RAPID COMMUNICATIONS

PHYSICAL REVIEW A VOLUME 60, NUMBER 5 NOVEMBER 1999

Universality of optimal measurements

Rolf Tarrach and Guifré Vidal*
Departament d'Estructura i Constituents de la Matèria, Universitat de Barcelona, Diagonal 647, E-08028 Barcelona, Spain

(Received 30 July 1999)

We present optimal and minimal measurements on identical copies of an unknown state of a quantum bit when the quality of measuring strategies is quantified with the gain of information (Kullback—or mutual information—of probability distributions). We also show that the maximal gain of information occurs, among isotropic priors, when the state is known to be pure. Universality of optimal measurements follows from our results: using the fidelity or the gain of information, two different figures of merits, leads to exactly the same conclusions for isotropic distributions. We finally investigate the optimal capacity of N copies of an unknown state as a quantum channel of information. [S1050-2947(99)51311-5]

PACS number(s): 03.67.−a, 03.65.Bz

Consider an unknown state of a two-level quantum system described by the density matrix $\rho(\vec{b})$, $\vec{b}$ being the Bloch vector, $b\equiv|\vec{b}|\leqslant 1$. The preparation device provides N identical copies of the system, so that the state at our disposal is $\rho(\vec{b})^{\otimes N}$. In the past few years the optimal measuring strategy, i.e., the most successful at revealing the identity of the unknown state, has been obtained, first for pure states [1–3] and then for mixed states [4]. Also the minimal ones among the optimal strategies, i.e., the ones with the smallest number of outcomes, have been constructed, both for pure states [5] and mixed states [4]. In the processing of information contained in quantum states, knowing the most efficient readout procedures, i.e., the optimal and least resource consuming ones, is of course of importance.

In all these contributions the quality of the measuring strategy, characterized by a resolution of the identity

$$\sum_i M_i=\mathbb{1}, \tag{1}$$

in terms of positive operators $M_i\geqslant 0$, has been quantified by the fidelity [6]. In other words, when outcome i (related to M_i) happens, one guesses the unknown state to be $\tilde{\rho}_i\equiv\rho(\vec{p}_i)$ and one quantifies the quality of the guess by

$$F(\rho(\vec{b}),\rho(\vec{p}_i))\equiv\{\mathrm{Tr}[\sqrt{\rho(\vec{b})^{1/2}\rho(\vec{p}_i)\rho(\vec{b})^{1/2}}]\}^2. \tag{2}$$

One can arrive at Eq. (2) from several different starting points. One of them is based on a measure of distinguishability of the probability distributions associated with ρ and ρ' by performing general positive operator valued measurements [as in Eq. (1)] on them [7] and minimizing,

$$F(\rho,\rho')=\min\left(\sum_j\sqrt{\mathrm{Tr}[\rho M_j]}\sqrt{\mathrm{Tr}[\rho' M_j]}\right)^2. \tag{3}$$

Another is based on the standard Hilbert-space scalar product of the two pure states, which, belonging to $\mathcal{C}^2\otimes\mathcal{C}^2$, lead to ρ and ρ' when reduced [8],

$$F(\rho,\rho')=\max|\langle\psi|\psi'\rangle|^2, \tag{4}$$

where maximization is performed over $\{|\psi\rangle,|\psi'\rangle\}/\rho=\mathrm{Tr}_a[|\psi\rangle\langle\psi|]$, $\rho'=\mathrm{Tr}_a[|\psi'\rangle\langle\psi'|]$.

