Correspondence

Designing Optimal Quantum Detectors Via Semidefinite Programming

Yonina C. Eldar, *Member, IEEE*, Alexandre Megretski, and George C. Verghese, *Fellow, IEEE*

***Abstract*—We consider the problem of designing an optimal quantum detector to minimize the probability of a detection error when distinguishing among a collection of quantum states, represented by a set of density operators. We show that the design of the optimal detector can be formulated as a semidefinite programming problem. Based on this formulation, we derive a set of necessary and sufficient conditions for an optimal quantum measurement. We then show that the optimal measurement can be found by solving a standard (convex) semidefinite program. By exploiting the many well-known algorithms for solving semidefinite programs, which are guaranteed to converge to the global optimum, the optimal measurement can be computed very efficiently in polynomial time within any desired accuracy. Using the semidefinite programming formulation, we also show that the rank of each optimal measurement operator is no larger than the rank of the corresponding density operator. In particular, if the quantum state ensemble is a pure-state ensemble consisting of (not necessarily independent) rank-one density operators, then we show that the optimal measurement is a pure-state measurement consisting of rank-one measurement operators.**

***Index Terms*—Duality, quantum detection, semidefinite programming.**

I. Introduction

In a quantum detection problem, a transmitter conveys classical information to a receiver using a quantum-mechanical channel. Each message is represented by preparing the quantum channel in a quantum state represented by a density operator, drawn from a collection of known states. At the receiver, the information is detected by subjecting the channel to a quantum measurement in order to determine the prepared state. If the quantum states are mutually orthogonal, then the state can be determined correctly with probability one by performing an optimal orthogonal (von Neumann) measurement [1]. However, if the given states are not orthogonal, then no measurement will distinguish perfectly between them. Our problem is, therefore, to construct a measurement that minimizes the probability of a detection error.

We consider a quantum state ensemble consisting of m density operators $\{\rho_i,\ 1 \leq i \leq m\}$ on an n-dimensional complex Hilbert space $\mathcal{H}$, with prior probabilities $\{p_i > 0,\ 1 \leq i \leq m\}$. A density operator ρ is a positive semidefinite (PSD) Hermitian operator with $\mathrm{Tr}(\rho) = 1$; we write $\rho \geq 0$ to indicate ρ is PSD. A pure-state ensemble is one in which each density operator ρ_i is a rank-one projector $|\phi_i\rangle\langle\phi_i|$, where the vectors $|\phi_i\rangle$, though evidently normalized to unit length, are not necessarily orthogonal.

For our *measurement*, we consider general positive operator-valued measures [2], [3], consisting of m PSD Hermitian operators $\{\Pi_i,\ 1 \leq i \leq m\}$ that form a resolution of the identity on $\mathcal{H}$. A pure-state measurement is one in which each measurement operator Π_i is a rank-one operator[1] $|\mu_i\rangle\langle\mu_i|$, where the vectors $|\mu_i\rangle$ are not necessarily orthogonal or normalized. An orthogonal measurement (i.e., a von Neumann measurement) is one in which the measurement operators Π_i are mutually orthogonal projection operators.

Necessary and sufficient conditions for an optimum measurement minimizing the probability of a detection error have been derived [4], [5]. However, except in some particular cases [2], [6]–[10], 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 [11] or *ad hoc* suboptimal measurements are used. A detection measurement that has many desirable properties and has been employed in many settings is the least-squares measurement [9], also known as the square-root measurement [12], [13].

Holevo [4] derives the necessary and sufficient conditions by considering infinitesimal transformations of the measurement operators Π_i that preserve their character as elements of a measurement. The drawback of this approach is that it does not readily lend itself to efficient computational algorithms. Yuen *et al.* [5] use the principle of duality in vector space optimization to derive the same necessary and sufficient conditions. Specifically, they show that the problem of finding the measurement that minimizes the probability of a detection error can be formulated as a generalized linear programming problem, with the positive orthant being replaced by the positive cone of PSD matrices. Although their approach leads to the same conditions derived by Holevo [4], their apparent suggestion that this formulation produces a standard finite-dimensional linear programming problem is not correct, because the cone of PSD matrices cannot be described by a finite set of linear inequalities.

In this correspondence, we derive the necessary and sufficient conditions for an optimal quantum measurement in a self-contained manner, again by exploiting duality arguments. The primary advantage of our formulation is that it readily lends itself to efficient computational methods. Specifically, we show that the optimal measurement can be found by solving a standard convex semidefinite program. By exploiting the many well-known algorithms for solving semidefinite programs [14]–[17], the optimal measurement can be computed very efficiently in polynomial time within any desired accuracy. Furthermore, in contrast to the iterative algorithm proposed by Helstrom [11] for solving the quantum detection problem, which is only guaranteed to converge to a local optimum, algorithms based on semidefinite programming are guaranteed to converge to the global optimum.

After a statement of the problem in Section II, we derive, in Section III, the necessary and sufficient conditions for the optimal measurement that minimizes the probability of a detection error, by for-

Manuscript received May 28, 2002; revised November 18, 2002. This work was supported in part by BAE Systems Cooperative Agreement RP6891 under Army Research Laboratory Grant DAAD19-01-2-0008, by the Army Research Laboratory Collaborative Technology Alliance through BAE Systems Subcontract RK78554, and by Texas Instruments through the TI Leadership University Consortium.

