Fortschr. Phys. **51**, No. 2–3, 150 – 156 (2003) / **DOI** 10.1002/prop.200310019

Quantum measurement and information

Z. Hradil* and **J. Řeháček**

Department of Optics, Palacký University, 17. listopadu 50, 77200 Olomouc, Czech Republic

Received 10 June 2002, accepted 19 June 2002
Published online 25 February 2003

PACS 03.65.Ta, 03.65.Wj, 03.67.-a, 89.70.+c

The operationally defined invariant information introduced by Brukner and Zeilinger is related to the problem of estimation of quantum states. It quantifies how the estimated states differ in average from the true states in the sense of Hilbert–Schmidt norm. This information evaluates the quality of the measurement and data treatment adopted. Its ultimate limitation is given by the trace of inverse of Fisher information matrix.

1 Introduction

Information is turning out to be the key issue of our civilization. Though its undisputed significance the term information is used in a rather ambiguous way. The first attempts to quantify the content of information are related to the concept of entropy in thermodynamics. Mathematical basis for rigorous approach comes from Shannon, who introduced Shannon entropy and mutual information. These concepts appeared to be very useful in communication and information science triggering also a new direction in quantum theory-quantum information. However, information is not solely related to the Shannon information. Recently a project has been undertaken to build up the kinematics of the quantum theory from the information-theoretical principles [1–4]. As a part of the project a new measure of the classical information gained from a measurement on a quantum system was introduced. This quantity summed over a complete set of mutually complementary observables (MCO) exhibits invariance with respect to unitary transformations applied to the state of the system and/or to the measured set of MCO. Moreover, when properly normalized, it also quantifies the (maximum) information content, and therefore evaluates the processing power of physical systems.

A nice feature of the invariant information of Brukner and Zeilinger is that its definition is operational. It is obtained by synthesizing the errors of a specially chosen set of measurements performed on the system. In this contribution, we analyze the invariant information from a different perspective. If certain observations are made on the system the obtained results can be in a natural way put together to form our estimate of the quantum state of the system. This hints on the existence of a tight link between the information gained from a particular set of measurements and the error of the reconstruction based on the obtained results. Being motivated by the estimation theory we will show how to synthesize information gained from individual measurements in more general situations, namely, when (i) a complete set of MCO is not available, and when (ii) the observables measured are not necessarily mutually complementary (still non-commutative but not “maximally” non-commutative). Special attention will be paid to the invariance properties of the total information gained from a measurement. We will also show an alternative interpretation of the invariant information and will discuss the role that MCO play in the state estimation.

* Corresponding author E-mail: hradil@risc.upol.cz

Let us consider a measurement of an observable $A = \sum_{j=1}^{p} a_j \Pi_j$ having a non-degenerate spectrum acting in the Hilbert space of dimension p. Each outcome j is detected with the probability $p_j = \mathrm{Tr}\bar{\rho}\Pi_j$, where $\bar{\rho}$ describes the ensemble of particles arriving at the apparatus. Brukner and Zeilinger's lack of information associated with such a measurement done on N particles is defined as a sum of variances of individual outcomes per particle, $E_A = \sum_j \sigma_j^2/N$ [2]. The total lack of information about the system is then obtained as a sum of those measures over a complete set of MCO,

$$E = \sum_{\alpha j} \sigma_{\alpha j}^2/N = \sum_{\alpha j} p_{\alpha j}(1 - p_{\alpha j}), \tag{1}$$

where $\alpha = 1 \dots p{+}1$ and $j = 1 \dots p$ label complementary observables and their eigenvectors, respectively. The invariant information is nothing else than a properly normalized complement of the total lack of information E [2]. Assume that a complete set of MCO exists for the given dimension p, which means there are $p(p+1)$ projectors $\Pi_{\alpha j}$ satisfying [5, 6]

$$\mathrm{Tr}\{\Pi_{\alpha j}\Pi_{\beta k}\} = \delta_{\alpha\beta}\delta_{jk} + (1 - \delta_{\alpha\beta})/p. \tag{2}$$

