acta physica slovaca vol. 48 No. 3, 133 – 140 June 1998

OPTIMAL COMPRESSION OF QUANTUM INFORMATION FOR ONE-QUBIT SOURCE AT INCOMPLETE DATA: A NEW ASPECT OF JAYNES PRINCIPLE [1]

Michał Horodecki[2], **Ryszard Horodecki**[3]
Institute of Theoretical Physics and Astrophysics
University of Gdańsk, 80–952 Gdańsk, Poland

Paweł Horodecki[4]
Faculty of Applied Physics and Mathematics
Technical University of Gdańsk, 80–952 Gdańsk, Poland

Received 22 May 1998, accepted 26 May 1998

We consider the problem of optimal processing of quantum information at incomplete experimental data characterizing quantum source. In particular, we then prove that for one-qubit quantum source the Jaynes principle offers a simple scheme for *optimal* compression of quantum information. According to the scheme one should process *as if* the density matrix of the source were actually *equal* to the matrix of the Jaynes state.

The techniques of quantum teleportation [1, 2], entanglement purification [3] as well as compression of quantum information (QIC) [4, 5] exemplify a basic goal of the domain which is to understand the kind of channel resources needed for storing and transmission of *intact* quantum states. A natural question which arises in this context is processing of quantum information at *incomplete* experimental data [6]. As one knows, the celebrated scheme of statistical inference is given by the Jaynes principle [7]. The latter provides a procedure for a partial reconstruction of quantum states based on mean values $\bar{a}_i$ of some *incomplete* set of observables $\{A_i\}$ [8]

$$\overline{a}_i = \langle A_i \rangle = \mathrm{Tr}(\varrho A_i). \tag{1}$$

According to the principle the most probable (or representative) state ϱ_J maximizes von Neumann entropy [9]

$$S(\varrho) = -\mathrm{Tr}\varrho \ln \varrho \tag{2}$$

under the constraint (1).

[1] Special Issue on Quantum Optics and Quantum Information

[2] E-mail address: michalh@iftia.univ.gda.pl
[3] E-mail address: fizrh@univ.gda.pl
[4] E-mail address: pawel@mifgate.pg.gda.pl

In spite of great number [10] of applications of Jaynes principle its status as well as interpretation still remain unclear [11]. The principle is the most rational inference scheme in the sense that it does not permit to draw any conclusions unwarranted by the experimental data. However, this argument making the principle plausible does not actually prove it [11]. The difficulties in understanding of the Jaynes inference scheme are due to the fact that the latter is just a *principle* and it was not derived within the quantum formalism. Recently it has been shown [6] that the Jaynes inference is not universal, as it cannot be used in the case of entanglement processing. However, the Jaynes principle could seem to be a natural tool for QIC, as it is just von Neumann entropy which indicates the maximal degree of compression [4].

The motivation of the present paper was an attempt to understand the Jaynes principle on the basis of quantum information theory. The impetus to the present consideration was given by the important work of Schumacher [4] who first pointed out the physical interpretation of von Neumann entropy as the measure of quantum information in the context of QIC.

The main purpose of this paper is to investigate the connection between the Jaynes principle and the problem of compression of quantum information produced by the source characterized by incomplete data. We show that the entropy of the Jaynes state (the maximum entropy) is a basic bound for the rate of QIC at incomplete data. We also show that for one-qubit source the Jaynes principle provides a scheme which offers *optimal* compression. According to the optimal protocol one should process as if the unknown density matrix of the ensemble of the source were just *equal* to the matrix of the Jaynes state.

To begin with, let us outline the problem of QIC [4, 5, 12]. Suppose we have a source generating state ϱ_i (called message) with probability p_i. The task is to transmit the states ϱ_i to receiver with asymptotically perfect fidelity by means of minimal number of 2-state quantum systems. The latter are called *qubits* and constitute basic units of quantum information. Alice, who is to compress the initial information represented by the states ϱ_i is allowed to operate over long sequences of input systems. After her compression procedure (which can be an arbitrary operation admitted within the quantum formalism) the emerging states are transformed onto qubits and sent to the receiver (Bob) who is to perform the inverse operation. To this end he flips the state of qubits again onto the systems identical to the ones emitted by the source, and performs decompression operation. Now the asymptotically faithful transmission means that the input states obtained by Bob are on average close to the states of input sequences provided the latter are sufficiently long. The closeness is quantified by means of fidelity of the form [13]

$$F(\varrho_{in}, \varrho_{out}) = \left[\mathrm{Tr}\sqrt{\sqrt{\varrho_{in}}\varrho_{out}\sqrt{\varrho_{in}}}\right]^2 . \tag{3}$$