Y. C. Eldar was with the Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, MA 02139 USA. She is now with the Technion–Israel Institute of Technology, Haifa 32000, Israel (e-mail: yonina@ee.technion.ac.il).

A. Megretski is with the Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, Cambridge, MA 02139 USA (e-mail: ameg@mit.edu).

G. C. Verghese is with the Laboratory for Electromagnetic and Electronic Systems, Massachusetts Institute of Technology, Cambridge, MA 02139 USA (e-mail: verghese@mit.edu).

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

Digital Object Identifier 10.1109/TIT.2003.809510

[1]In this correspondence, when we say rank-one operator we mean an operator that can be expressed in the form $\Pi_i = |\mu_i\rangle\langle\mu_i|$ for some $|\mu_i\rangle \in \mathcal{H}$. Note, however, that $|\mu_i\rangle$ may be equal to 0 in which case the operator actually has rank zero.

mulating our problem as a semidefinite program. Using this formulation, in Section IV, we prove that if the quantum state ensemble is a pure-state ensemble consisting of rank-one density operators $\rho_i = |\phi_i\rangle\langle\phi_i|$, then the optimal measurement is a pure-state measurement consisting of rank-one measurement operators $\Pi_i = |\mu_i\rangle\langle\mu_i|$. This generalizes a previous result by Kennedy [18], which establishes that for *linearly independent* vectors $|\phi_i\rangle$ the optimal measurement is a (necessarily orthogonal) pure-state measurement. We also show that for a mixed quantum state ensemble, the rank of each optimal measurement operator Π_i is no larger than the rank of the corresponding density matrix ρ_i. In Section V, we consider efficient iterative algorithms that are guaranteed to converge to the globally optimum measurement.

Throughout the correspondence, we use the Dirac bra-ket notation of quantum mechanics. In this notation, the elements of $\mathcal{H}$ are "ket" vectors, denoted, e.g., by $|x\rangle \in \mathcal{H}$. The corresponding "bra" vector $\langle x|$ is the conjugate transpose of $|x\rangle$. The inner product of two vectors is a complex number denoted by $\langle x|y\rangle$. An outer product of two vectors such as $|x\rangle\langle y|$ is a rank-one matrix, which takes $|z\rangle \in \mathcal{H}$ to $\langle y|z\rangle|x\rangle \in \mathcal{H}$.

II. Optimal Detection of Quantum States

Assume that a quantum channel is prepared in a quantum state drawn from a collection of given states. The quantum states are represented by a set of m PSD Hermitian density operators $\{\rho_i,\ 1 \le i \le m\}$ on an n-dimensional complex Hilbert space $\mathcal{H}$. At the receiver, a measurement is constructed, comprising m PSD Hermitian measurement operators $\{\Pi_i,\ 1 \le i \le m\}$ on $\mathcal{H}$. The problem is to choose the measurement operators to minimize the probability of detection error, i.e., the probability of incorrect detection of the transmitted state.

We assume without loss of generality that the eigenvectors of the density operators $\{\rho_i,\ 1 \le i \le m\}$ span[2] $\mathcal{H}$. In this case, to constitute a measurement, the measurement operators Π_i must satisfy

$$\sum_{i=1}^{m} \Pi_i = I \tag{1}$$

where I is the identity operator on $\mathcal{H}$.

We seek the PSD measurement operators $\{\Pi_i,\ 1 \le i \le m\}$ satisfying (1) that minimize the probability of a detection error, or equivalently, maximize the probability of correct detection. Given that the transmitted state is ρ_j, the probability of correctly detecting the state using measurement operators $\{\Pi_i,\ 1 \le i \le m\}$ is $\mathrm{Tr}(\rho_j\Pi_j)$. Therefore, the probability of correct detection is given by

$$P_d = \sum_{i=1}^{m} p_i \mathrm{Tr}(\rho_i\Pi_i) \tag{2}$$

where $p_i > 0$ is the prior probability of ρ_i, with $\sum_i p_i = 1$. Denoting by $\mathcal{B}$ the set of Hermitian operators on $\mathcal{H}$ and defining $\rho_i' = p_i\rho_i$, our problem reduces to the maximization problem

$$\max_{\Pi_i \in \mathcal{B}} \sum_{i=1}^{m} \mathrm{Tr}(\rho_i'\Pi_i) \tag{3}$$

subject to the constraints

$$\Pi_i \ge 0, \qquad 1 \le i \le m \tag{4}$$

$$\sum_{i=1}^{m} \Pi_i = I. \tag{5}$$

[2]Otherwise, we can transform the problem to a problem equivalent to the one considered in this correspondence by reformulating the problem on the subspace spanned by the eigenvectors of $\{\rho_i,\ 1 \le i \le m\}$.

Denoting by Λ the set of all ordered sets $\Pi = \{\Pi_i\}_{i=1}^m$, $\Pi_i \in \mathcal{B}$, satisfying (4) and (5), and defining $J(\Pi) = \sum_{i=1}^m \mathrm{Tr}(\rho_i'\Pi_i)$, we can express our maximization problem as

$$\max_{\Pi \in \Lambda} J(\Pi). \tag{6}$$

We refer to Λ as the feasible set, and to any $\Pi \in \Lambda$ as a feasible point. Since Λ is a compact set and $J(\Pi)$ is a continuous linear functional, there exist an optimal $\hat{\Pi} \in \Lambda$ and an optimal value $\hat{J}$ defined by

$$\hat{J} = J(\hat{\Pi}) \ge J(\Pi), \qquad \forall\, \Pi \in \Lambda. \tag{7}$$

Equipped with the standard operations of addition and multiplication by real numbers, $\mathcal{B}$ is an n^2-dimensional *real* vector space. By choosing an appropriate basis for $\mathcal{B}$, the problem of (3)–(5) can be put in the form of a standard semidefinite programming problem, which is a convex optimization problem; for a detailed treatment of semidefinite programming problems see, e.g., [15]–[17], [14]. By exploiting the many well-known algorithms for solving semidefinite programs [14], e.g., interior point methods[3] [17], [15], the optimal measurement can be computed very efficiently in polynomial time.