One easily finds that the total error has already the desired invariance under the choice of the complete set of mutually complementary measurements, and under unitary transformation of the true state $\bar{\rho}$ of the system:

$$E = p - \mathrm{Tr}\bar{\rho}^2. \tag{3}$$

The main argument for just summing up the errors of complementary observations is that MCO are independent in the sense that a measurement of one of them gives no information about the rest [7]. If a particular observable α is measured, the Bayesian posterior distributions associated with the observables $\beta \neq \alpha$ complementary to it are flat. For this reason, MCO are sometimes called "unbiased observables" [8]. As will be shown in the following such a classical interpretation of MCO variables should be modified.

2 Quantum estimation theory

There is a close link between the total error E and the error of the standard quantum tomography. Instead of discussing the errors of measured projections separately, one can make a synthesis of all the results by forming an estimate of the unknown quantum state, and then evaluate the average error. Let us apply this strategy to the detection of a complete set of MCO. A generic quantum state may be decomposed in the basis of MCO as follows

$$\bar{\rho} = \sum_{\alpha j} \bar{w}_{\alpha j}\Pi_{\alpha j}, \tag{4}$$

where $\Pi_{\alpha j}$ are projectors on the eigenspaces of a complete set of MCO, the coefficients $\bar{w}_{\alpha j}$ being determined by the true probabilities $p_{\alpha j} = \mathrm{Tr}\{\bar{\rho}\Pi_{\alpha j}\}$ as follows [5, 10],

$$\bar{w}_{\alpha j} = p_{\alpha j} - \frac{1}{p+1}, \tag{5}$$

Suppose now that the complementary observations are done with N particles each. Due to fluctuations, the registered relative frequencies $f_{\alpha j}$ will generally differ from the true probabilities $p_{\alpha j}$ and so the estimated state

$$\rho = \sum_{\alpha j} w_{\alpha j}\Pi_{\alpha j}. \tag{6}$$

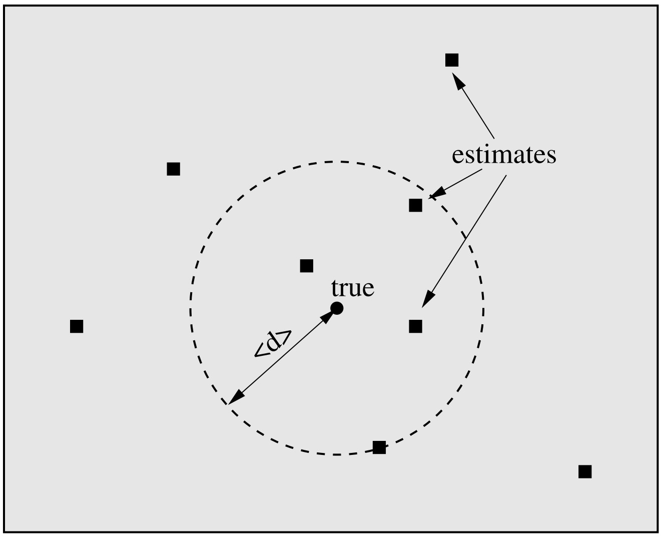

Fig. 1 The lack of information as the mean distance between the true and estimated states.

Let us see how much. Perhaps the most simple estimation strategy is the direct inversion based on the above mentioned independence of MCO,

$$w_{\alpha j} = f_{\alpha j} - \frac{1}{p+1}. \tag{7}$$

The error will be quantified by evaluating the Hilbert–Schmidt distance between the true and the estimated state

$$d = \mathrm{Tr}\{(\rho - \bar{\rho})^2\}. \tag{8}$$

The geometrical meaning is sketched in Fig. 1. The mean distance (error) is then given by averaging d over many repetitions of the estimation procedure, each yielding slightly different estimates of $\bar{w}_{\alpha j}$,

$$E_{\mathrm{est}} = \langle d \rangle. \tag{9}$$

Using Eqs. (4)–(8) and the definition of complementarity (2) in Eq. (9) we get,

$$\langle d \rangle = \sum_{\alpha j} \langle (\Delta w_{\alpha j})^2 \rangle = \frac{1}{N} \sum_{\alpha j} p_{\alpha j}(1 - p_{\alpha j}). \tag{10}$$

