}$$

where M_{jk} are the elements of a 4×4 unitary matrix $M(4)$. In the Schrödinger picture, the *in* and *out* states are related by

$$|\psi\rangle_{out}=U|\psi\rangle_{in}. \tag{4.3}$$

It can be shown [13] that when using single-photon states representation, the matrix element M_{il} is the same as the matrix element of U between the single-particle states $|i\rangle=a_i^{\dagger}|0\rangle$ and $|l\rangle=a_l^{\dagger}|0\rangle$, i.e.,

$$\langle i|U|l\rangle=M_{il}. \tag{4.4}$$

To design the desired eight-port interferometer, we first calculate the optimal value of q_i. Then from Eq. (2.5) and the fact that our failure space is one dimensional, the vectors $|\phi_i\rangle$ are given by

$$|\phi_i\rangle=\sqrt{q_i}|1^{\mathcal{A}}\rangle=\sqrt{q_i}a_4^{\dagger}|0\rangle, \tag{4.5}$$

where the state $|1^{\mathcal{A}}\rangle$ denotes one-photon state in the failure space, which is just one photon in mode 4. Once the vectors $|\phi_i\rangle$ are determined, the inner products $\langle\psi_i'|\psi_j'\rangle$ $(i,j=1,2,3)$ are given by

$$\langle\psi_i'|\psi_j'\rangle=\langle\psi_i|\psi_j\rangle_{in}-\langle\phi_i|\phi_j\rangle. \tag{4.6}$$

We then have to find vectors $|\psi_i'\rangle$ that satisfy this equation. The answer is not unique, and one way of proceeding is the following. If we define the Hermitian matrix L to be

$$L_{ij}=\langle\psi_i|\psi_j\rangle_{in}-\langle\phi_i|\phi_j\rangle, \tag{4.7}$$

then we note from Eq. (2.7) that $L_{12}=L_{13}=0$. This implies that the simplest choice for $|\psi_1'\rangle$ is a vector with only one nonzero component. Then the vectors $|\psi_2'\rangle$ and $|\psi_3'\rangle$ will have nonzero components in only their other two places. The obvious choice is

$$|\psi_1'\rangle=\begin{pmatrix}\sqrt{p_1}\\0\\0\\0\end{pmatrix}. \tag{4.8}$$

In this column vector, the first entry is the amplitude of the photon to be in mode 1, the second is the amplitude to be in mode 2, etc. Mode 4 corresponds to the failure space, $\mathcal{A}$. The vectors $|\psi_2'\rangle$ and $|\psi_3'\rangle$ will have nonzero components in only their second and third places, and if their overlap is real, we can choose

$$|\psi_2'\rangle=\begin{pmatrix}0\\\sqrt{p_2}\cos\theta\\\sqrt{p_2}\sin\theta\\0\end{pmatrix},\quad |\psi_3'\rangle=\begin{pmatrix}0\\\sqrt{p_3}\cos\theta\\-\sqrt{p_3}\sin\theta\\0\end{pmatrix}, \tag{4.9}$$

where

$$\theta = \frac{1}{2}\cos^{-1}\left(\frac{L_{23}}{\sqrt{p_2 p_3}}\right). \tag{4.10}$$

This simple choice works for the last example in this section [see Eq. (4.22), below]. For the first, somewhat more general, example we are forced to choose the second component of $|\psi_1'\rangle$ to be nonzero and then the first and third components of the other two success vectors are different from zero. They can be obtained by simply interchanging the first and second components in the above expressions of the vectors $|\psi_i'\rangle$ [see Eq. (4.13), below].

Once we have the input and output vectors, the unitary transformation, U, which maps the input states onto the output states then can be chosen, and this, as shown by Eq. (4.4), gives the explicit form of $M(4)$. Furthermore, $M(4)$ can be factorized as a product of two-dimensional $U(2)$ transformations [13,17], and any $U(2)$ transformations can be implemented by a lossless beam splitter and a phase shifter with appropriate parameters. A beam splitter with a phase shifter at one output port transforms the input operators into output operators as

$$\begin{pmatrix} a_1 \\ a_2 \end{pmatrix}_{out} = \begin{pmatrix} e^{i\phi}\sin\omega & e^{i\phi}\cos\omega \\ \cos\omega & -\sin\omega \end{pmatrix}\begin{pmatrix} a_1 \\ a_2 \end{pmatrix}_{in}, \tag{4.11}$$

where a_1, a_2 are the annihilation operators of modes 1 and 2, respectively, ω describes the reflectivity and transmittance of the beam splitter, and ϕ describes the effect of the phase shifter (in the factorization method given by Reck *et al.* [17], the phase shifters described by ϕ should be placed at the input ports). Therefore, we can use appropriate beam splitters, phase shifters, and a mirror to construct the desired eight-port interferometer.