Recently, methods based on semidefinite programming have been employed in a variety of different problems in quantum detection and quantum information [19]–[24]. The fact that the optimal quantum detector can be found by solving a semidefinite program was pointed out independently in [19]. Here we provide a more general development. In particular, rather than relying on results that are scattered throughout the literature in various forms, in what follows we present a self-contained and direct derivation of the necessary and sufficient conditions for the optimal measurement. As we will see, this derivation also leads to efficient methods for computing the optimal measurement in cases in which an analytical solution is not known.

In the next section, we derive the necessary and sufficient conditions on the measurement operators by formulating a *dual problem*. The dual problem will also be used in Section V to develop efficient computational algorithms.

III. Dual Problem Formulation

Our objective is to formulate a *dual problem* whose optimal value serves as a certificate for $\hat{J}$. Specifically, we will formulate a minimization problem of the form $\min_X T(X)$ for some linear functional T such that for all feasible values of $X \in \mathcal{B}$, i.e., values of $X \in \mathcal{B}$ that satisfy a certain set of constraints, and for any $\Pi \in \Lambda$, we shall have $T(X) \ge J(\Pi)$. The dual problem, therefore, provides an upper bound on the optimal value of the original (primal) problem. In addition, we would like the minimal value of T, denoted $\hat{T}$, to be equal to $\hat{J}$. The equality $\hat{J} = \hat{T}$ will then lead to conditions of optimality on the measurement operators. Furthermore, in this case, instead of solving the primal problem, we can find $\hat{J}$ and the optimal measurement by solving the dual problem, which turns out to have far fewer decision variables.

A. Constructing the Dual Problem

A general method for deriving a dual problem is to invoke the separating hyperplane theorem [25], which states that two disjoint convex sets[4] can always be separated by a hyperplane. We will take one convex set to be the point 0, and then carefully construct another convex set

[3]Interior point methods are iterative algorithms that terminate once a prespecified accuracy has been reached. A worst case analysis of interior point methods shows that the effort required to solve a semidefinite program to a given accuracy grows no faster than a polynomial of the problem size. In practice, the algorithms behave much better than predicted by the worst case analysis, and in fact in many cases the number of iterations is almost constant in the size of the problem.

[4]A set C is convex if for any $x,\ y \in C$, $\alpha x + (1-\alpha)y \in C$ for all $\alpha \in [0,\ 1]$.

that does not contain 0. This set will capture the equality constraints in the primal problem and the fact that for any primal feasible point, the value of the primal function is no larger than the optimal value. The dual variables will then emerge from the parameters of the separating hyperplane.

In our problem, we have one equality constraint $\sum_{i=1}^m \Pi_i = I$, and we know that $\hat{J} \geq J(\Pi)$. Our constructed convex set will accordingly consist of matrices of the form $-I + \sum_{i=1}^m \Pi_i$ where $\Pi_i \in \mathcal{B}$ and $\Pi_i \geq 0$, and scalars of the form $r - J(\Pi)$ where $r > \hat{J}$. We thus consider the (n^2+1)-dimensional real vector space

$$\mathcal{L} = \mathcal{B} \times \mathcal{R} = \{(S, x): S \in \mathcal{B}, x \in \mathcal{R}\}$$

where $\mathcal{R}$ denotes the reals, with inner product defined by

$$\langle (W, y), (S, x) \rangle = \mathrm{Tr}(WS) + yx. \tag{8}$$

Note that since $W, S \in \mathcal{B}$, $\mathrm{Tr}(WS) \in \mathcal{R}$.

We now define the subset Ω of $\mathcal{L}$ by

$$\Omega = \left\{ \left(-I + \sum_{i=1}^m \Pi_i, r - \sum_{i=1}^m \mathrm{Tr}(\Pi_i \rho_i') \right) : \Pi_i \in \mathcal{B}, \Pi_i \geq 0, r \in \mathcal{R}, r > \hat{J} \right\}. \tag{9}$$

It is easily verified that Ω is convex, and $0 \notin \Omega$. Therefore, by the separating hyperplane theorem, there exists a *nonzero* vector $(Z, a) \in \mathcal{L}$ such that $\langle (Z, a), (Q, b) \rangle \geq 0$ for all $(Q, b) \in \Omega$, i.e.,

$$\mathrm{Tr}\left(Z \left(-I + \sum_{i=1}^m \Pi_i \right) \right) + a \left(r - \sum_{i=1}^m \mathrm{Tr}(\Pi_i \rho_i') \right) \geq 0 \tag{10}$$

for all $\Pi_i \in \mathcal{B}$ and $r \in \mathcal{R}$ such that $\Pi_i \geq 0$, $r > \hat{J}$. It will turn out that the hyperplane parameters (Z, a) define the optimal dual point. We first show that these parameters have to satisfy certain constraints, which lead to the formulation of the dual problem.

