

Iterative procedure for computing accessible information in quantum communication

Jaroslav Řeháček,¹ Berthold-Georg Englert,² and Dagomir Kaszlikowski²

¹*Department of Optics, Palacky University, 17. listopadu 50, 772 00 Olomouc, Czech Republic*

²*Department of Physics, National University of Singapore, Singapore 117542, Singapore*

(Received 2 September 2004; published 26 May 2005)

We present an iterative algorithm that finds the optimal measurement for extracting the accessible information in any quantum communication scenario. The maximization is achieved by a steepest-ascent approach toward the extremal point, following the gradient uphill in sufficiently small steps. We apply it to a simple *ad hoc* example, as well as to a problem with a bearing on the security of a tomographic protocol for quantum key distribution.

DOI: 10.1103/PhysRevA.71.054303

PACS number(s): 03.67.Hk, 89.70.+c

When studying problems in quantum information science, such as investigating strategies for eavesdropping on quantum communication lines or calculating the capacity of a quantum channel, one often needs to search for the quantum measurement that is optimal for the purpose at hand. Typically, the quantity to be maximized is the accessible information associated with a set of states. We describe here a numerical procedure for finding this optimum and illustrate it with two examples.

We consider the following quantum communication scenario. Alice sends quantum objects to Bob in one of J states, specified by subnormalized statistical operators ρ_j ($j=1, \dots, J$) whose traces $\text{tr}\{\rho_j\}$ are the probabilities with which they are sent. The state in which Bob receives the objects is therefore given by

$$\rho = \sum_{j=1}^J \rho_j \quad \text{with } \text{tr}\{\rho\} = 1. \quad (1)$$

He performs a positive-operator-valued measurement (POVM), composed of K positive operators Π_k that decompose the identity,

$$\sum_{k=1}^K \Pi_k = 1 \quad \text{with } \Pi_k \geq 0, \quad (2)$$

which he wishes to choose such that the joint probabilities

$$p_{jk} = \text{tr}\{\rho_j \Pi_k\}, \quad \sum_{j,k} p_{jk} = 1 \quad (3)$$

are most informative about the ρ_j that happens to be the case for the particular quantum object under consideration. Bob's figure of merit is the *accessible information*

$$I = \sum_{j,k} p_{jk} \ln \left(\frac{p_{jk}}{p_{j \cdot} p_{\cdot k}} \right), \quad (4)$$

where $p_{j \cdot}$ and $p_{\cdot k}$ are the marginal probabilities,

$$p_{j \cdot} = \sum_k p_{jk} = \text{tr}\{\rho_j\}, \quad p_{\cdot k} = \sum_j p_{jk} = \text{tr}\{\rho \Pi_k\}. \quad (5)$$

Here, $p_{j \cdot}$ is the probability that Alice sends the j th state, $p_{\cdot k}$ is the probability that Bob gets the k th measurement outcome,

and p_{jk} is the probability that she sends the j th state and he gets the k th outcome.

Since Alice's ρ_j 's are given, the accessible information of Eq. (4) is a nonlinear convex functional of Bob's POVM, $\Pi \equiv \{\Pi_1, \dots, \Pi_K\} \rightarrow I(\Pi)$. The convexity,

$$I(\Pi^{(\lambda)}) \leq (1-\lambda)I(\Pi^{(1)}) + \lambda I(\Pi^{(2)}) \quad (6)$$

for $\Pi_k^{(\lambda)} = (1-\lambda)\Pi_k^{(1)} + \lambda\Pi_k^{(2)}$ with $0 \leq \lambda \leq 1$, is well known (see, e.g., [1]). It follows immediately from

$$\left(\frac{\partial}{\partial \lambda} \right)^2 I(\Pi^{(\lambda)}) = \sum_{j,j',k} \frac{(p_{jk}^{(1)} p_{j'k}^{(2)} - p_{j'k}^{(1)} p_{jk}^{(2)})^2}{2p_{jk} p_{j'k} p_{\cdot k}} \geq 0, \quad (7)$$

where $p_{jk} = (1-\lambda)p_{jk}^{(1)} + \lambda p_{jk}^{(2)}$.