These equivalent definitions of the fidelity, plus the following properties that characterize it further, make it a unique quantification of the comparison of two general quantum states: (i) $0\leqslant F(\rho,\rho')=F(\rho',\rho)\leqslant 1$. (ii) $F(\rho,\rho')=1\Leftrightarrow\rho=\rho'$; $F(\rho,\rho')=0\Leftrightarrow\rho\rho'=0$. (iii) $F(U\rho U^\dagger,U\rho' U^\dagger)=F(\rho,\rho')$, $UU^\dagger=U^\dagger U=\mathbb{1}$. (iv) $F(|\psi\rangle\langle\psi|,\rho)=\langle\psi|\rho|\psi\rangle$. (v) $F(\rho\otimes\sigma,\rho'\otimes\sigma')=F(\rho,\rho')F(\sigma,\sigma')$. (vi) $F[\rho,p\rho_1+(1-p)\rho_2]\geqslant pF(\rho,\rho_1)+(1-p)F(\rho,\rho_2), 0\leqslant p\leqslant 1$. In Refs. [2,3,5] the unknown state was known to be pure, $b=1$, but no knowledge of the direction of the Bloch vector was assumed. In reference [4] the unknown state was a mixed state drawn stochastically from a known isotropic distribution $f(b)$, and although the best guess $\tilde{\rho}_i$ depended on $f(b)$, the optimal measuring strategy, that is, the set $\{M_i\}$ of positive operators of the different outcomes, did not. For isotropic distributions optimal measurements are thus distribution, i.e., $f(b)$, independent.

However, proposing an outcome-dependent guess and evaluating its quality through the fidelity are only two of the criteria that could have been used to define optimal measurements. A sound alternative, the one we shall investigate in this work and probably the most sensible choice in the context of quantum information theory, consists of quantifying the quality of measuring strategies through the gain of information about the unknown state. In fact, information theory already supplies a universally accepted, unambiguous scheme for this purpose, which we shall follow. It is based on the Bayes formula, which provides a conditional (outcome-dependent), posterior distribution $f_c(\vec{b}|i)$ from the (here isotropic) prior distribution $f(b)$, and on the Kullback formula, which quantifies the gain of information acquired when replacing $f(b)$ with $f_c(\vec{b}|i)$.

More specifically, if $P_i(\vec{b})\equiv\mathrm{Tr}[\rho(\vec{b})M_i]$ is the probability of outcome i when the unknown state is $\rho(\vec{b})$ and

$$P_{ap}(i)\equiv\int d^3bf(b)P_i(\vec{b})\left(\int d^3bf(b)=1\right) \tag{5}$$

is the *a priori* probability of outcome i, then the Bayes formula states that the posterior distribution $f_c(\vec{b}|i)$, the one that collects our knowledge about the unknown state $\rho(\vec{b})$ after measuring when the initial knowledge is given by $f(b)$, reads

*Electronic address: guifre@ecm.ub.es

1050-2947/99/60(5)/3339(4)/$15.00

$$f_c(\vec{b}|i)=f(b)P_i(\vec{b})/P_{ap}(i). \tag{6}$$

The gain of information about $\rho(\vec{b})$, ΔI, is then given, in bits, by the Kullback formula of $f_c(\vec{b}|i)$ relative to $f(b)$ [12]

$$K_i[f_c/f]\equiv\int d^3b f_c(\vec{b}|i)\log_2[f_c(\vec{b}|i)/f(b)]. \tag{7}$$

This expression, the only one satisfying a series of intuitively reasonable conditions [13], is well-defined for continuous distributions (it has no dependence on the measure in the space of quantum states), and its average over possible outcomes,

$$\overline{K}[f_c/f]\equiv\sum_i P_{ap}(i)K_i[f_c/f], \tag{8}$$

is precisely the difference of the *a priori* and average *a posteriori* entropies H of the corresponding probability distributions of states,

$$H[f]-\overline{H}[f_c]\equiv-\int d^3b\, f(b)\log_2 f(b)+\sum_i P_{ap}(i) \times\int d^3b\, f_c(\vec{b}|i)\log_2 f_c(\vec{b}|i), \tag{9}$$

as can be checked by considering Eqs. (6)–(8) and that $\Sigma_i P_i(\vec{b})=1$ [9]. This quantification is therefore equivalent to the one already used in previous works on quantum-state estimation with discrete distributions (see, e.g., Ref. [10]).