Note that (10) with $\Pi_i = 0$, $r \to \hat{J}$ implies

$$a\hat{J} \geq \mathrm{Tr}(Z). \tag{11}$$

Similarly, (10) with $r = \hat{J} + 1, \Pi_j = 0$ for $j \neq i, \Pi_i = t|x\rangle\langle x|$ where $|x\rangle \in \mathbb{C}^n$ is fixed and $t \to +\infty$ yields $\langle x|Z - a\rho_i'|x\rangle \geq 0$. Since $|x\rangle$ and i are arbitrary, this implies

$$Z \geq a\rho_i', \qquad 1 \leq i \leq m. \tag{12}$$

With $\Pi_i = 0$, $r \to +\infty$, (10) implies $a \geq 0$. If $a = 0$, then (12) yields $Z \geq 0$, and (11) yields $0 \geq \mathrm{Tr}(Z)$, which together means $Z = 0$. However, this would contradict the assumption that $(Z, a) \neq 0$. Therefore, we conclude that $a > 0$, and define $\hat{X} = Z/a$. Then (11) implies that

$$T(\hat{X}) \leq \hat{J} \tag{13}$$

where $T(X) = \mathrm{Tr}(X)$, and (12) implies that $\hat{X} \geq \rho_i'$ for $1 \leq i \leq m$. Let Γ be the set of $X \in \mathcal{B}$ satisfying $X \geq \rho_i'$, $1 \leq i \leq m$. Then, for any $X \in \Gamma$, $\Pi \in \Lambda$, we have

$$T(X) - J(\Pi) = \sum_{i=1}^m \mathrm{Tr}(\Pi_i (X - \rho_i')) \geq 0. \tag{14}$$

Since $\hat{X} \in \Gamma$, from (13) and (14) we conclude that $T(\hat{X}) = \hat{J}$.

Thus, we have proven that the dual problem associated with (3)–(5) is

$$\min_{X \in \mathcal{B}} T(X) \tag{15}$$

where $T(X) = \mathrm{Tr}(X)$, subject to

$$X \geq \rho_i', \qquad 1 \leq i \leq m. \tag{16}$$

Furthermore, we have shown that there exists an optimal $\hat{X} \in \Gamma$ and an optimal value $\hat{T}$ defined by

$$\hat{T} = T(\hat{X}) \leq T(X), \qquad \forall X \in \Gamma \tag{17}$$

such that

$$\hat{T} = \hat{J}. \tag{18}$$

B. Optimality Conditions

Let $\hat{\Pi}_i$ denote the optimal measurement operators that maximize (3) subject to (4) and (5), and let $\hat{X}$ denote the optimal X that minimizes (15) subject to (16). Then from (18) it follows that

$$\sum_{i=1}^m \mathrm{Tr}\left(\hat{\Pi}_i (\hat{X} - \rho_i') \right) = 0. \tag{19}$$

Since $\hat{X} \geq \rho_i'$ and $\Pi_i \geq 0$, (19) is satisfied if and only if

$$(\hat{X} - \rho_i')\hat{\Pi}_i = \hat{\Pi}_i (\hat{X} - \rho_i') = 0, \qquad 1 \leq i \leq m. \tag{20}$$

Once we find the optimal $\hat{X}$ that minimizes the dual problem (15), the constraint (20) is a necessary and sufficient condition on the optimal measurement operators $\hat{\Pi}_i$. We have already seen that this condition is necessary. To show that it is sufficient, we note that if a set of measurement operators Π_i satisfies (20), then $\sum_{i=1}^m \mathrm{Tr}(\Pi_i(\hat{X} - \rho_i')) = 0$ so that $J(\Pi) = T(\hat{X}) = \hat{J}$.

Note that the dual problem involves many fewer decision variables than the primal maximization problem. Specifically, in the dual problem, we have n^2 real decision variables while the primal problem has mn^2 real decision variables. Therefore, it is advantageous to solve the dual problem and then use (20) to determine the optimal measurement operators, rather than solving the primal problem directly. In Section V, we develop efficient algorithms that follow this strategy.

Using (1), (20), and (16) leads to the conditions

$$\sum_{i=1}^m \rho_i' \hat{\Pi}_i = \sum_{i=1}^m \hat{\Pi}_i \rho_i' \tag{21}$$

$$\sum_{i=1}^m \rho_i' \hat{\Pi}_i \geq \rho_j', \qquad 1 \leq i \leq m. \tag{22}$$

Thus, any optimal measurement $\hat{\Pi} = \{\hat{\Pi}_i\}_{i=1}^m$ must satisfy (21) and (22). These conditions are also derived in [5], [4]. However, as noted in the Introduction, the approach taken here lends itself to fast iterative algorithms, as we will see in Section V, and also provides additional insight into the optimal measurement operators, as we show in Section IV.

In [5], it was established that the conditions (21) and (22) together with (4) and (5) are also sufficient. For completeness, we repeat the argument here. Suppose that the measurement operators $\hat{\Pi}_i$ satisfy (21) and (22). Then $\hat{X} = \sum_{i=1}^m \hat{\Pi}_i \rho_i' \in \Gamma$. It then follows from (14) that for any set of measurement operators $\Pi_i \in \Lambda$

$$\sum_{i=1}^m \mathrm{Tr}(\Pi_i \rho_i') \leq \mathrm{Tr}(\hat{X}) = \sum_{i=1}^m \mathrm{Tr}(\hat{\Pi}_i \rho_i') \tag{23}$$

with equality for $\Pi_i = \hat{\Pi}_i$. Therefore, the measurement operators $\hat{\Pi}_i$ are optimal.

