

Invariant Information and Quantum State Estimation

J. Řeháček and Z. Hradil

Department of Optics, Palacky University, 17. listopadu 50, 772 00 Olomouc, Czech Republic
(Received 30 November 2001; published 14 March 2002)

The invariant information, introduced by Č. Brukner and A. Zeilinger [Phys. Rev. Lett. **83**, 3354 (1999)], is reconsidered from the point of view of quantum state estimation. We show that this quantity is directly related to the mean error of the standard reconstruction from the measurement of a complete set of mutually complementary observables. We give its generalization in terms of the Fisher information. Provided that the optimum reconstruction is adopted, the information loses its invariant character.

DOI: 10.1103/PhysRevLett.88.130401

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

Invariants are important concepts of physics, because referring to them, physicists build up theoretical models making it possible to explain consistently their observations. 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 [2] 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 noncommutative but not “maximally” noncommutative). 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.

Let us consider a measurement of an observable $A = \sum_{j=1}^p a_j \Pi_j$ having a nondegenerate spectrum acting in the Hilbert space of dimension p . Each outcome j is detected with the probability $p_j = \text{Tr} \bar{\rho} \Pi_j$, 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,

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

where $\alpha = 1, \dots, p + 1$ and $j = 1, \dots, p$ label complementary observables and their eigenvectors, respectively. The invariant information is nothing other 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]

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

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:

$$E = p - \text{Tr} \bar{\rho}^2. \quad (3)$$

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.

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:

$$\bar{\rho} = \sum_{\alpha j} \bar{w}_{\alpha j} \Pi_{\alpha j}, \quad (4)$$

where Π_{α_j} are projectors on the eigenspaces of a complete set of MCO, the coefficients \bar{w}_{α_j} being determined by the true probabilities $p_{\alpha_j} = \text{Tr}\{\bar{\rho}\Pi_{\alpha_j}\}$ as follows [5,9]:

$$\bar{w}_{\alpha_j} = p_{\alpha_j} - \frac{1}{p+1}. \quad (5)$$

Now suppose that the complementary observations are done with N particles each. Because of fluctuations, the registered relative frequencies f_{α_j} will generally differ from the true probabilities p_{α_j} and so the estimated state

$$\rho = \sum_{\alpha_j} w_{\alpha_j} \Pi_{\alpha_j}. \quad (6)$$

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

$$w_{\alpha_j} = f_{\alpha_j} - \frac{1}{p+1}. \quad (7)$$

The error will be quantified by evaluating the Hilbert-Schmidt distance between the true and the estimated states [10]

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

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}_{α_j} ,

$$E_{\text{est}} = \langle d \rangle. \quad (9)$$

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

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

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 $\text{Tr}\bar{\rho}^2$. This is one of the key results of this contribution. It shows an alternative interpretation of the total error E and the invariant information: 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.

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, we 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: $\text{Tr}\{\Gamma_j \Gamma_k\} = \delta_{jk}$. 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,

$$\langle d \rangle = \sum_k \langle (\Delta a_k)^2 \rangle. \quad (11)$$

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 [13]

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

where \mathbf{F} is the Fisher information matrix defined as [14,15]

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

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. For large N , one is allowed to replace the multinomial distribution P by its Gaussian approximation. Keeping the notation of MCO registrations we have

$$\log P(\mathbf{n} | \mathbf{a}) \approx -N^2 \sum_{\alpha_j} \frac{(f_{\alpha_j} - p_{\alpha_j})^2}{\sigma_{\alpha_j}^2}, \quad (14)$$

where $\sigma_{\alpha_j}^2 = N p_{\alpha_j} (1 - p_{\alpha_j})$ as before. Under this approximation the Fisher matrix becomes

$$F_{kl} = N^2 \sum_{\alpha_j} \frac{\text{Tr}\{\Gamma_k \Pi_{\alpha_j}\} \text{Tr}\{\Gamma_l \Pi_{\alpha_j}\}}{\sigma_{\alpha_j}^2}. \quad (15)$$

The main formal result of the paper is obtained from Eqs. (11) and (12): Operational information defined as the mean error of the optimal estimation from the measurement of the chosen set of observables is given by the trace of the inverse of the Fisher matrix [16],

$$\langle d \rangle_{\text{opt}} = \text{Tr} \mathbf{F}^{-1}. \quad (16)$$

It quantifies the maximum amount of information about an unknown state gained by the measurement performed. Based on this information, one can easily compare the performances of various tomographic schemes. It is interesting to notice that the definition (16) is similar to the definition of the reciprocal ‘‘intrinsic accuracy’’ of a multiparameter estimation [15], which reads $e = 1/\text{Tr} \mathbf{F}$. Although the two definitions look similar, their motivation

and physical assumptions leading to them are different. In particular, the approach in [15] 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 a fundamental role in our considerations.