First, the question of which are the optimal measurements according to this information theoretically based criterion will be addressed. We will check explicitly for $N=1$ and $N=2$, and provide clues for any N, that optimal—and also minimal—measuring strategies are universal, i.e., independent of whether the fidelity or the increase of information is used for their quantification [11], and will compute the corresponding optimal gain of information ΔI. Then we will move to consider which is the isotropic prior $f(b)$ for which optimal measurements extract most information, so that it corresponds to the optimal (isotropic) quantum channel of information. After introducing a reversible compression procedure we conclude that the optimal amount of extractable information is, as $N\rightarrow\infty$, of one bit per effective quantum bit (qubit) isotropic distributions.

In order to find an optimal measuring strategy, i.e., a set of operators M_i as in Eq. (1) maximizing the gain of information [Eq. (8)], the following theorem and subsequent corollaries, valid for any number of copies N, will be very useful.

Theorem. Let the positive operator $M_i\geqslant 0$ be such that its probability $P_i(\vec{b})=\mathrm{Tr}[M_i\rho(\vec{b})^{\otimes N}]$ can be written, for any $\vec{b}$, as the sum of two contributions of the form $P_{i,k}(\vec{b})\equiv\mathrm{Tr}[M_{i,k}\rho(\vec{b})^{\otimes N}]$, $k=1,2$, where the operators $M_{i,1},M_{i,2}$ are also positive (and $M_{i,1}+M_{i,2}$ is not necessarily equal to M_i). Let us introduce corresponding prior probabilities $P_{ap}(i,k)$ and posterior distributions $f_c(\vec{b}|i,k)$ as in Eqs. (5) and (6). Then,

$$P_{ap}(i)K_i[f_c/f]\leqslant\sum_{k=1}^{2}P_{ap}(i,k)K_{i,k}[f_c/f]. \tag{10}$$

Proof. It follows from the inequality

$$(x_1+x_2)\ln\frac{x_1+x_2}{y_1+y_2}\leqslant x_1\ln\frac{x_1}{y_1}+x_2\ln\frac{x_2}{y_2}, \tag{11}$$

$\forall\ x_1,x_2,y_1,y_2\geqslant 0$. ■

Corollary 1. An optimal measuring strategy with rank-1 operators always exists (cf. [14]).

Proof. Indeed, suppose $\Sigma_i M_i=\mathbb{1}$ corresponds to an optimal measurement. Then, if $M_i=\Sigma_k|i,k\rangle\langle i,k|$ is the spectral decomposition of M_i, it follows from the theorem that the rank-1 positive operator valued measurement $\Sigma_{i,k}|i,k\rangle\langle i,k|=\mathbb{1}$ is also optimal. ■

We can already consider the case $N=1$, that is, when only one copy of the unknown state is available. One can convince oneself immediately that an optimal (and also minimal) measurement is just a standard von Neumann measurement. In fact, any will do because of the isotropy of $f(b)$. Suppose that we measure σ_z. Then, for $\vec{b}=(b\sin\theta\cos\phi,b\sin\theta\sin\phi,b\cos\theta)$, we have

$$f_c(\vec{b}|\pm)=(1\pm b\cos\theta)f(b) \tag{12}$$

and the gain of information is

$$\Delta I^{(1)}=\pi\int_0^1 db\, b^2 f(b)[(1+b)^2/b]\log_2(1+b) -[(1-b)^2/b]\log_2(1-b)-\log_2 e/2. \tag{13}$$

The function in square brackets in Eq. (13) is monotonically increasing, so that the distribution for which the absolute increase in knowledge is maximal is

$$f_m^{(1)}(b)=(1/4\pi)\delta(b-1), \tag{14}$$

i.e., an isotropic distribution of pure states.