We summarize our results in the following theorem.

Theorem 1: Let $\{\rho_i, 1 \leq i \leq m\}$ denote a set of density operators with prior probabilities $\{p_i > 0, 1 \leq i \leq m\}$, and let $\{\rho_i' =$

$p_i\rho_i,\ 1 \le i \le m\}$. Let Λ denote the set of all ordered sets of Hermitian measurement operators $\Pi = \{\Pi_i\}_{i=1}^m$ that satisfy $\Pi_i \ge 0$ and $\sum_{i=1}^m \Pi_i = I$, and let Γ denote the set of Hermitian matrices X such that $X \ge \rho_i,\ 1 \le i \le m$. Consider the problem $\max_{\Pi\in\Lambda} J(\Pi)$ and the dual problem $\min_{X\in\Gamma} T(X)$, where $J(\Pi) = \sum_{i=1}^m \mathrm{Tr}(\rho_i'\Pi_i)$ and $T(X) = \mathrm{Tr}(X)$. Then

1) for any $X \in \Gamma$ and $\Pi \in \Lambda$, $T(X) \ge J(\Pi)$;
2) there is an optimal Π, denoted $\hat{\Pi}$, such that $\hat{J} = J(\hat{\Pi}) \ge J(\Pi)$ for any $\Pi \in \Lambda$;
3) there is an optimal X, denoted $\hat{X}$, such that $\hat{T} = T(\hat{X}) \le T(X)$ for any $X \in \Gamma$;
4) $\hat{T} = \hat{J}$;
5) given $\hat{X}$, a necessary and sufficient condition on the optimal measurement operators $\hat{\Pi}_i$ is $(\hat{X} - \rho_i')\hat{\Pi}_i = 0,\ 1 \le i \le m$.

IV. Rank-One Ensembles

Suppose now that the density operators ρ_i are rank-one operators of the form $\rho_i = |\phi_i\rangle\langle\phi_i|$ for some $|\phi_i\rangle \in \mathcal{H}$. In this case, it seems intuitively plausible that the optimal measurement will consist of rank-one measurement operators of the form $\hat{\Pi}_i = |\mu_i\rangle\langle\mu_i|$ for some $|\mu_i\rangle \in \mathcal{H}$.

There are some particular cases in which an analytical solution to the quantum detection problem is known [2], [6]–[10]. In all of these cases, when the density operators are rank-one operators, the optimal measurement also has rank one. In the special case in which the vectors $|\phi_i\rangle$ are *linearly independent*, Kennedy [18] showed that the optimal measurement is always a rank-one measurement. However, this implication has not been proven in the general case. Using the conditions for optimality we derived in the previous section, we now prove this implication for an arbitrary rank-one ensemble.

We have seen that the optimal measurement operators $\hat{\Pi}_i$ can be determined by solving (20), where $\hat{X}$ is the optimal matrix that minimizes (15) subject to (16). Thus, the measurement operators $\hat{\Pi}_i$ must lie in the null space of $\hat{X} - \rho_i'$, denoted $\mathcal{N}(\hat{X} - \rho_i')$, and consequently, $\mathrm{rank}(\hat{\Pi}_i) \le \dim(\mathcal{N}(\hat{X} - \rho_i'))$.

Since $\hat{X} \ge \rho_i,\ 1 \le i \le m$, it follows that $\hat{X}$ is positive definite on $\mathcal{H}$. Indeed, since the eigenvectors of the matrices ρ_i span $\mathcal{H}$, for any $h \in \mathcal{H}$ there exists an i such that $\langle h|\rho_i'|h\rangle > 0$, which implies that $\langle h|\hat{X}|h\rangle > 0$ for any $h \in \mathcal{H}$, so that $\mathcal{N}(\hat{X}) = \{0\}$. Now, for any two matrices Z_1 and Z_2, $\mathrm{rank}(Z_1 + Z_2) \ge \mathrm{rank}(Z_1) - \mathrm{rank}(Z_2)$, so that

$$\dim(\mathcal{N}(Z_1 + Z_2)) \le \dim(\mathcal{N}(Z_1)) + \mathrm{rank}(Z_2). \tag{24}$$

With $Z_1 = \hat{X}$ and $Z_2 = -\rho_i'$, (24) yields

$$\dim(\mathcal{N}(\hat{X} - \rho_i')) \le \mathrm{rank}(\rho_i') = \mathrm{rank}(\rho_i) \tag{25}$$

and

$$\mathrm{rank}(\hat{\Pi}_i) \le \dim(\mathcal{N}(\hat{X} - \rho_i')) \le \mathrm{rank}(\rho_i), \quad 1 \le i \le m. \tag{26}$$

In the special case in which the operators $\rho_i = |\phi_i\rangle\langle\phi_i|$ have rank-one, it follows immediately from (26) that the optimal measurement operators also have rank-one, so that they have the form $\hat{\Pi}_i = |\mu_i\rangle\langle\mu_i|$ for some $|\mu_i\rangle \in \mathcal{H}$.