Now, since $I(\Pi)$ is convex, it acquires its global maximum at the boundary of the convex set of all POVMs. The challenge is then to find the maximizing POVM, and it is the objective of this article to describe an iteration procedure for an efficient numerical search.

We observe that $I(\Pi) = \sum_k \text{tr}\{R_k \Pi_k\}$ with

$$R_k = \sum_j \rho_j \ln \left(\frac{p_{jk}}{p_{j \cdot} p_{\cdot k}} \right), \quad (8)$$

and the response of $I(\Pi)$ to a variation of the POVM is given by $\delta I = \sum_k \text{tr}\{R_k \delta \Pi_k\}$ because there is no net contribution from the induced changes of the R_k 's. The variations $\delta \Pi_k$ are subject to the constraints of Eq. (2), which we enforce by first writing $\Pi_k = A_k^\dagger A_k$ and then imposing

$$\sum_k \delta \Pi_k = \sum_k (A_k^\dagger \delta A_k + \delta A_k^\dagger A_k) = 0 \quad (9)$$

on the variations δA_k . The most general form for these variations is

$$\delta A_k = i \sum_l \epsilon_{kl} A_l \quad \text{with } \epsilon_{kl}^\dagger = \epsilon_{lk}, \quad (10)$$

where the ϵ_{kl} 's are *arbitrary* infinitesimal operators. So,

$$\delta I = i \sum_{k,l} \text{tr} \{ \epsilon_{kl} A_l (R_k - R_l) A_k^\dagger \}, \quad (11)$$

and the POVMs at the stationary points of $I(\Pi)$ must necessarily be such that $A_l R_k A_k^\dagger = A_l R_l A_k^\dagger$ or

$$\Pi_l R_k \Pi_k = \Pi_l R_l \Pi_k \quad \text{for all } k, l. \quad (12)$$

Upon summing over k or l we arrive at an equivalent set of equations

$$\Pi_l \Lambda = \Pi_l R_l, \quad R_k \Pi_k = \Lambda \Pi_k, \quad (13)$$

which are adjoint statements of each other because

$$\Lambda = \sum_k R_k \Pi_k = \sum_l \Pi_l R_l \quad (14)$$

is hermitian. Mathematically speaking, Λ is the Lagrange multiplier of the constraint (9), and its physical significance is revealed by noting that $I = \text{tr} \{ \Lambda \}$.

Equations (12)–(14) have been investigated by Holevo [2]. They are not solved directly in our approach. Rather, we exploit the observation that Eq. (11) identifies the gradient in the POVM space.

Accordingly, our numerical procedure for finding the solution of Eq. (12) is an iteration method that realizes a steepest ascent toward the maximum of $I(\Pi)$. In each round we proceed in the direction of the gradient by putting $\epsilon_{kl} = -i\alpha A_k (R_k - R_l) A_l^\dagger$ in Eq. (10) whereby $\alpha > 0$ controls the step size. As an immediate consequence of Eq. (11), the first-order (in α) change of $I(\Pi)$ is then assuredly positive,

$$\left. \frac{\partial I}{\partial \alpha} \right|_{\alpha=0} = \sum_{k,l} \text{tr} \{ (R_k - R_l) \Pi_k (R_k - R_l) \Pi_l \} \geq 0, \quad (15)$$

and vanishes only at the stationary points where Eq. (12) holds. But we must correct for the second-order terms that give a nonzero value to the sum in Eq. (9).

A round of the iteration procedure thus consists of the following three steps.

Step 1: Use the existing approximate POVM to calculate p_{jk} , $p_{.k}$, and R_k in accordance with Eqs. (3), (5), and (8).