It is interesting to point out that, if instead of using in Ref. [4] the mean average fidelity $\overline{F}^{(1)}$ we had used the mean average increase in fidelity,

$$\Delta F^{(1)}\equiv\overline{F}^{(1)}-F_{ap}^{(1)}, \tag{15}$$

with the optimal guess $\tilde{\rho}_0\equiv\rho(0)$ if no measurement was performed, so that

$$F_{ap}^{(1)}=\tfrac{1}{2}+I_{1/2}=F_{ap}^{(N)} \tag{16}$$

with (cf. [4])

$$I_\alpha\equiv 4\pi\int_0^1 db\, b^2 f(b)[(1-b^2)/4]^\alpha \tag{17}$$

$(I_0=1,I_\alpha\geqslant 4I_{\alpha+1})$, we would have obtained

$$\Delta F^{(1)}=\sqrt{I_{1/2}^2+\tfrac{1}{36}(1-4I_1)^2}-I_{1/2}. \tag{18}$$

It is then easily verified that the maximum value of $\Delta F^{(1)}$ also corresponds to the distribution equation (14). Thus, for $N=1$, quantifying with the fidelity or with the Kullback information leads to the same (for $N=1$ somewhat obvious) optimal and minimal measuring strategy and to the same distribution that maximizes $\Delta I^{(1)}$ and $\Delta F^{(1)}$. Is this also true for $N=2$?

In order to answer this question we need to present a second corollary. Notice first that with the following notation (borrowed from [4]) for the composite Hilbert space of N copies of the unknown state $\rho(\vec{b})$,

$$\mathcal{H}^{(N)}\equiv\mathcal{H}_A\otimes\mathcal{H}_B\otimes\ldots\mathcal{H}_N, \tag{19}$$

for the corresponding local spin operators,

$$\begin{aligned} \vec{S}_A &\equiv \frac{1}{2}\vec{\sigma}\otimes I^{\otimes N-1}, \\ \vec{S}_B &\equiv \frac{1}{2} I\otimes\vec{\sigma}\otimes I^{\otimes N-2}, \\ &\vdots \\ \vec{S}_N &\equiv \frac{1}{2} I^{\otimes N-1}\otimes\vec{\sigma}, \end{aligned} \tag{20}$$

and for the partial and total spin operators,

$$\vec{S}_{(\alpha)} \equiv \sum_{\beta=A}^{\alpha} \vec{S}_\beta, \quad \alpha = A,B,\dots,N, \quad \vec{S}\equiv\vec{S}_{(\alpha=N)}, \tag{21}$$

the following spin invariances hold [4]:

$$[\vec{S}^2_{(\alpha)}, \rho^{\otimes N}] = 0, \quad \alpha = A,\dots,N, \tag{22}$$

and since

$$[\vec{S}^2_{(\alpha)}, \vec{S}^2_{(\beta)}] = 0, \quad \forall \;\; \alpha,\beta, \tag{23}$$

the total Hilbert space can be written as a direct sum

$$\mathcal{H}^{(N)} = \oplus_{\{s_{(\alpha)}\}} E_{\{s_{(\alpha)}\}}, \tag{24}$$

where $E_{\{s_{(\alpha)}\}}$ are the simultaneous eigenspaces of all the operators $\vec{S}^2_{(\alpha)}, \forall \alpha \neq A$, with corresponding eigenvalues $\{s_\alpha(s_\alpha+1)\}$, ordered with decreasing α (see [4] for more details). For instance, for $N=2$ only $\vec{S}^2_{(B)}$ ($s_{(B)}$) is relevant, i.e., $E_{\{s_{(\alpha)}\}} = E_{s_{(B)}}$, and the decomposition reads

$$\mathcal{H}^{(N=2)} = E_1 \oplus E_0, \tag{25}$$

where E_1 is the triplet or symmetric (under exchange of copies) subspace, with total spin $s\equiv s_{(B)}=1$, whereas E_0 is the singlet or antisymmetric subspace, with total spin $s=0$. Then, we have the following.

Corollary 2. There always exists an optimal measuring strategy consisting only of rank-1 operators of the form $|\{s_{(\alpha)}\}\rangle\langle\{s_{(\alpha)}\}|$, where the not necessarily normalized vector $|\{s_{(\alpha)}\}\rangle$ is an eigenvector of all partial and total spin operators, i.e.,

$$\vec{S}^2_{(\beta)}|\{s_{(\alpha)}\}\rangle = s_{(\beta)}(s_{(\beta)}+1)|\{s_{(\alpha)}\}\rangle, \quad \forall \;\; \beta, \tag{26}$$

and thus it belongs to the subspace $E_{\{s_{(\alpha)}\}}$.