Finally, photon detection is performed at the four output ports. We can design the total transformation in such a way that if the photon is detected at the first output port, we claim with certainty that the initial state was $|\psi_1\rangle$, if the photon is detected at the second or the third output port, we claim with certainty that the initial state was either $|\psi_2\rangle$ or $|\psi_3\rangle$, but we do not know which of these two states it was. If the photon is detected at the fourth output port, we obtain no information about the input state.

We shall now consider two examples. The first is more general than the second, but the second has the advantage that it is simple and the eight-port interferometer that it requires consists of only two 50-50 beam splitters. In the first example, all of the input vectors have the same overlap, which is given by s, and we shall consider the case $0 < s \leqslant 1/\sqrt{2}$. The optimal failure probabilities for this case are given in Eq. (3.3). For the input vectors we shall take

$$|\psi_1\rangle_{in} = \begin{pmatrix} \frac{1}{\sqrt{3}}(1+2s)^{1/2} \\ \sqrt{\frac{2}{3}}(1-s)^{1/2} \\ 0 \\ 0 \end{pmatrix},$$

$$|\psi_2\rangle_{in} = \begin{pmatrix} \frac{1}{\sqrt{3}}(1+2s)^{1/2} \\ -\frac{1}{\sqrt{6}}(1-s)^{1/2} \\ \frac{1}{\sqrt{2}}(1-s)^{1/2} \\ 0 \end{pmatrix},$$

$$|\psi_3\rangle_{in} = \begin{pmatrix} \frac{1}{\sqrt{3}}(1+2s)^{1/2} \\ -\frac{1}{\sqrt{6}}(1-s)^{1/2} \\ -\frac{1}{\sqrt{2}}(1-s)^{1/2} \\ 0 \end{pmatrix}. \tag{4.12}$$

The output vectors, $|\psi_i\rangle_{out} = |\psi_i'\rangle + |\phi_i\rangle$, can be computed by the method outlined above. Doing so gives us

$$|\psi_1\rangle_{out} = \begin{pmatrix} 0 \\ (1-\sqrt{2}s)^{1/2} \\ 0 \\ (s\sqrt{2})^{1/2} \end{pmatrix},$$

$$|\psi_2\rangle_{out} = \begin{pmatrix} [(1+s-s\sqrt{2})/2]^{1/2} \\ 0 \\ [(1-s)/2]^{1/2} \\ (s/\sqrt{2})^{1/2} \end{pmatrix},$$

$$|\psi_3\rangle_{out} = \begin{pmatrix} [(1+s-s\sqrt{2})/2]^{1/2} \\ 0 \\ -[(1-s)/2]^{1/2} \\ (s/\sqrt{2})^{1/2} \end{pmatrix}. \tag{4.13}$$

Our next step is to determine the transformation U that describes the eight-port interferometer, or, more specifically, the matrix $M(4)$ that describes its action in the one-photon subspace. It must satisfy $|\psi_i\rangle_{out} = U|\psi\rangle_{in}$, and, in addition, it must map the vector that is orthogonal to all three input vectors, onto the vector that is orthogonal to all three output vectors,

$$\frac{1}{A}\begin{pmatrix} -(s\sqrt{2})^{1/2}B \\ -(s\sqrt{2})^{1/2}C \\ 0 \\ BC \end{pmatrix} = M(4)\begin{pmatrix} 0\\0\\0\\1 \end{pmatrix}, \tag{4.14}$$