Step 2: Next, choose a “small” positive value for α in G_k

$$= 1 + \alpha \left(R_k - \sum_l R_l \Pi_l \right) \text{ and compute } \tilde{\Pi}_k = G_k^\dagger \Pi_k G_k.$$

Step 3: Finally, sum up these $\tilde{\Pi}_k$, $S = \sum_l \tilde{\Pi}_l$, and take

$$S^{-1/2} \tilde{\Pi}_k S^{-1/2} \text{ as the new, improved approximation for } \Pi_k. \quad (16)$$

We evaluate $I(\Pi)$ at the end of each round to verify that an acceptable value for α was chosen in step 2. A decrease of I , rather than an increase, would indicate an overshooting and thus tell us that α was too large [3].

Any randomly chosen POVM can be used as the starting point for the iteration (16), except for the maximally ignorant POVMs for which each member is simply a multiple of the identity, $\Pi_k = p_{.k}$. Since $p_{jk} = p_{j \cdot} p_{.k}$ for these POVMs, they re-

sult in $I(\Pi) = 0$ and thus mark the global minima of the accessible information. These minima are unstable fixpoints of the iteration (16), and so it is enough to perturb them slightly by admixing a small fraction of a randomly generated POVM.

As is the case for all steepest-ascent methods, there is the possibility of convergence toward a local, rather than a global, maximum. There is no absolute protection against this danger, but in practice one can fight it efficiently by running the iteration several times for comparison, with different starting points. This worked very well whenever we applied the iteration scheme (16).

Prior to any iteration, a choice must be made for the value of K —i.e., the number of elements in the POVM. According to a theorem by Davies [1], one never needs more than $K = r^2$ members, where r is the rank of ρ . But very often fewer elements will do. For example, if it is possible to represent all ρ_j by real matrices, then $K = \frac{1}{2}r(r+1)$ members suffice, as Sasaki *et al.* have shown [4]. A particular example that has a bearing on the security analysis of certain schemes for quantum cryptography is rank-1 ρ_j 's that form a so-called “acute pyramid,” for which one knows that $K = J$ or $K = J + 1$ will do, depending on the volume of the pyramid [5].

One approach is, therefore, to choose $K = r^2$ for general ρ_j 's and $K = \frac{1}{2}r(r+1)$ if the ρ_j 's have joint real matrix representations. After the iterative optimization, one would then look for equivalent members and combine them into one new member, thereby reducing the value of K . Members Π_{k_1} and Π_{k_2} are equivalent if $p_{jk_1} p_{j'k_2} = p_{j'k_1} p_{jk_2}$ for all j and j' , for then $R_{k_1} = R_{k_2}$, and the pair $(\Pi_{k_1} + \Pi_{k_2}, 0)$ is as good as the pair (Π_{k_1}, Π_{k_2}) .

Another approach is to begin with a small K value, with $K = J$ suggesting itself. Then, after optimizing for this K , one would add a randomly chosen Π_{K+1} and so increase K by 1, with the proper normalization to unit sum achieved analogously to step 3 of (16). The optimal POVM has been found when the increase of K becomes virtual—i.e., when the eventual reduction of equivalent members decreases K .

We note that the iteration scheme (16) can be used for the optimization of other functionals as well, as long as there are given ρ_j 's and a procedure for calculating the R_k 's from the ρ_j 's and Π_k 's. Indeed, the set of R_k 's is the functional gradient of $I(\Pi)$ and can be regarded as defining $I(\Pi)$ up to a Π -independent constant.

An example is Helstrom's classic problem of minimum-error discrimination [6], where one has $K = J$ and wishes to maximize $\sum_j p_{jj}$. Therefore, we have $R_k = \rho_k$ for Helstrom's problem and can iterate as described above without further ado. In view of the very different R_k 's, the optimal POVM for minimum-error discrimination will, as a rule, be different from the POVM that maximizes the accessible information of Eq. (4). Especially, when the states to be discriminated are pure states that are nearly collinear, $\text{tr} \{ \rho_j \rho_{j'} \} \approx p_{j \cdot} p_{j' \cdot}$, ρ is almost a pure state itself, and this tends to bias the optimal POVM for accessible information away from the one that minimizes the discrimination error. This is well illustrated by the acute pyramids mentioned above [5].