If, in addition, the vectors $\{|\phi_i\rangle,\ 1 \le i \le m\}$ are linearly independent, then the vectors $\{|\mu_i\rangle,\ 1 \le i \le m\}$ must also be linearly independent since $\sum_{i=1}^m |\mu_i\rangle\langle\mu_i|$ is equal to the identity on $\mathcal{H}$, where now $\mathcal{H}$ is the m-dimensional space spanned by the vectors $|\phi_i\rangle$. Then, for $1 \le j \le m$

$$|\mu_j\rangle = \sum_{i=1}^m \langle\mu_i|\mu_j\rangle|\mu_i\rangle. \tag{27}$$

Since the vectors $|\mu_i\rangle$ are linearly independent, we must have that $\langle\mu_i|\mu_j\rangle = \delta_{ij}$ so that the vectors $|\mu_i\rangle$ are mutually orthogonal. We, therefore, recover the statement by Kennedy [18], that for a pure-state ensemble with linearly independent vectors, the optimal measurement is an orthogonal pure-state measurement.

We summarize our results in the following theorem.

Theorem 2: Let $\{\rho_i,\ 1 \le i \le m\}$ be a quantum-state ensemble consisting of density operators ρ_i with prior probabilities $p_i > 0$. Then, the optimal measurement consists of measurement operators $\{\Pi_i,\ 1 \le i \le m\}$ with $\mathrm{rank}(\Pi_i) \le \mathrm{rank}(\rho_i)$. In particular, if $\{\rho_i = |\phi_i\rangle\langle\phi_i|,\ 1 \le i \le m\}$ is a pure-state quantum ensemble, then the optimal measurement is a pure-state measurement consisting of measurement operators of the form $\{\Pi_i = |\mu_i\rangle\langle\mu_i|,\ 1 \le i \le m\}$.

V. Computational Aspects

In the general case, there is no closed-form analytical solution to the maximization problem (3) or the minimization problem (15). However, since (3) and (15) are convex optimization problems, there are very efficient methods for their solution. In particular, the optimal matrix $\hat{X}$ and the optimal measurement operators $\hat{\Pi}_i$ can be computed in Matlab using the linear matrix inequality (LMI) toolbox. A convenient interface for using the LMI toolbox is the Matlab package[5] IQCβ. The algorithm is guaranteed to converge to the global optimum within any desired accuracy in polynomial time.

Since (15) involves fewer decision variables than (3), in many cases it is computationally more efficient to first find the optimal matrix $\hat{X}$ minimizing $\mathrm{Tr}(X)$ subject to (16), and then determine the optimal measurement operators $\hat{\Pi}_i$ using (20), (4) and (5). Following this strategy, in the next section, we develop a procedure for computing the optimal measurement operators for rank-one ensembles. The case of mixed state ensembles is considered in Section V-C.

A. Rank-One Ensembles

If the density operators ρ_i have rank one, then, from Theorem 2, the optimal measurement operators $\hat{\Pi}_i$ also have rank one. From (20) and (4) it then follows that $\hat{\Pi}_i$ can be expressed as

$$\hat{\Pi}_i = a_i|q_i\rangle\langle q_i| \tag{28}$$

where $a_i \ge 0$, and $|q_i\rangle$ is a normalized vector that spans $\mathcal{N}(\hat{X} - \rho_i')$. To determine the vector $|q_i\rangle$ we may use the eigendecomposition of $\hat{X} - \rho_i'$.

To satisfy (5) we must have

$$\sum_{i=1}^m a_i|q_i\rangle\langle q_i| = I. \tag{29}$$

Let $|e\rangle = \mathrm{vec}(I)$ and $|y_i\rangle = \mathrm{vec}(|q_i\rangle\langle q_i|)$, where $|v\rangle = \mathrm{vec}(V)$ denotes the vector obtained by stacking the columns of V. Then we can express (29) as

$$Y|a\rangle = |e\rangle \tag{30}$$

where Y is the matrix of columns $|y_i\rangle$ and $|a\rangle$ is the vector with components a_i. If the matrix Y has full column rank, then the unique solution to (30) is

$$|a\rangle = (Y^*Y)^{-1}Y^*|e\rangle. \tag{31}$$

In the general case, Y will not have full column rank and there will be many solutions $|a\rangle$ to (30). Each such vector defines a corresponding set of optimal measurement operators $\hat{\Pi}_i$ via (28). To find a unique

[5]This software was created by A. Megretski, C.-Y. Kao, U. Jönsson, and A. Rantzer and is available at http://web.mit.edu/ameg/www/index.html.

solution we may seek the vector[6] $|a\rangle \geq 0$ that satisfies (30), and such that

$$\sum_{i=1}^{m} \mathrm{Tr}(\hat{\Pi}_i) = \sum_{i=1}^{m} a_i$$

is minimized. Our problem therefore reduces to

$$\min \langle 1|a\rangle \tag{32}$$

where $|1\rangle$ denotes the vector with components that are all equal 1, subject to

$$\begin{aligned} Y|a\rangle &= |e\rangle; \\ |a\rangle &\geq 0. \end{aligned} \tag{33}$$

The problem of (32), (33) is just a standard linear programming problem that can be solved very efficiently using standard linear programming tools [26], for example, the LMI toolbox in Matlab.

B. Example

We now consider an example illustrating the computational steps involved in computing the optimal measurement for a rank-one ensemble.