Notice that for the given number of input particles the mean distance of the estimated state from the true state is proportional to the total error (1), which depends on $\bar{\rho}$ only through $\mathrm{Tr}\bar{\rho}^2$. This shows an alternative interpretation of the total error E and the invariant information [9]: The total lack of the information as defined by Brukner and Zeilinger determines the mean error of the standard reconstruction based on the measurement of a complete set of MCO.

3 Fisher information

Given the same data, the accuracy of our estimate based on them strongly depends on the chosen reconstruction procedure. Direct inversion (7) is simple and straightforward because it implicitly uses the apparent statistical independence of MCO. But this is not the only possibility there. Keeping in mind that we want to characterize the information gained through the measurement one should use the best reconstruction method available. To evaluate the error of the optimal estimation procedure it is more convenient to decompose the true density matrix in a basis of *orthogonal* observables rather than MCO, $\bar{\rho} = 1/p + \sum_k \bar{a}_k \Gamma_k$, where the $p^2 - 1$ unknown parameters $\bar{a}_k$ provide a minimal representation of the state $\bar{\rho}$, and Γ_j are orthonormal: $\mathrm{Tr}\{\Gamma_j \Gamma_k\} = \delta_{jk}$. When the measurement is over, an estimate ρ of the true state is formed, $\rho = 1/p + \sum_k a_k(f_k)\Gamma_k$. The parameters a_k specifying the estimated state depend on the registered frequencies according to the given reconstruction procedure. Now we calculate the mean distance between the estimated and true states. Trivial calculation shows that it is given by a sum of the variances of the estimated parameters,

$$\langle d \rangle = \sum_k \langle (\Delta a_k)^2 \rangle. \tag{11}$$

To proceed, the variances appearing in (11) must somehow be determined. This may be a nontrivial problem because, generally, the estimated parameters a_k might depend on the measured frequencies f_k in a very complicated way. However, for our purpose, it is enough to evaluate the performance of the optimum estimation. It is known that the variances of estimated parameters cannot be less than the Cramer-Rao lower bound [11]. Further, it is known that maximum-likelihood estimators attain the bound asymptotically (for large N) [12]. In our case the Cramer-Rao bound reads:

$$\langle (\Delta a_k)^2 \rangle \geq [\mathbf{F}^{-1}]_{kk}, \tag{12}$$

where $\mathbf{F}$ is the Fisher information matrix defined as [13, 14]

$$\mathrm{F}_{kl} = \left\langle \frac{\partial}{\partial a_k} \log P(\mathbf{n}|\mathbf{a}) \frac{\partial}{\partial \mathbf{a_l}} \log \mathbf{P}(\mathbf{n}|\mathbf{a}) \right\rangle. \tag{13}$$

Here $P(\mathbf{n}|\mathbf{a})$ denotes the probability of registering data $\mathbf{n}$ provided the true state is $\mathbf{a}$, and the averaging is done over the registered data. Bound (12) may overestimate actual errors in some cases because we ignore the positivity constraint which the estimated ρ must obey. However, for large N, the probability of getting an unphysical estimate ρ from the measured data gets very small, and inequality (12) becomes a good error estimate. For large N, one is allowed to replace the multinomial distribution P by its Gaussian approximation. Keeping the notation of MCO registrations we have,

$$\log P(\mathbf{n}|\mathbf{a}) \approx -\mathbf{N^2} \sum_{\alpha \mathbf{j}} \frac{(\mathbf{f}_{\alpha \mathbf{j}} - \mathbf{p}_{\alpha \mathbf{j}})^{\mathbf{2}}}{\sigma^{\mathbf{2}}_{\alpha \mathbf{j}}}, \tag{14}$$

where $\sigma^2_{\alpha j} = N p_{\alpha j}(1 - p_{\alpha j})$ as before. Under this approximation the Fisher matrix becomes,

$$\mathrm{F}_{kl} = N^2 \sum_{\alpha j} \frac{\mathrm{Tr}\{\Gamma_k \Pi_{\alpha j}\} \mathrm{Tr}\{\Gamma_l \Pi_{\alpha j}\}}{\sigma^2_{\alpha j}}. \tag{15}$$