Equation (16) has the following simple geometrical interpretation: On registering counts \mathbf{n} , the probability $P(\mathbf{n} | \mathbf{a})$ characterizes the likelihood \mathcal{L} of various states \mathbf{a} . In terms of the Fisher matrix it reads

$$\log \mathcal{L} = \log P(\mathbf{n} | \mathbf{a}) \approx - \sum_{kl} (a_k - \tilde{a}_k)(a_l - \tilde{a}_l) F_{kl}, \quad (17)$$

where $\tilde{\mathbf{a}}$ specifies the maximum-likelihood solution. Let us define the error volume as the set of density matrices whose likelihoods do not drop below a certain threshold, $\mathcal{L} \geq \text{const}$. According to (17) the error volume is an ellipsoid, the lengths of its axes being inversely proportional to the eigenvalues of the Fisher matrix. The optimal estimation error $\langle d \rangle_{\text{opt}}$ can thus be interpreted as the sum of half-axes of the error ellipsoid which corresponds to the chosen set of measurement. From this point of view, the synthesis of all the quantum observations is equivalent to the registration of the orthogonal observables Γ' defining those axes. In this new representation the Fisher matrix attains the diagonal form, which means that the estimates of the transformed quantum-state “coordinates” $\tilde{a}'_k = \text{Tr}\{\tilde{\rho} \Gamma'_k\}$ fluctuate independently. Their statistical independence justifies adding their variances as it appears in Eq. (16).

This geometrical construction also shows that the structure of uncertainties related to the quantum tomography is richer than what a single number (16) might indicate. The chosen measurement might provide different resolutions in different directions, depending on the shape and orientation of the noise ellipsoid. Let us mention in passing that the resolution optimized over all measurements has been shown to provide a natural statistical distance between two quantum states [17].

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 nontrivial 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 $\langle d \rangle_{\text{opt}}$ for $p = 3$. Straightforward, but rather lengthy calculation of

$\text{Tr} \mathbf{F}^{-1}$ in this case yields

$$\langle d \rangle_{\text{opt}} = \frac{1}{N^2} \sum_{\alpha j} \sigma_{\alpha j}^2 - \frac{1}{N^2} \sum_{\alpha} \frac{\sigma_{\alpha 1}^4 + \sigma_{\alpha 2}^4 + \sigma_{\alpha 3}^4}{\sigma_{\alpha 1}^2 + \sigma_{\alpha 2}^2 + \sigma_{\alpha 3}^2}. \quad (18)$$

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 losing its invariant character (universality). Or conversely, one can have an invariant reconstruction at the expense of losing precision.

Let us now address the broader aspects of optimal information processing discussed above. It seems that the noninvariant character of $\langle d \rangle_{\text{opt}}$ originates from the “inhomogeneity” of the measured quorum of MCO. Two projectors drawn from the quorum can be either “complementary,” $\text{Tr}\{\Pi_1 \Pi_2\} = 1/p$, or orthogonal, $\text{Tr}\{\Pi_1 \Pi_2\} = 0$, 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 [18].

The invariance of the optimum estimation error can be restored provided a more “homogeneous” quorum of observables is measured. This can be illustrated on the thought measurement of the following observables:

$$\Pi_j = \frac{1}{2} + \frac{\sqrt{2}}{2} \Gamma_j, \quad j = 1, \dots, p^2 - 1, \quad (19)$$

where the normalization and the factor of $1/2$ ensure that $0 \leq \Pi_j \leq 1$, so that Π_j are actually elements of positive operator-valued measure which can be associated probabilities

$$p_j = \frac{1}{2} + \frac{\sqrt{2}}{2} \tilde{a}_j, \quad (20)$$

and \tilde{a}_j specify the true state as before. Because we now have $\text{Tr}\{\Pi_j \Gamma_k\} = \delta_{jk}$, the Fisher matrix (15) assumes the diagonal form. Tracing its inverse we get an expression resembling the invariant total error (1), $\langle d \rangle_{\text{opt}} = 2 \sum_j \sigma_j^2 / N^2$. In terms of density matrices the mean error of the optimal estimation from orthogonal observations reads

$$N \langle d \rangle_{\text{ortho}} = \frac{p^2 - 1}{2} + \frac{1}{p} - \text{Tr} \tilde{\rho}^2. \quad (21)$$