where

$$A=[(1-s)(1+2s)]^{1/2},$$

$$B=(1-s\sqrt{2})^{1/2},$$

$$C=(1+s-s\sqrt{2})^{1/2}. \tag{4.15}$$

These equations determine $M(4)$ and it is given by $M(4)=$

$$\begin{pmatrix} \sqrt{\frac{2}{3}}\frac{C}{\sqrt{1+2s}} & -\frac{C}{\sqrt{3(1-s)}} & 0 & -\frac{B}{A}(s\sqrt{2})^{1/2} \\ \frac{B}{\sqrt{3(1+2s)}} & \sqrt{\frac{2}{3}}\frac{B}{\sqrt{1-s}} & 0 & -\frac{C}{A}(s\sqrt{2})^{1/2} \\ 0 & 0 & 1 & 0 \\ \frac{(\sqrt{2}+1)(s\sqrt{2})^{1/2}}{\sqrt{3(1+2s)}} & \frac{(\sqrt{2}-1)(s\sqrt{2})^{1/2}}{\sqrt{3(1-s)}} & 0 & \frac{BC}{A} \end{pmatrix}. \tag{4.16}$$

This matrix can be expressed as the product of three matrices, each of which corresponds to a beam splitter. In particular, we have that

$$M(4)=T_{2,4}T_{1,4}T_{1,2}, \tag{4.17}$$

where the matrix $T_{p,q}$ represents the action of a beam splitter that mixes only modes p and q. The 4×4 matrix for $T_{p,q}$ can be obtained from that of a 4×4 identity matrix, I, by replacing the matrix elements I_{pp} and I_{qq} by the transmissivity of the beam splitter, t, replacing I_{pq} by the reflectivity, r, and replacing I_{qp} by $-r$. The transmissivities and reflectivities for beam splitters in Eq. (4.17) are

$$T_{2,4}:\ \ t=B,\ \ r=-(s\sqrt{2})^{1/2},$$

$$T_{1,4}:\ \ t=\frac{C}{A},\ \ r=-(s\sqrt{2})^{1/2}\frac{B}{A}, \tag{4.18}$$

$$T_{1,2}:\ \ t=\sqrt{\frac{2(1-s)}{3}},\ \ r=-\sqrt{\frac{1+2s}{3}}.$$

This constitutes a complete description of the optical network that optimally discriminates between $|\psi_1\rangle_{in}$ and $\{|\psi_2\rangle_{in},|\psi_3\rangle_{in}\}$, where these input states are given in Eq. (4.12), and it is shown schematically in Fig. 3.

An especially simple network will suffice for our second example. The input vectors are

$$|\psi_1\rangle_{in}=\begin{pmatrix}\sqrt{2/3}\\0\\1/\sqrt{3}\\0\end{pmatrix},$$

$$|\psi_2\rangle_{in}=\begin{pmatrix}0\\1/\sqrt{3}\\\sqrt{2/3}\\0\end{pmatrix},$$

$$|\psi_3\rangle_{in}=\begin{pmatrix}0\\-1/\sqrt{3}\\\sqrt{2/3}\\0\end{pmatrix}. \tag{4.19}$$

These input states have the property that

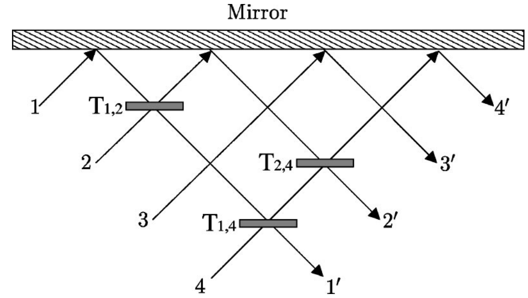

FIG. 3. The eight-port interferometer described by Eq. (4.16) can be constructed from three beam splitters and a mirror.

$$ {}_{in}\langle\psi_1|\psi_2\rangle_{in} = {}_{in}\langle\psi_1|\psi_3\rangle_{in} = \frac{\sqrt{2}}{3}, $$

$$ {}_{in}\langle\psi_2|\psi_3\rangle_{in} = \frac{1}{3}. \tag{4.20} $$

The optimal failure probabilities are found to be $q_1 = 2/3$ and $q_2 = q_3 = 1/3$. Using Eqs. (3.7) and (3.8) this gives

$$ Q = \frac{4}{9}, \tag{4.21} $$

for the minimum average failure probability of this kind of generalized measurement. This is to be compared to 13/27, the average failure probability of a von Neumann type projective measurement, from Eq. (2.26).