Consequently, the optimized operational information is given by the trace of the inverse of the Fisher matrix [9]

$$\langle d \rangle_{\mathrm{opt}} = \mathrm{Tr} \mathbf{F}^{-1}. \tag{16}$$

It does not depend on a particular choice of operators $\{\Gamma_j\}$ quantifying the maximum amount of information about an unknown state gained by the measurement performed.

It is interesting to notice that the definition (16) is similar to the definition of the reciprocal "intrinsic accuracy" of a multi-parameter estimation [14], which reads $e = 1/\mathrm{Tr}\mathbf{F}$. Although the two definitions look similar, their motivation and physical assumptions leading to them are different. In particular, the approach in [14] hinges upon the assumption of the independence of data, meaning that $\mathbf{F}$ is diagonal, which usually does not hold in quantum tomography. This is why $\mathbf{F}^{-1}$ rather than $\mathbf{F}$ itself plays fundamental role in our considerations.

4 Comparison

Based on this operational definition of information, the performances of various schemes may be easily compared. Now, we will go back to the measurement of a complete set of MCO. It is not difficult to see that the invariance of the error (10) of the direct inversion (7) stems from the fact that the projectors measured are considered as statistically independent observables. Each parameter a_k in the decomposition (6) is determined by measuring only one projector – the rest contributes nothing. As will be shown below, such data treatment cannot be optimal. This leads us to the following important question: Will the invariance of the estimation error survive if the direct inversion is replaced by the optimum data treatment? It turns out that the answer depends on the dimension p of the Hilbert space. In the simplest non trivial case of $p = 2$ the answer is in the affirmative. However for $p \geq 3$ a deviation appears. Let us have a look at the behavior of $\langle d \rangle_{\rm opt}$ for $p = 3$. Straightforward, but rather lengthy calculation of $\mathrm{Tr}\mathbf{F}^{-1}$ in this case yields:

$$\langle d \rangle_{\rm opt} = \frac{1}{N^2} \sum_{\alpha j} \sigma_{\alpha j}^2 - \frac{1}{N^2} \sum_{\alpha} \frac{\sigma_{\alpha 1}^4 + \sigma_{\alpha 2}^4 + \sigma_{\alpha 3}^4}{\sigma_{\alpha 1}^2 + \sigma_{\alpha 2}^2 + \sigma_{\alpha 3}^2}. \tag{17}$$

Notice that the first expression on the right-hand side is exactly the invariant total error (10). The second term then quantifies the improvement of the optimum estimation upon the simple linear inversion (7). One can easily check that this term is not an invariant quantity. So it turns out that the error of the optimum estimation from the measurement of a complete set of MCO is not invariant with respect to unitary transformations. This means that the maximum amount of information about the true state that can be gained from such measurements might differ even for true states of the same degree of purity! One pays for the optimality of the reconstruction procedure by loosing its invariant character (universality). Or conversely, one can have an invariant reconstruction at the expense of loosing precision. This behavior is plotted in Fig. 2 for several mixed quantum states generated randomly. The invariant result (3) represents the worst resolution corresponding to the upper edge of the plotted region.