Consider the case in which the ensemble consists of three rank-one density operators $\rho_i = |\phi_i\rangle\langle\phi_i|$, $1 \leq i \leq 3$, where

$$|\phi_1\rangle = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad |\phi_2\rangle = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad |\phi_3\rangle = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \tag{34}$$

with prior probabilities

$$p_1 = 0.1, \qquad p_2 = 0.6, \qquad p_3 = 0.3. \tag{35}$$

To find the optimal measurement operators, we first find the optimal matrix $\hat{X}$ that minimizes $\mathrm{Tr}(X)$ subject to $X \geq \rho_i'$ with $\rho_i' = p_i \rho_i$. The matrix $\hat{X}$ is computed using the IQCβ toolbox on Matlab. To this end, we generate the following code (see the bottom of this page). The optimal $\hat{X}$ is given by

$$\hat{X} = \begin{bmatrix} 0.352 & 0.217 \\ 0.217 & 0.434 \end{bmatrix}. \tag{36}$$

Using the eigendecomposition of $\hat{X} - \rho_i'$, we conclude that, as expected from Theorem 2, $\mathcal{N}(\hat{X} - \rho_i')$ has dimension 1 for each i and is spanned by the vector $|q_i\rangle$ where

$$|q_1\rangle = \begin{bmatrix} -0.833 \\ 0.554 \end{bmatrix}, \quad |q_2\rangle = \frac{1}{\sqrt{2}} \begin{bmatrix} 0.850 \\ 0.527 \end{bmatrix}, \quad |q_3\rangle = \begin{bmatrix} -0.525 \\ 0.851 \end{bmatrix}. \tag{37}$$

The optimal measurement operators are therefore given by

$$\hat{\Pi}_i = a_i |q_i\rangle\langle q_i| = |\mu_i\rangle\langle\mu_i|$$

[6]The inequality is to be understood as a component-wise inequality.

with $|\mu_i\rangle = \sqrt{a_i}|q_i\rangle$ and a_i denoting the ith component of $|a\rangle$. From (30), $|a\rangle$ must satisfy

$$\begin{bmatrix} 0.693 & 0.722 & 0.276 \\ -0.461 & 0.448 & -0.447 \\ -0.461 & 0.448 & -0.447 \\ 0.306 & 0.278 & 0.724 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \end{bmatrix}. \tag{38}$$

Since the matrix in (38) has full column rank, there is a unique solution

$$|a\rangle = \begin{bmatrix} 0.007 \\ 0.999 \\ 0.994 \end{bmatrix}. \tag{39}$$

The optimal measurement vectors are then given by $|\mu_i\rangle = \sqrt{a_i}\,|q_i\rangle$ which yields

$$|\mu_1\rangle = \begin{bmatrix} -0.067 \\ 0.046 \end{bmatrix}, \quad |\mu_2\rangle = \begin{bmatrix} 0.849 \\ 0.527 \end{bmatrix}, \quad |\mu_3\rangle = \begin{bmatrix} -0.524 \\ 0.849 \end{bmatrix}. \tag{40}$$

We can immediately verify that the measurement operators $\hat{\Pi}_i = |\mu_i\rangle\langle\mu_i|$ with $|\mu_i\rangle$ given by (40) together with $\hat{X}$ given by (36) satisfy the necessary and sufficient conditions (4), (5), and (20). Furthermore, we have that the probability of correct detection is given by

$$\mathrm{Tr}(\hat{X}) = \sum_{i=1}^{m} p_i \mathrm{Tr}(\hat{\Pi}\rho_i) = 0.78. \tag{41}$$

In Fig. 1, we plot the weighted state vectors $|\psi_i\rangle = \sqrt{p_i}\,|\phi_i\rangle$ given by (34) and (35), together with the optimal measurement vectors $|\mu_i\rangle$ given by (40). For comparison, we also plot the least-squares measurement vectors $|\chi_i\rangle$ which are given by [9]

$$|\chi_i\rangle = (\Psi\Psi^*)^{-1/2}|\psi_i\rangle \tag{42}$$

where Ψ is the matrix of columns $|\psi_i\rangle$ and $(\cdot)^{1/2}$ is the unique symmetric square root of the corresponding matrix. Note, that since the vectors $|\phi_i\rangle$ span $\mathcal{H}$, $\Psi\Psi^*$ is invertible. The probability of correct detection using the least-squares measurement vectors is

$$\sum_{i=1}^{m} p_i |\langle\chi_i|\phi_i\rangle|^2 = 0.71.$$

As expected, this probability is smaller than the probability of correct detection using the optimal measurement vectors which from (41) is equal to 0.78.

```
>> abst_init_lmi                  % Initializing the LMI toolbox
>> X = symmetric(2);              % Defining a symmetric 2 x 2 variable X
>> X > p1 * R1;                   % Imposing the inequality constraints:
>> X > p2 * R2;                   % Here p1 = p1, p2 = p2, p3 = p3 and
>> X > p3 * R3;                   % R1 = rho1, R2 = rho2, R3 = rho3
>> lmi_mincx_tbx(trace(X));       % Minimizing Tr (X) subject to the constraints
>> X=value(X)                     % Getting the optimal value of X.
```