This is, obviously, an invariant quantity. However, one pays for the invariance by having much larger uncertainty about the state of the quantum system; even the simple direct inversion from MCO measurements is better if $p > 2$; see Eq. (3)—indeed, measurements (19) are rather noisy if $p > 2$.

Finally, let us mention that optimal quantum reconstruction schemes provide invariant estimation errors. Such an optimal measurement would consist in measuring the “unknown” quantum state 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: $NE_{\text{opt}} = \sum_j \lambda_j(1 - \lambda_j) = 1 - \text{Tr}\bar{\rho}^2$. 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 be attained only asymptotically as N goes to infinity.

To conclude, we have shown that the invariant information introduced by Brukner and Zeilinger is related to the problem of the estimation of a quantum state. It quantifies how the 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. Usually, the measurement of a complete set of mutually complementary observables is assumed. In this case, the measured state can be easily found via a linear inversion. The average distance between the estimated and true state is given directly by the invariant information.

The amount of extracted information may be increased provided that data are used in an optimal way achieving the ultimate limit given by (the inverse of) the Fisher information. When doing this, the unitary invariance is sacrificed in favor of the accuracy of estimation. Consequently, some “directions” can be identified easier than another. This seems to be just an example of the interplay between the universality and effectiveness of a particular method.

Finally, we touched the concept of mutually complementary observables. One advantage of such observations is that the data can be more easily processed to get information about the signal via the direct inversion. This is possible, unitary invariant but not efficient—one can do something better. The inefficiency of such an approach can be recognized in the structure of the Brukner-Zeilinger information where the variances of the measured projectors are simply added regardless of their commutativity or noncommutativity. This is not optimal. In the light of this, the often discussed issue of the existence or non-existence of a complete set of mutually complementary observables in some dimensions seems to be just an academic

question without essential consequences for the problem of the state determination and information in the above-mentioned sense.

We thank Č. Brukner, A. Zeilinger, and B.-G. Englert for pertinent comments. This work was supported by Grants No. LN00A015 and No. J14/98 of the Czech Ministry of Education, and by the East-West program of the Austrian Government.

-
- [1] A. Zeilinger, *Found. Phys.* **29**, 631 (1999).
 - [2] Č. 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, e-print arXiv:quant-ph/0106119.
 - [5] I. D. Ivanović, *J. Phys. A* **14**, 3241 (1981).
 - [6] J. Schwinger, *Quantum Mechanics*, edited by B.-G. Englert (Springer-Verlag, Berlin, 2001).
 - [7] W. K. Wootters 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] B.-G. Englert and Y. Aharonov, *Phys. Lett. A* **284**, 1 (2001).
 - [10] For its Wigner representation, see U. Leonhardt, *J. Mod. Opt.* **44**, 2271 (1997).
 - [11] C. R. Rao, *Bull. Calcutta Math. Soc.* **37**, 81 (1945); H. Cramér, *Mathematical Methods of Statistics* (Princeton University Press, Princeton, NJ, 1946).
 - [12] R. A. Fisher, *Proc. Cambridge Philos. Soc.* **22**, 700 (1925).
 - [13] 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.
 - [14] C. W. Helstrom, *Quantum Detection and Estimation Theory* (Academic Press, New York, 1976).
 - [15] B. R. Frieden, *Physics from Fisher Information: A Unification* (Cambridge University Press, Cambridge, U.K., 1999).
 - [16] Needless to say, this quantity does not depend on a particular choice of operators $\{\Gamma_j\}$.
 - [17] W. K. Wootters, *Phys. Rev. D* **23**, 357 (1981); S. L. Braunstein and C. M. Caves, *Phys. Rev. Lett.* **72**, 3439 (1994).
 - [18] Z. Hradil, *Phys. Rev. A* **55**, 1561(R) (1997); J. Řeháček, Z. Hradil, and M. Ježek, *Phys. Rev. A* **63**, 040303(R) (2001).