Let us now address the broader aspects of optimal information processing discussed above. It seems that the non-invariant character of $\langle d \rangle_{\rm opt}$ originates from the "inhomogeneity" of the measured quorum of MCO. Two projectors drawn from the quorum can be either "complementary", $\mathrm{Tr}\{\Pi_1 \Pi_2\} = 1/p$, or orthogonal, $\mathrm{Tr}\{\Pi_1 \Pi_2\} = 0$, to each other. Consequently, the mesh established by MCO is not as regular as it might have seemed at the beginning. Provided that this is ignored, the observations may be regarded as "independent" on average, but such treatment is obviously not optimal. Some additional information can be gained provided those observations are treated as noise dependent observations [16].

For the sake of comparison let us evaluate resolution of several schemes which may be adopted for quantum state estimation. Optimal measurement would consist in measuring the "unknown" quantum state

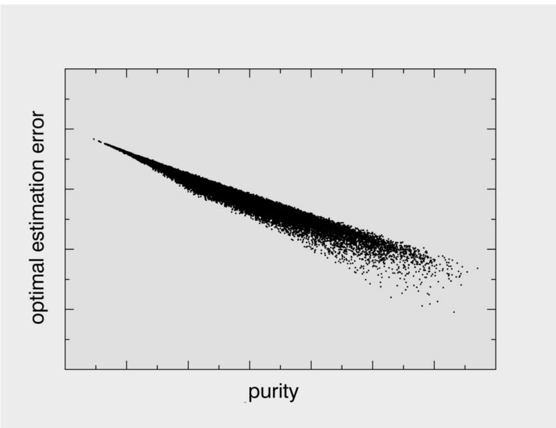

Fig. 2 The optimal estimation error for $p = 3$ as a function of purity.

along its eigenvectors, $\bar{\rho}\Pi_j = \lambda_j \Pi_j$, representing mutually exclusive outcomes due to their orthogonality – maximum likelihood estimation would always coincide with the deterministic data inversion. This gives the minimal achievable value of the average error: $NE_{\rm opt} = \sum_j \lambda_j(1-\lambda_j) = 1 - {\rm Tr}\bar{\rho}^2$. It should be clear that in the case of a generic quantum state whose diagonalizing basis is unknown and should be estimated together with $\{\lambda_j\}$, the optimal information can only be attained asymptotically as N goes to infinity. This optimal quantum reconstruction scheme provides invariant estimation errors.

This may be compared with the performance of MCO scheme. Realizing that total number of particles in the relation (3) is $(p+1)$ times larger than for single observable, the average error reads

$$NE = p(p+1) - (p+1){\rm Tr}\bar{\rho}^2. \tag{18}$$

For comparison let us consider simple measurement which is complete in the sense of completeness relation but not informationally complete [17]. Using the closure relation

$$\sum_i^p |e_i\rangle\langle e_i| = \hat{1}_p, \tag{19}$$

the diagonal elements of density matrix only could be inferred. Here the kets $|e_i\rangle$ represent an orthonormal basis on the p-dimensional Hilbert space. Straightforward calculation gives the average error

$$NE = N\sum_{i\neq j} |\rho_{ij}|^2 + 1 - \sum_i \bar{\rho}_{ii}^2, \tag{20}$$

which is nearly independent of the number of particles used for the measurement. This is due to the fact that nondiagonal elements are not registered. An informationally complete measurement will be obtained by considering also all the pairs (i,j) in the closure relations

$$\tfrac{1}{2}(|e_i\rangle + |e_j\rangle)(\langle e_i| + \langle e_j|) + \tfrac{1}{2}(|e_i\rangle - |e_j\rangle)(\langle e_i| - \langle e_j|) = \hat{1}_{ij} \tag{21}$$

$$\tfrac{1}{2}(|e_i\rangle + i|e_j\rangle)(\langle e_i| - i\langle e_j|) + \tfrac{1}{2}(|e_i\rangle - i|e_j\rangle)(\langle e_i| + i\langle e_j|) = \hat{1}_{ij}, \tag{22}$$