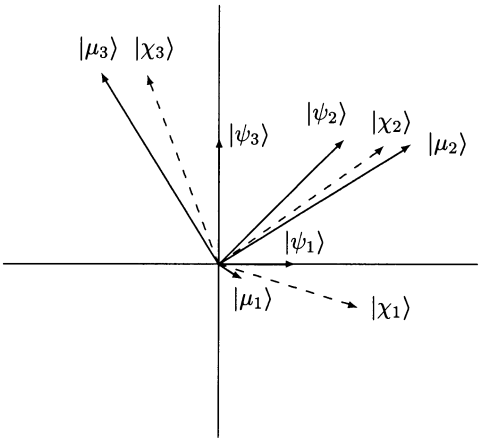

Fig. 1. Illustration of the optimal measurement vectors. The weighted state vectors are $|\psi_i\rangle = \sqrt{p_i}\,|\phi_i\rangle$ where the vectors $|\phi_i\rangle$ and the probabilities p_i are given by (34) and (35), respectively. The optimal measurement vectors $|\mu_i\rangle$ are given by (40). The least-squares measurement vectors $|\chi_i\rangle$ are plotted in dashed lines for comparison, and are given by (42).

C. Mixed State Ensembles

We now consider the case in which at least one of the density operators ρ_i has rank larger than 1. From (20) and (5), it follows that given $\hat{X}$, the optimal measurement operators $\hat{\Pi}_i$ that maximize (3) must satisfy

$$\begin{bmatrix} \hat{X}-\rho_1' & 0 & 0 & \cdots & 0 \\ 0 & \hat{X}-\rho_2' & 0 & \cdots & 0 \\ & & \vdots & & \\ 0 & 0 & \cdots & 0 & \hat{X}-\rho_m' \\ I & I & \cdots & I & I \end{bmatrix} \begin{bmatrix} \hat{\Pi}_1 \\ \hat{\Pi}_2 \\ \vdots \\ \hat{\Pi}_m \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ I \end{bmatrix}. \tag{43}$$

Conversely, any set of operators $\hat{\Pi}_i$ that satisfy (43) and in addition are Hermitian and PSD, maximize (3).

If the left-hand matrix in (43) has full column rank, then there are unique operators $\hat{\Pi}_i$ that satisfy (43). In this case, we are guaranteed that $\hat{\Pi}_i$ are Hermitian and PSD and are, therefore, the optimal measurement operators. If, on the other hand, the left-hand matrix in (43) does not have full column rank, then there are many possible operators satisfying (43), some of which may not be Hermitian and PSD. Thus, in this case, from all possible operators satisfying (43), we need to find a set of operators that is Hermitian and PSD. Alternatively, in this case, we may solve the primal problem directly.

REFERENCES

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

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

[3] A. Peres, "Neumark's theorem and quantum inseparability," *Found. Phys.*, vol. 20, no. 12, pp. 1441–1453, 1990.

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

[5] 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.

[6] 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.

[7] 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.

[8] M. Ban, K. Kurokawa, 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.

[9] Y. C. Eldar and G. D. Forney, Jr., "On quantum detection and the square-root measurement," *IEEE Trans. Inform. Theory*, vol. 47, pp. 858–872, Mar. 2001.

[10] Y. C. Eldar, A. Megretski, and G. C. Verghese. Optimal detection of symmetric mixed quantum states. [Online] Available: http://www.arXiv.org/abs/quant-ph/0211111

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

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

[13] 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.

[14] L. Vandenberghe and S. Boyd, "Semidefinite programming," *SIAM Rev.*, vol. 38, no. 1, pp. 40–95, Mar. 1996.

[15] F. Alizadeh, "Combinatorial optimization with interior point methods and semi-definite matrices," Ph.D. dissertation, Univ. Minn., Minneapolis, Oct. 1991.

[16] F. Alizadeh, "Optimization over the positive-definite cone: Interior point methods and combinatorial applications," in *Advances in Optimization and Parallel Computing*, P. Pardalos, Ed. Amsterdam, The Netherlands: North-Holland, 1992.

[17] Y. Nesterov and A. Nemirovski, *Interior-Point Polynomial Algorithms in Convex Programming*. Philadelphia, PA: SIAM, 1994.

[18] R. S. Kennedy, "On the optimum receiver for the M-ary linearly independent pure state problem," MIT Res. Lab. Electron. Quart. Progr. Rep., Tech. Rep. 110, July 1973.

[19] M. Ježek, J. Řeháček, and J. Fiurášek, "Finding optimal strategies for minimum-error quantum-state discrimination," *Phys. Rev. A*, vol. 65, pp. 060301–060306, June 2002.

[20] A. C. Doherty, P. A. Parrilo, and F. M. Spedalieri. Distinguishing separable and entangled states. [Online]. Available: http://www.arXiv.org/abs/quant-ph/0112007

[21] E. M. Rains, "A semidefinite program for distillable entanglement," *IEEE Trans. Inform. Theory*, vol. 47, pp. 2921–2933, Nov. 2001.

[22] K. Audenaert and B. De Moor. Optimizing completely positive maps using semidefinite programming. [Online]. Available: http://www.arXiv.org/abs/quant-ph/0109155

[23] Y. C. Eldar, "A semidefinite programming approach to optimal unambiguous discrimination of quantum states," *IEEE Trans. Inform. Theory*, vol. 49, pp. 446–456, Feb. 2003.

[24] Y. C. Eldar, "Mixed quantum state detection with inconclusive results," *Phys. Rev.*, to be published.

[25] D. G. Luenberger, *Optimization by Vector Space Methods*. New York: Wiley, 1968.

[26] D. Bertsimas and J. Tsitsiklis, *Introduction to Linear Optimization*. Belmont, MA: Athena Scientific, 1997.