Proof. Let $\Sigma_i M_i = \mathbb{1}$ correspond to an optimal measurement with rank-1 operators $M_i = |i\rangle\langle i|$ (where the $|i\rangle$ do not need to be orthogonal nor normalized) and let $\Pi_{\{s_\alpha\}} = \Pi^2_{\{s_\alpha\}}$ be a projector onto the whole subspace $E_{\{s_\alpha\}}$. Then it follows from Eq. (24) that

$$\sum_{\{s_\alpha\}} \Pi_{\{s_\alpha\}} \mathbb{1} \Pi_{\{s_\alpha\}} = \sum_{\{s_\alpha\}} \Pi_{\{s_\alpha\}} = \mathbb{1}, \tag{27}$$

so that if we replace $\mathbb{1}$ with $\Sigma_i M_i$ in the left-hand side of this equation, we obtain a new measurement

$$\sum_{i,\{s_\alpha\}} |i,\{s_\alpha\}\rangle\langle i,\{s_\alpha\}| = \mathbb{1}, \quad |i,\{s_\alpha\}\rangle \equiv \Pi_{\{s_\alpha\}}|i\rangle. \tag{28}$$

Now, since Eq. (22) implies that for each $|i\rangle$,

$$\mathrm{Tr}[\rho(\vec{b})^{\otimes N}|i\rangle\langle i|] = \sum_{\{s_\alpha\}} \mathrm{Tr}[\rho(\vec{b})^{\otimes N}|i,\{s_\alpha\}\rangle\langle i,\{s_\alpha\}|], \tag{29}$$

the theorem guarantees that the measurement of Eq. (28) is also optimal. ■

(Notice that exactly the same conclusion was also achieved, for any N, when the fidelity was used as a criterion for optimality [4], this being indicative of the universality we are considering here.)

Thus, in order to find an optimal measuring strategy for $N=2$ we can always choose the pure states on which the measurement projects to be symmetric or antisymmetric under the exchange of the two qubits. Let us next compute $\Delta I^{(2)}$ for the optimal strategy of Ref. [4], that is, corresponding to a resolution of the identity of the form

$$\mathbb{1} = |\sigma\rangle\langle\sigma| + \frac{3}{4}\sum_{i=1}^{4} (|\hat{n}_i\rangle\langle\hat{n}_i|)^{\otimes 2}, \tag{30}$$

where $|\sigma\rangle$ is the (normalized) singlet state, $\vec{\sigma}\cdot\hat{n}|\hat{n}\rangle = |\hat{n}\rangle$ ($\langle\hat{n}|\hat{n}\rangle = 1$) and the four unitary vectors $\hat{n}_i$ point to the four directions of the vertices of a regular tetrahedron. One readily obtains

$$f_c(\vec{b}|\sigma) = [(1-b^2)/4]\,\frac{f(b)}{P_{ap}(\sigma)}, \quad P_{ap}(\sigma) = I_1, \tag{31}$$

$$\begin{aligned} f_c(\vec{b}|\hat{n}) &= \tfrac{3}{16}(1+\vec{b}\cdot\hat{n})^2[f(b)/P_{ap}(\hat{n})], \\ P_{ap}(\hat{n}) &= \tfrac{1}{2}(1-I_1), \end{aligned} \tag{32}$$

so that

$$\begin{aligned} \Delta I^{(2)} = \pi\int_0^1 db\, b^2 f(b)\{&[(1+b)^3/b]\log_2(1+b) \\ &-[(1-b)^3/b]\log_2(1-b) + (1-b^2)\log_2(1-b^2)\} \\ &-(1-I_1)\{2\log_2 e/3 + \log_2[(1-I_1/3]\} \\ &-I_1\log_2 I_1 - 2. \end{aligned} \tag{33}$$