If the input state is pure ($\varrho_{in} = |a_{in}\rangle\langle a_{in}|$) then the fidelity takes the familiar form $F(\varrho_{in}, \varrho_{out}) = \langle a_{in}|\varrho_{out}|a_{in}\rangle$. In this case F can be interpreted as probability that the output state ϱ_{out} passes the test of being the state ϱ_{in}. The overall scheme of

compression-decompression protocol is the following

$$\varrho_{i_1}\otimes\ldots\otimes\varrho_{i_N} \xrightarrow[\Lambda_A]{\text{Alice's compression}} \tilde{\varrho}_{i_1,\ldots,i_N} \xrightarrow{\substack{\text{transmission}\\ \text{by means of qubits}}} \tilde{\varrho}_{i_1,\ldots,i_N} \xrightarrow[\Lambda_B]{\text{Bob's decompression}} \varrho^{out}_{i_1,\ldots,i_N} \tag{4}$$

with the condition

$$\lim_{N\to\infty}\sum_{i_1,\ldots i_N} p_{i_1}\ldots p_{i_N} F(\varrho_{i_1}\otimes\ldots\otimes\varrho_{i_N},\varrho^{out}_{i_1,\ldots,i_N})=1. \tag{5}$$

Thus the average fidelity must tend to 1 for sufficiently long input sequences. Now the basic problem is to find the protocol with minimal number of qubits per message needed to carry the ensemble of states $\tilde{\varrho}_{i_1,\ldots,i_N}$. In other words, the dimension of the Hilbert space $\mathcal{H}_{\tilde{\varrho}}$ spanned by the eigenvectors of the total density matrix $\tilde{\varrho}$ of the ensemble should be as small as possible. Then also the needed number R of qubits per message given by

$$R=\lim_{N\to\infty}\frac{1}{N}\log\dim\mathcal{H}_{\tilde{\varrho}} \tag{6}$$

will take the minimal value.

The outlined problem of QIC was first raised by Schumacher [4]. For ensemble of pure states he showed that it is possible to reduce the needed number of qubits R to the value of the von Neumann entropy of the total density matrix of ensemble $\varrho=\sum_i p_i\varrho_i$. The proposed protocol was then simplified by Jozsa and Schumacher [5] (we will refer to it as SJ protocol). Later on, Barnum *et al.* [12] showed that any possible compression protocol cannot compress the signal better than the SJ protocol. Thus for ensemble of pure states we have

$$R_{\min}=S(\varrho). \tag{7}$$

For ensemble of mixed states the problem is more complicated and in general remains still open [14, 15].

Let us now briefly recall the SJ compression scheme. Here the Alice's operation goes as follows. First, she subjects the initial sequence of states to a measurement with two outcomes $0,1$ corresponding to some projectors P and $P^{\perp}=I-P$ respectively. Obtained outcome 1 she does nothing else, otherwise (i.e. if an "error" occurred) she replaces the resulting state of sequence of systems with some arbitrarily established state $|0\rangle\langle 0|$ where $|0\rangle$ belongs to the subspace $\mathcal{H}$ determined by the projector P. After such operation the resulting ensemble lies solely within the subspace $\mathcal{H}$ and the needed number of qubits to carry it is equal to $\log\dim\mathcal{H}$.

Now there is fidelity lemma [5] which says that for any projector P if the probability of error

$$p=\mathrm{Tr}\varrho^{\otimes N}P^{\perp},\quad \varrho^{\otimes N}=\underbrace{\varrho\otimes\ldots\otimes\varrho}_{N} \tag{8}$$

asymptotically vanishes then the condition of faithful transmission (5) is fulfilled with Bob decompression being trivial (he needs do nothing apart from flipping the signal from qubits onto systems identical with the ones emitted by the source) [16].