where $i \neq j$. These relations determine the real and imaginary parts of nondiagonal elements ρ_{ij}, respectively. Consider now the following scheme for the detection: The diagonal elements are sampled according to the relation (19) with m particles, and the closure relations (21) and (22) are sampled with m_{ij} particles. The total number of particles used is $N = m + \sum_{ij} 2m_{ij}$. Straightforward calculation yields the following formula for the average error

$$d = \frac{1}{m}\sigma^2 + \sum_{i\neq j}\frac{1}{m_{ij}}\sigma_{ij}^2, \tag{23}$$

$$\sigma^2 = 1 - \sum_i \bar{\rho}_{ii}^2, \quad \sigma_{ij}^2 = \tfrac{1}{2}(\bar{\rho}_{ii} + \bar{\rho}_{jj}) - \tfrac{1}{4}(\bar{\rho}_{ii} + \bar{\rho}_{jj})^2 - \tfrac{1}{2}|\bar{\rho}_{ij}|^2. \tag{24}$$

The error optimized with respect to the parameters m, m_{ij} reads

$$NE = \left(\sigma + \sqrt{2}\sum_{i\neq j}\sigma_{ij}\right)^2. \tag{25}$$

Rough numerical estimations show that the average error is worse comparing to MCO (18). Further improvement may be achieved, since information about diagonal elements of density matrix may be inferred also from the measurements (21) and (22).

5 Conclusion

The invariant information introduced by Brukner and Zeilinger is related to the problem of the estimation of a quantum state. It quantifies how an estimated state differs on average from the true state in the sense of the Hilbert–Schmidt norm. It depends on the quality of the measurement and on the data treatment adopted. Provided that data are treated in optimal way, the amount of extracted information may achieve the ultimate limit given by Fisher information. This scheme my be used for the evaluation of the performance of the measurement.

Acknowledgements This work was supported by grants No. LN00A015 and J14/98 of the Czech Ministry of Education.

References

[1] A. Zeilinger, Found. Phys. **29**, 631 (1999).
[2] C. Brukner and A. Zeilinger, Phys. Rev. Lett. **83**, 3354 (1999).
[3] Č. Brukner and A. Zeilinger, Acta Phys. Slovaca **49**, 647 (1999).
[4] Č. Brukner, M. Zukowski, and A. Zeilinger, eprint arXiv:quant-ph/0106119.
[5] I. D. Ivanović, J. Phys. A Math. Gen. **14**, 3241 (1981).
[6] J. Schwinger, Quantum Mechanics, edited by B.-G. Englert (Springer, Berlin, 2001).
[7] W. K. Wootter and B. D. Fields, Ann. Phys. (N.Y.) **191**, 363 (1989); see also [8].
[8] To be precise, a measurement of one such an observable usually gives some information about the observables complementary to it because of the positivity constraint which the density operator must obey. However, all the possible values of those observables that are consistent with this constraint are equally likely; in this sense, MCO are unbiased.
[9] J. Řeháček and Z. Hradil, Phys. Rev. Lett. **88**, 130401 (2002).
[10] B.-G. Englert and Y. Aharonov, Phys. Lett. A **284**, 1 (2001).
[11] C. R. Rao, Bull. Calcutta Math. Soc. **37**, 81 (1945).
H. Cramér, Mathematical methods of statistics (Princeton University Press, 1946).
[12] R. A. Fisher, Proc. Camb. Philos. Soc. **22**, 700 (1925).
[13] C. W. Helstrom, Quantum Detection and Estimation Theory (Academic Press, New York, 1976).
[14] B. R. Frieden, Physics from Fisher Information: A Unification (Cambridge University Press, 1999).
[15] W. K. Wootters, Phys. Rev. D **23**, 357 (1981).
S. L. Braunstein and C. M. Caves, Phys. Rev. Lett. **72**, 3439 (1994).
[16] Z. Hradil, Phys. Rev. A **55**, R1561 (1997).
J. Řeháček, Z. Hradil, and M. Ježek, Phys. Rev. A **63**, 040303(R) (2000).
[17] Ch. A. Fuchs, eprint arXiv:quant-ph/0205039