As a simple example let us consider $J = 2$ and ρ_1, ρ_2 with the following matrix representations:

TABLE I. Hellstrom's success rate (SR) $\sum_j p_{jj}$ for the example of Eqs. (17), as obtained for a particular iteration sequence that begins with a randomly chosen zeroth POVM, given for various numbers of iteration rounds (NIR). The right column lists the accessible information (AI) obtained for the successive POVMs that optimize SR iteratively.

NIR	SR	AI
0	0.5	0.0
10	0.838 445 2747	0.434 433 2664
20	0.840 825 0481	0.447 716 5615
30	0.840 886 8797	0.448 081 4579
40	0.840 888 4135	0.448 090 5243
50	0.840 888 4515	0.448 090 7489
∞	0.840 888 4524	0.448 090 7546

$$\rho_1 \triangleq \frac{1}{30} \begin{pmatrix} 0 & 0 & 0 \\ 0 & 5 & 2 \\ 0 & 2 & 10 \end{pmatrix}, \quad \rho_2 \triangleq \frac{1}{60} \begin{pmatrix} 5 & 2 & 0 \\ 2 & 25 & 0 \\ 0 & 0 & 0 \end{pmatrix}. \quad (17)$$

These are rank-2 matrices with $r=3$ for the rank of $\rho = \rho_1 + \rho_2$. The Helstrom problem is solved by Π_1 and Π_2 projecting on the subspaces associated with the positive and negative eigenvalues of $\rho_1 - \rho_2$, respectively, so that $\max \sum_j p_{jj} = 0.840 888 4524$. As shown in Table I the iteration converges to this value in a few dozen rounds.

The respective iterations for maximizing the accessible information are summarized in Fig. 1 for $K=2$ and $K=3$. No improvement is found for $K=4, 5$, or 6 , so that the optimal POVM has three members in this example. No attempt has been made, for Table I or Fig. 1, to optimize parameter α of (16). A judicious choice could reduce the necessary number of iteration rounds by much.

The algorithm of (16) provides *numerical* answers for the accessible information and identifies the optimal POVMs. This is already sufficient for many quantum information tasks, such as the checking of the security of noisy quantum communication or the calculation of the maximal information yield per quantum system sent. The applications are not limited to these, however. Sometimes it is possible, upon restricting the set of measurements over which the accessible information is maximized to some special class of measurements, to derive analytical forms for the optimal POVM. But its *global* optimality can often be only conjectured, without having a solid proof. Our algorithm can be used, in such cases, to check the optimality of the POVM found. Such checks were indeed performed for the low-dimensional "pyramids" of Ref. [5].

Another, even more important potential application is the use of the algorithm for finding the *analytical* solutions to optimization problems. For instance, it is typical in quantum communication protocols that the states that are to be distinguished by the eavesdropper constitute a highly symmetric family of quantum states. The optimal POVM then also tends to be relatively simple, which makes it sometimes possible to reconstruct its analytical form from the result of the numerical search.

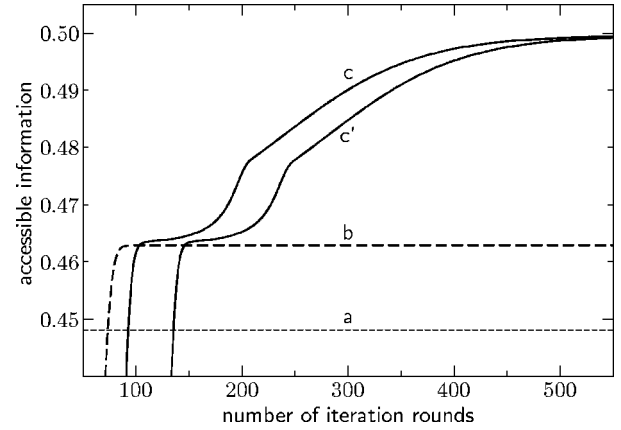


FIG. 1. Values of the accessible information for the example of Eqs. (17), as a function of the number of iteration rounds. The thin dashed horizontal line *a* shows the value obtained for the POVM that optimizes Hellstrom's success rate; see Table I. The thick dashed line *b* is for $K=2$, with the POVM of line *a* as the zeroth approximation. The optimal POVM is composed of a rank-1 projector and a rank-2 projector. Curves *c* and *c'* are for $K=3$ with different random choices for the zeroth approximation. The optimal POVM is a von Neumann measurement; i.e., it is composed of three rank-1 projectors. These iterations use prechosen values of α without any optimization of the step size. It is also worth mentioning that numerical experience indicates that the total number of iteration rounds does not depend much on the complexity of the problem—i.e., on the values of J, K , and the rank of ρ .