Moreover, the eigenvalues of ϱ can be divided into two parts: an amount of approximately $2^{NS(\varrho)}$ typical eigenvalues carrying almost all "weight" of the matrix ϱ and the remaining eigenvalues (atypical) the sum of which vanishes for large N. The subspace $\mathcal{H}_t$ spanned by the eigenvectors corresponding to the typical eigenvalues is called typical one. Now in the SJ protocol the projector P is chosen to project onto the typical subspace. Then, by the fidelity lemma, the faithful transmission is possible, and the signal is compressed down to the value of $S(\varrho)$ qubits per message (as $\dim \mathcal{H}_t =$ the number of typical eigenvalues $\approx 2^{NS(\varrho)}$).

Consider now the case of incomplete data. Namely, suppose that Alice (who is to compress the signal states) knows neither the states ϱ_i generated by the source nor the probabilities p_i. Instead, let she know mean values a_i of some incomplete set of observables A_i measured on a large subensemble of the systems produced by the source. As the set is incomplete, Alice is not able to recover the density matrix of the ensemble. Suppose now that she wants to compress the signal, basing on that incomplete information. However, there are many ensembles which are in agreement with the data. Then her strategy must be so clever that the Bob decompression could be faithful for *any* ensemble satisfying the data. The basic question is: what is the maximal compression rate which allow for faithful decompression if only incomplete data are measured? So far, in the problem of QIC the form of the ensemble generated by the source was supposed to be known, hence the maximal compression rate was a function of the ensemble. Here, the only characteristics of the source is contained in the measured data, so that the maximal rate (or its bounds) is a function of the observables A_i and the mean values $\bar{a}_i$.

Note first that the basic limit for the compression rate at incomplete data can be found by means of the Jaynes principle: the minimal number of qubits *cannot* be lower than the maximum entropy

$$R_{\min}(\{A_i; \bar{a}_i\}) \geq S_J. \tag{9}$$

where $S_J = S(\varrho_J)$. Indeed, the actual ensemble of the source could have its density matrix just equal to the Jaynes one (as the latter is in agreement with the data by definition). It could also consist of pure states, as the mean values of observables say nothing about components of ensemble. Then according to the the mentioned result of Barnum *et al.* [12], any protocol which compresses the signal to the value less than the maximum entropy does not allow for faithful decompression.

Here a very natural question arises: is it that the minimal number of qubits per message is in fact *equal* to the Jaynes entropy? Below we will show that in the case of one-qubit source the answer is "yes". The bound (9) will be reached by a scheme (we will call it Jaynes compression) according to which Alice and Bob apply to the ensemble the SJ protocol *as if* its density matrix were *equal* to the Jaynes state. We will show that for one-qubit source satisfying the data the Jaynes compression allows for faithful decompression. Thus, Alice and Bob can faithfully process, imaging that the real state is the Jaynes one, even if in fact it is not the case!

Suppose that Alice has measured only one (nondegenerate) observable A and obtained mean value $\bar{a}$. We will show that the optimal compression is provided by the

Jaynes scheme. For this purpose write the spectral decomposition of the observable

$$A = a_1|v\rangle\langle v| + a_2|w\rangle\langle w|, \tag{10}$$

where a_i are eigenvalues and $|v\rangle, |w\rangle$ are eigenvectors. Let us write the density matrix ϱ of input ensemble write in the basis $|v\rangle, |w\rangle$

$$\varrho = \varrho_{11}|v\rangle\langle v| + \varrho_{12}|v\rangle\langle w| + \varrho_{21}|w\rangle\langle v| + \varrho_{22}|w\rangle\langle w|. \tag{11}$$

The diagonal elements of ϱ can be expressed in terms of the mean value $\bar{a}$ and eigenvalues λ_1, λ_2 as follows

$$\varrho_{11} = \frac{\bar{a} - a_2}{a_1 - a_2}, \quad \varrho_{22} = 1 - \varrho_{11} = \frac{a_1 - \bar{a}}{a_1 - a_2}. \tag{12}$$

Note that density matrices satisfying the constraint $\langle A\rangle = \bar{a}$ can differ from each other only by off-diagonal elements. As one knows [11] discarding the off-diagonal elements cannot decrease entropy, so that the Jaynes state ϱ_J (which has maximal entropy) must be equal to

$$\varrho_J = \varrho_{11}|v\rangle\langle v| + \varrho_{22}|w\rangle\langle w|. \tag{13}$$

Hence ϱ_{11} and ϱ_{22} are eigenvalues of ϱ_J.