Can we do better, i.e., is there another resolution of the identity that leads to a larger $\Delta I^{(2)}$? Let us prove that there is none. Because of corollary 2, the whole question boils down to whether symmetric entangled states could do better than the symmetric product states $|\hat{n}_i\rangle|\hat{n}_i\rangle$ used in Eq. (30). Consider therefore a general symmetric state of Schmidt decomposition

$$|\psi\rangle = \sqrt{p}|+\rangle|+\rangle + \sqrt{1-p}|-\rangle|-\rangle, \quad p\in[0,1], \tag{34}$$

where the isotropy of $f(b)$ has been taken into account in choosing the basis. One can readily obtain the average Kullback information corresponding to this state,

$$\begin{aligned} \Delta I^{(2)}_\psi &= \frac{1}{2}\int_0^1 db\, b^2 f(b)\int_0^{2\pi} d\phi \int_{-1}^{1} d\mu\, h\log_2 h/[(1-I_1)/3], \\ h &\equiv k + l\cos 2\phi, \quad l\equiv 2\sqrt{p(1-p)}b^2(1-\mu^2), \\ k &\equiv 1 + b^2\mu^2 + (2p-1)2b\mu, \end{aligned} \tag{35}$$

which after integration of ϕ gives

$$\begin{aligned} \Delta I^{(2)}_\psi = \frac{\pi}{2}\int_0^1 db\, b^2 f(b)\int_{-1}^{1} d\mu\{&(1+b^2\mu^2) \\ &\times\log_2[3e/2(1-I_1)] + k\log_2(k+\sqrt{k^2-l^2})\}. \end{aligned} \tag{36}$$

This is a function of p that we want to maximize. Only $k\log_2(k+\sqrt{k^2-l^2})$ depends on p. The part $-l^2$ is maximized for $p=0$ and $p=1$. The other part, too, as one can see easily neglecting the term l^2. Thus $\Delta I_{\psi}^{(2)}$ is maximized when $|\psi\rangle$ is a product state and the resolution of Eq. (30) is indeed optimal.

As we did for $N=1$, it is interesting to recall, with the help of Ref. [4], the average increase in fidelity for $N=2$

$$\Delta F^{(2)}=\sqrt{(I_{1/2}-I_{3/2})^2+\tfrac{1}{16}(1-4I_1)^2}+I_{3/2}-I_{1/2}. \tag{37}$$

One can now check that both $\Delta I^{(2)}$ and $\Delta F^{(2)}$ are again maximized for the distribution equation (14). For $\Delta I^{(2)}$ this follows by observing that the part in square brackets in Eq. (33) is an increasing function of b and that the other part, which depends on I_1, increases as I_1 goes towards zero.

We have thus checked for $N=1$ and $N=2$ that both the fidelity and the Kullback information lead to the same optimal measuring strategy and to the same, pure-state distribution that maximizes their increases. We conjecture, while not foreseeing any feature that could jeopardize extending the proof to $N>2$, that the universality of optimal measurements holds for any number N of copies of the unknown state [15]. Corollary 2 makes this conjecture very plausible. The precise optimal strategy is in fact determined to a great extent by the isotropy of the prior distribution, the symmetries of the state $\rho(\vec{b})^{\otimes N}$ that allow us to choose each positive operator M_i to act only on one of the subspaces $E_{\{s(\alpha)\}}$, and the fact that both the fidelity and the Kullback information favor strategies with outcomes i whose normalized probability of occurrence $\mathrm{Tr}[\rho(\vec{b})^N M_i]/\mathrm{Tr}[M_i]$ spans the largest possible range as a function of the direction of $\vec{b}$.