The output vectors, $|\psi_i\rangle_{out} = |\psi_i'\rangle + |\phi_i\rangle$, can again be computed by the method outlined previously. Doing so gives us

$$ |\psi_1\rangle_{out} = \begin{pmatrix} 1/\sqrt{3} \\ 0 \\ 0 \\ \sqrt{2/3} \end{pmatrix}, $$

$$ |\psi_2\rangle_{out} = \begin{pmatrix} 0 \\ 1/\sqrt{3} \\ 1/\sqrt{3} \\ 1/\sqrt{3} \end{pmatrix}, $$

$$ |\psi_3\rangle_{out} = \begin{pmatrix} 0 \\ -1/\sqrt{3} \\ 1/\sqrt{3} \\ (1/\sqrt{3} \end{pmatrix}. \tag{4.22} $$

The matrix $M(4)$ can be chosen to be

$$ M(4) = \begin{pmatrix} 1/\sqrt{2} & 0 & 0 & -1/\sqrt{2} \\ 0 & 1 & 0 & 0 \\ -1/2 & 0 & 1/\sqrt{2} & -1/2 \\ 1/2 & 0 & 1/\sqrt{2} & 1/2 \end{pmatrix}, \tag{4.23} $$

and it can be expressed as

$$ M(4) = T_{3,4} T_{1,4}. \tag{4.24} $$

In this case, both $T_{1,4}$ and $T_{3,4}$ represent 50-50 beam splitters, and they are given explicitly by

$$ T_{1,4}: \quad t = \frac{1}{\sqrt{2}}, \quad r = -\frac{1}{\sqrt{2}}, $$

$$ T_{3,4}: \quad t = \frac{1}{\sqrt{2}}, \quad r = -\frac{1}{\sqrt{2}}. \tag{4.25} $$

This last example constitutes what is probably the simplest choice of the set of parameters for a possible experimental realization.

V. CONCLUSIONS

The usual problem considered when trying to unambiguously discriminate among quantum states is to correctly identify which state a given system is in when one knows the set of possible states in which it can be prepared. Here we have considered a different problem. The set of possible states is divided into two subsets, and we only want to know to which subset the quantum state of our given system belongs. As this is a less ambitious task than actually identifying the state, we expect that our probability to be successful will be greater for attaining this more limited goal.

We considered the simplest instance of this problem, the situation in which we are trying to discriminate between a set containing one quantum state and another containing two. A method for finding the optimal strategy for discriminating between these two sets was presented, and analytical solutions for particular cases were given. In addition, we have shown that if the quantum states are single-photon states, where the photon can be split among several modes, the optimal discrimination strategy can be implemented by using a linear optical network.

These ideas can be extended in a number of different ways. One possibility is to consider the situation in which one is given N qubits, each of which is in either the state $|\psi_1\rangle$ of $|\psi_2\rangle$, where these states are not orthogonal. What we would like to know is how many of the qubits are in the state $|\psi_1\rangle$. In order to phrase this problem in a way that makes its connection to the problems considered in this paper clear, we note that the total set of possible states for this problem consists of 2^N states (the states are strings of N qubits), and this can be divided up into the subsets S_n, where the members of S_n are sequences of N qubits in which n are in the state $|\psi_1\rangle$. For a given sequence of qubits, our problem is to determine to which of the sets S_n it belongs. Another possibility is to use these methods to compare strings of qubits in order to find out if they are identical or not. Again, suppose that we have strings of N qubits in which each qubit is in one of the two nonorthogonal states, $|\psi_1\rangle$ or $|\psi_2\rangle$. We are given two of these strings and want to know if they are the same or not. In this case, our set of possible states consists of pairs of strings, and hence has 2^{2N} members. This is divided into two subsets, the first, S_{equal}, consisting of pairs of identical N-qubit strings (2^N members), and its complement, $\overline{S}_{equal}$, consisting of everything else. Our task, when given two sequences of N qubits, is to decide if they are in S_{equal} or in $\overline{S}_{equal}$ [18]. More detailed consideration of these problems remains for future research.