As an illustration, we consider the qubit version of the fully tomographic protocols for quantum key distribution of Ref. [7], an extension of the classic four-state protocol of Bennett and Brassard [8] or, rather, its six-state generalization [9], with some features of Ekert's protocol [10], which itself involves partial tomography. In this protocol, eavesdropper Eve controls the source that distributes qubit pairs to Alice and Bob, one pair at a time, in such a way that the statistical operator of these pairs is the singlet state $|\psi\rangle$ with an admixture of unbiased noise,

$$\rho_{A\&B} = |\psi\rangle\langle\psi| (1 - \epsilon) + \frac{\epsilon}{4} \mathbb{I} \quad (18)$$

with $0 \leq \epsilon \leq 1$.

Both Alice and Bob measure the six-element POVM consisting of the rank-1 operators $\frac{1}{6}(1 \pm \sigma_\zeta)$ with $\zeta = x, y, z$, where the σ_ζ 's are the respective basic Pauli operators for their qubits. The mutual information of the resulting joint probabilities is

$$I_{A\&B} = \frac{1}{6} [\epsilon \log_2 \epsilon + (2 - \epsilon) \log_2 (2 - \epsilon)], \quad (19)$$

where we now adopt the conventions of information theory and employ the binary logarithm rather than the natural logarithm that is more convenient in Eqs. (4)–(16).

At the source, each pair is entangled with an ancilla, which remains in Eve's possession as a quantum record of what has been sent to Alice and Bob. As shown in Ref. [7] (see also the Appendix in [5]), the best Eve can do is to use

another qubit pair for the ancilla and prepare the pure state

$$|\Psi\rangle = |\psi_{12}\psi_{34}\rangle a + |\psi_{13}\psi_{24}\rangle b, \quad (20)$$

where qubits 1 and 2 are Alice's and Bob's, respectively, qubits 3 and 4 make up the ancilla, and ψ_{jk} symbolizes the singlet for qubits j and k . The constraint $\text{tr}_{34}\{|\Psi\rangle\langle\Psi|\} = \rho_{A\&B}$ requires $|b|^2 = \epsilon$ and $|2a+b|^2 = 4-3\epsilon$, and it is convenient to choose the complex phases of a and b such that the product $\bar{\epsilon} \equiv (2a+b)b^* = \sqrt{4\epsilon-3\epsilon^2}$ is positive.

For each of Alice's (or Bob's) six measurement results there is a corresponding ancilla state

$$\rho_{\zeta^\pm} = \frac{1}{24} \left[1 \mp \frac{1}{2}(\epsilon + \bar{\epsilon})\sigma_\zeta^{(3)} \mp \frac{1}{2}(\epsilon - \bar{\epsilon})\sigma_\zeta^{(4)} - (1 - \epsilon)\vec{\sigma}^{(3)} \cdot \vec{\sigma}^{(4)} \right], \quad (21)$$

with $\zeta=x,y,z$. For Eve it is, therefore, as if Alice were sending her these subnormalized statistical operators, and what Eve can know about Alice's measurement results is measured by the accessible information associated with this sextet.

Eve's optimal POVM can be found numerically by the iteration procedure of (16). One finds that it has six elements which, upon careful inspection of the numerical results, are identified as the rank-1 operators,

$$\Pi_{\zeta^\pm} = \frac{1}{6} \left[1 \mp \frac{\sqrt{3}}{2}(\sigma_\zeta^{(3)} - \sigma_\zeta^{(4)}) - \frac{3}{2}\sigma_\zeta^{(3)}\sigma_\zeta^{(4)} + \frac{1}{2}\vec{\sigma}^{(3)} \cdot \vec{\sigma}^{(4)} \right], \quad (22)$$

with $\zeta=x,y,z$. Note the remarkable simplicity of the optimal POVM: it has relatively few elements and does not depend on the noise parameter ϵ . This is clearly a consequence of the symmetry of the sextet (21), which consists of unitarily equivalent rank-2 operators.