Now, suppose we want to use the N qubits as a quantum channel of classical information. Alice prepares N copies of a given state $\rho(\vec{b})$ (the classical information being encoded in the vector $\vec{b}$) and sends them to Bob, who will perform a collective measurement in order to recover as much information about $\vec{b}$ as possible. The previous results single out, using, when restricted to isotropic prior distributions, only pure states ($b=1$) to encode classical information as the optimal method. We can then easily compute the optimal capacity of this isotropic quantum channel for any N, to find that

$$\Delta I^{(N)}=\log_2(N+1)-[N/(N+1)]\log_2 e, \tag{38}$$

which for large N gives $\log_2 N/N$ bits carried per qubit. Notice that this is a purely quantum channel, no additional flow of classical information being required at any stage. Its poor capacity can be exponentially enhanced without spoiling this fact if we take into account that a pure state $\phi^{\otimes N}$ belongs to the symmetric subspace $\mathcal{S}^{(N)}$ of the whole Hilbert space $\mathcal{H}^{(N)}$. Since the dimension of $\mathcal{S}^{(N)}$ is $N+1$, which corresponds to the dimension of a Hilbert space $\mathcal{H}^{(M)}$ of $M \equiv \log_2(N+1)$ qubits, Alice can always compress, by means of a state-independent, unitary (and thus fully reversible) transformation, the state $\phi^{\otimes N}$ to fit in M qubits, which will then be transferred to Bob. In this case the capacity increases up to $1-O(1/\log N)$ bits per qubit, which is asymptotically the classical one (as expected, since for any two inequivalent states ϕ and ϕ', $\phi^{\otimes N}$, and $\phi'^{\otimes N}$ become orthogonal as $N\rightarrow\infty$), and which is consistent with the Levitin-Holevo bound [16] for the classical capacity of a quantum channel.

Summarizing, using the gain of information as a guide, we have constructed optimal and minimal measurements on $N=1,2$ identical copies and have shown that for isotropic distributions the maximal gain of information is achieved for pure states. Also the universality of optimal measurements has been proven, since these measurements exactly coincide with those obtained in previous work, where the fidelity was taken as the figure of merit. We conjecture that also for $N \geqslant 3$ the most informative measurements are the most faithful ones, and vice versa.

G.V. acknowledges CIRIT Grant No. 1997FI-00068 PG. Financial support from CIRYT, Contract No. AEN98-0431; CIRIT, Contract No. 1998SGR-00026; and from the ESF-QIT program is also acknowledged. We thank Antonio Acin and Chris Fuchs for their comments.

[1] A.S. Holevo, *Probabilistic and Statistical Aspects of Quantum Theory* (North-Holland, Amsterdam, 1982), Chap. IV.

[2] S. Massar and S. Popescu, Phys. Rev. Lett. **74**, 1259 (1995).

[3] R. Derka *et al.*, Phys. Rev. Lett. **80**, 1571 (1998).

[4] G. Vidal *et al.*, Phys. Rev. A **60**, 126 (1999).

[5] J.I. Latorre *et al.*, Phys. Rev. Lett. **81**, 1351 (1998).

[6] A. Uhlmann, Rep. Math. Phys. **9**, 273 (1976).

[7] C.A. Fuchs and C.M. Caves, Open Syst. Inform. Dynam. **3**, 345 (1995); also e-print quant-ph/9604001.

[8] R. Josza, J. Mod. Opt. **41**, 2315 (1994).

[9] Furthermore, and due to the symmetry in the i and $\vec{b}$ distributions of Bayes' formula, Eq. (6), the expressions in Eqs. (8) and (9) are also equal to the corresponding expressions for which f and f_c have been traded for P_{ap} and P.

[10] A. Peres and W.K. Wootters, Phys. Rev. Lett. **66**, 1119 (1991).

[11] Holevo [1] presented examples of the universal character of optimal measurements in the context of phase estimation, and related universality to convexity properties of the figures of merits (as in our theorem).

[12] S. Kullback and R.A. Leibler, Ann. Math. Stat. **22**, 78 (1951); S. Kullback, *Information Theory and Statistics* (Wiley, New York, 1959).

[13] A. Hobson, J. Stat. Phys. **1**, 383 (1969).

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

[15] We have also been able to check, for an arbitrary number N of copies of the unknown state, that the optimal measurements according to the fidelity (as presented in [4]) are at least locally optimal (that is, better than any other measurement that follows from infinitesimally perturbing the former) for the Kullback information.

[16] L.B. Levitin (unpublished); A.S. Holevo, Probl. Inf. Transm. **9**, 177 (1973).