ACKNOWLEDGMENTS

We want to thank S. Barnett, A. Chefles, U. Herzog, and I. Jex for helpful discussions of various aspects of this problem. This research was supported by the Office of Naval Research (Grant No. N00014-92J-1233), by the National Science Foundation (Grant No. PHY-9970507), by the Hungarian Science Research Fund (OTKA, Grant No. T 030671), and by a grant from PSC-CUNY as well as by a CUNY collaborative grant.

APPENDIX: DERIVATION OF THE OPTIMAL SOLUTION VIA THE METHOD OF LAGRANGE MULTIPLIERS

In this appendix, we shall show that by using the method of Lagrange multipliers, we can derive the conclusions contained in Eqs. (2.23)–(2.28) rigorously, starting from the fact that for optimal discrimination, the vectors $|\phi_i\rangle$ must be linearly dependent. To express this statement in a compact form we define the positive semidefinite matrix C, where $C_{ij} = \langle \phi_i|\phi_j\rangle$. Then, in general, if $|\phi_i\rangle (i=1,\dots,n)$ are linearly dependent, the determinant of matrix C must vanish, $\Delta = \det(C)=0$ [13]. With the help of Eqs. (2.5) and (2.7), we can eliminate two of the three overlaps from the matrix C and obtain explicitly

$$\Delta = \begin{vmatrix} q_1 & O_{12} & O_{13} \\ O_{12}^* & q_2 & re^{i\theta} \\ O_{13}^* & re^{-i\theta} & q_3 \end{vmatrix} = q_1q_2q_3 - r^2q_1 - |O_{13}|^2q_2 - |O_{12}|^2q_3 + 2|O_{12}||O_{13}|r\cos(\theta-\alpha)=0. \tag{A1}$$

Here O_{ij} again denotes $\langle\psi_i|\psi_j\rangle$, $re^{i\theta}=\langle\phi_2|\phi_3\rangle$ is the remaining overlap where r and θ are to be determined from the conditions for optimum, and $\alpha = -\arg(O_{12}O_{13}^*)$. Since C is positive semidefinite, all the diagonal subdeterminants of Δ must be non-negative.

We now wish to minimize the average probability of failure Q, Eq. (2.1), subject to the constraint in Eq. (A1). This can be done by minimizing the quantity

$$\tilde{Q} = \sum_i^3 \eta_i q_i + \lambda\Delta, \tag{A2}$$

where λ is a Lagrange multiplier. The conditions for minimum with respect to r and θ, $\partial\tilde{Q}/\partial r=0$ and $\partial\tilde{Q}/\partial\theta=0$, lead immediately to

$$|O_{12}||O_{13}|\cos(\theta-\alpha) - q_1 r = 0, \tag{A3}$$

$$r|O_{12}||O_{13}|\sin(\theta-\alpha)=0. \tag{A4}$$

The solutions of these equations, corresponding to the minimum of Q, are

$$\theta = \alpha \tag{A5}$$

and

$$q_1 r = |O_{12}||O_{13}|. \tag{A6}$$

Next, we perform the optimization with respect to the remaining variables. Notice that the derivative of $\tilde{Q}$ with respect to λ returns Eq. (A1). Therefore, we use the optimal values of r and θ in Eq. (A1) and in the conditions for minimum with respect to the failure probabilities, $\partial\tilde{Q}/\partial q_i = 0$ for $i=1,2,3$. After some algebra we obtain the following set of equations:

$$q_1\Delta = \Delta_{12}\Delta_{13} = 0, \tag{A7}$$

$$q_1^2\frac{\partial\tilde{Q}}{\partial q_1} = \eta_1 q_1^2 + \lambda(\Delta_{12}\Delta_{13} + |O_{12}|^2\Delta_{13} + |O_{13}|^2\Delta_{12}) = 0, \tag{A8}$$

$$\frac{\partial\tilde{Q}}{\partial q_2} = \eta_2 + \lambda\Delta_{13} = 0, \tag{A9}$$

$$\frac{\partial\tilde{Q}}{\partial q_3} = \eta_3 + \lambda\Delta_{12} = 0, \tag{A10}$$