The accessible information gained by Eve from this POVM is given by

$$I_{A\&E} = I_{A\&B}(\epsilon \rightarrow 1 - \sqrt{3/4\bar{\epsilon}}), \quad (23)$$

with $I_{A\&B}(\epsilon)$ as in Eq. (19). Accordingly, the critical ϵ value, for which $I_{A\&E} = I_{A\&B}$, is given by

$$\epsilon_{\text{crit}} = \left(\frac{5}{2} + \sqrt{3} \right)^{-1} \approx 0.2363. \quad (24)$$

For $\epsilon < \epsilon_{\text{crit}}$, we have $I_{A\&B} > I_{A\&E}$, so that the information that Bob has about Alice's bit values exceeds Eve's information about them. Then, the Csiszár-Körner theorem [11] ensures that they can generate a private, secure cryptographic key by one-way communication.

We must not fail to mention that the POVM of Eq. (22) is only optimal for $\epsilon < \frac{2}{3}$ —that is, in the interesting parameter range where $\rho_{A\&B}$ is not separable and the joint probabilities between Alice and Bob contain nonclassical correlations. For $\epsilon \geq \frac{2}{3}$, Eve can blend $\rho_{A\&B}$ from product states and thereby obtain $I_{A\&E} = \frac{1}{3}$ right away.

In conclusion, we have presented an iterative numerical procedure for finding the POVM that optimally extracts the accessible information from a given set of states received in a quantum communication scenario. The method is a steepest-ascent approach toward the maximum; it follows the gradient in steps that are sufficiently small to avoid overshooting, so that the accessible information increases monotonically in each iteration step. We have illustrated the method at a simple *ad hoc* example. A second example, which has a bearing on quantum key distribution, shows how the analytical answer can be established, once crucial insight is gained from the numerical solution.

We wish to thank Frederick Willeboordse and Shiang Yong Looi for valuable discussions. This work was supported by Projects No. LN00A015 and MSM6198959213 of the Czech Ministry of Education, by A*Star Grant No. 012-104-0040, and by NUS Grant WBS: R-144-000-089-112 and R-144-000-109-112.

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

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

[3] One may speed up the convergence by adjusting the value of α . For example, after finding the new values I_1, I_2 of $I(\Pi)$ for two trial values α_1, α_2 , a good choice is

$$\alpha = \frac{1(I_1 - I_0)\alpha_2^2 + (I_0 - I_2)\alpha_1^2}{2(I_1 - I_0)\alpha_2 + (I_0 - I_2)\alpha_1},$$

where I_0 is the old value of $I(\Pi)$.

[4] M. Sasaki, S. M. Barnett, R. Jozsa, M. Osaki, and O. Hirota, Phys. Rev. A **59**, 3325 (1999).

[5] D. Kaszlikowski, A. Gopinathan, Y. C. Liang, L. C. Kwek, and B.-G. Englert, Phys. Rev. A **70**, 032306 (2004).

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

[7] Y. C. Liang, D. Kaszlikowski, B.-G. Englert, L. C. Kwek, and C. H. Oh, Phys. Rev. A **68**, 022324 (2003).

[8] C. H. Bennett and G. Brassard, in *Proceedings of the IEEE Conference on Computers, Systems, and Signal Processing, Bangalore, India*, December 1984 (IEEE, New York, 1984), p. 175.

[9] D. Bruß and C. Macchiavello, Phys. Rev. Lett. **88**, 127901 (2002).

[10] A. K. Ekert, Phys. Rev. Lett. **67**, 661 (1991).

[11] I. Csiszár and J. Körner, IEEE Trans. Inf. Theory **24**, 339 (1978).