where Δ_{12} and Δ_{13} are the diagonal subdeterminants of Δ,

$$\Delta_{12} = q_1q_2 - |O_{12}|^2, \tag{A11}$$

$$\Delta_{13} = q_1q_3 - |O_{13}|^2. \tag{A12}$$

We now have four variables q_1, q_2, q_3, and λ, and four equations, Eqs. (A7)–(A10), to find them. Equation (A7) tells us that at least one of the diagonal subdeterminants vanishes. With no loss of generality we can assume this to be $\Delta_{12}=0$. Comparing this to Eq. (A10) we see that λ must be singular. The singularity, however, is tractable since the same equation tells us that the product $\lambda\Delta_{12}$ is finite. Then it follows from the singular behavior of λ and Eq. (A9) that the other diagonal subdeterminant also vanishes, $\Delta_{13}=0$, but the product $\lambda\Delta_{12}$ also remains finite. Using these finite values from Eqs. (A9)–(A10) in Eq. (A8), we can summarize our findings as follows:

$$\Delta_{12} = \Delta_{13} = 0, \tag{A13}$$

which is just Eq. (2.8), and

$$\eta_1 q_1^2 - \eta_2|O_{12}|^2 - \eta_3|O_{13}|^2 + \lambda\Delta_{12}\Delta_{13} = 0. \tag{A14}$$

Multiplying Eq. (A9) by Δ_{12} [or Eq. (A10) by Δ_{13}] and taking into account Eq. (A13) gives that the singularity in λ is such that $\lambda\Delta_{12}\Delta_{13}=0$. Using this in Eq. (A14) we finally obtain

$$\eta_1 q_1^2 - \eta_2|O_{12}|^2 - \eta_3|O_{13}|^2 = 0. \tag{A15}$$

This is the solution found in Sec. II, Eq. (2.11), and the rest of Sec. II follows from here and Eq. (A13).

For the sake of completeness we also give the expression for $1/\lambda$,

$$\frac{1}{\lambda} = -\sqrt{\frac{\Delta_{12}\Delta_{13}}{\eta_2 \eta_3}}, \tag{A16}$$

which exhibits no singularity. In fact, $1/\lambda = 0$ when $\Delta_{12} = \Delta_{13} = 0$, as expected. Finally, let us note that Eq. (A13), which is identical to Eq. (2.8), implies that all of the failure vectors, $|\phi_i\rangle$, are parallel to each other, i.e., they lie in a space, $\mathcal{A}$, of dimension one.

[1] I.D. Ivanovic, Phys. Lett. A **123**, 257 (1987).
[2] D. Dieks, Phys. Lett. A **126**, 303 (1988).
[3] A. Peres, Phys. Lett. A **128**, 19 (1988).
[4] G. Jaeger and A. Shimony, Phys. Lett. A **197**, 83 (1995).
[5] J. Bergou, M. Hillery, and Y. Sun, J. Mod. Opt. **47**, 487 (2000).
[6] A. Chefles and S.M. Barnett, Phys. Lett. A **236**, 177 (1997).
[7] B. Huttner, A. Muller, J.D. Gautier, H. Zbinden, and N. Gisin, Phys. Rev. A **54**, 3783 (1996).
[8] R.B.M. Clarke, A. Chefles, S.M. Barnett, and E. Riis, Phys. Rev. A **63**, 040305(R) (2001).
[9] A. Peres and D. Terno, J. Phys. A **31**, 7105 (1998).
[10] A. Chefles, Phys. Lett. A **239**, 339 (1998).
[11] A. Chefles and S.M. Barnett, Phys. Lett. A **250**, 223 (1998).
[12] L.-M. Duan and G.-C. Guo, Phys. Lett. A **261**, 25 (1999).
[13] Y. Sun, M. Hillery, and J.A. Bergou, Phys. Rev. A **64**, 022311 (2001); and J. Bergou, M. Hillery, and Y. Sun, Fortschr. Phys. **49**, 915 (2001), where a preliminary version of the current problem was first formulated.
[14] A. Chefles and S.M. Barnett, J. Mod. Opt. **45**, 1295 (1998).
[15] A. Chefles, Contemp. Phys. **41**, 401 (2000).
[16] U. Herzog and J. Bergou, Phys. Rev. A **65**, 050305(R) (2002).
[17] M. Reck, A. Zeilinger, H.J. Bernstein, and P. Bertani, Phys. Rev. Lett. **73**, 58 (1994).
[18] The case $N=1$ is presently being investigated by ourselves and independently by S.M. Barnett, A. Chefles, and I. Jex (unpublished).