Compare now the density matrix $\varrho^{\otimes N}$ of ensemble of sequences of signal states and the N-fold tensor product of the Jaynes matrix $\varrho_J^{\otimes N}$. The latter one has eigenvalues equal to the diagonal elements of the former one, hence for any projector P onto the subspace spanned by any collection of eigenvectors of $\varrho_J^{\otimes N}$, we have

$$\mathrm{Tr}\varrho_J^{\otimes N} P = \mathrm{Tr}\varrho^{\otimes N} P. \tag{14}$$

The above equality says that the probability of error for any ensemble satisfying the data is equal to the probability of error for the ensemble with density matrix ϱ_J. Now, if Alice performs the measurement by means of projector onto typical subspace of the state $\varrho_J^{\otimes N}$ then by virtue of the fidelity lemma the faithful transmission is possible for *any* ensemble satisfying the data. Thus in this case we have

$$R_{\min}(A; \bar{a}) = S_J. \tag{15}$$

The result incorporates the case of ensemble of mixed states as such ensemble can also be compressed by means of SJ protocol [14].

Let us now analyse the case when Alice knows mean values of two observables A, B

$$\mathrm{Tr}(\varrho A) = \bar{a}, \quad \mathrm{Tr}(\varrho B) = \bar{b}. \tag{16}$$

Let us write the observables in eigenbasis $|v\rangle, |w\rangle$ of A

$$A = \begin{bmatrix} a_1 & 0 \\ 0 & a_2 \end{bmatrix}, \quad B = \begin{bmatrix} b_1 & c \\ c & b_2 \end{bmatrix}, \tag{17}$$

where the relative phase of the base vectors is chosen so that c is real (of course a_i and b_i are real due to hermiticity of A and B). Applying the constraint (16) we see that the most general form of ϱ is

$$\varrho = \begin{bmatrix} \varrho_{11} & d + i\gamma \\ d - i\gamma & \varrho_{22} \end{bmatrix}, \tag{18}$$

where the diagonal elements are of the form (12) and $d = \bar{b} - (b_1\varrho_{11} + b_2\varrho_{22})/(2c)$ so that the only free parameter is γ (due to positivity of ϱ γ must satisfy inequality $\gamma^2 \leq \varrho_{11}\varrho_{22} - d^2$). The eigenvalues of ϱ are the closest together (hence ϱ has the largest entropy) if $\gamma = 0$ hence the Jaynes state is of the form

$$\varrho_J = \begin{bmatrix} \varrho_{11} & d \\ d & \varrho_{22} \end{bmatrix}. \tag{19}$$

Then we obtain

$$\varrho = \varrho_J + i\gamma \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}. \tag{20}$$

Now, the matrix ϱ_J has real entries, so that it can be diagonalized by a real rotation. Note that the matrix in the second term of the equation is rotation by $\pi/2$). As the rotations on the plane commute, if we apply the ones diagonalizing ϱ_J to the matrix ϱ, the second term in (20) will not be affected. Then, in the eigenbasis of ϱ_J, both ϱ and ϱ_J have the same diagonal elements. Thus, we can apply the same argument as in the case of one observable data so that

$$R_{\min}(A, B; \bar{a}, \bar{b}) = S_J. \tag{21}$$

As a matter of fact, the above analysis completes the proof that the Jaynes compression allows faithful transmission of quantum information. In fact, suppose Alice knows mean values of three observables A, B, C. If they are linearly independent, they constitute complete data, so that the state of ensemble is uniquely determined. If, instead, one of them (e.g. C) can be written as linear combination of the others, then the mean value of C is determined by the means of A and B so that we turn back to the case of two observables. To be careful, one should also consider the case of the so-called generalized observables (Positive Operator Valued measures) [17]. It appears [18], that in the case of one-qubit source this can be easily reduced to the case of ordinary von Neumann observables.

In conclusion, we have shown that the Jaynes principle puts bound for maximal compression rate. Moreover, for one-qubit source it provides a very simple scheme of *optimal* degree of compression. To obtain it, one should process as if the density matrix of the source were actually equal to the Jaynes matrix. The results shed new light on the status of the Jaynes principle, as they allow to hope that from the point of view of quantum information theory the principle seems to be a consequence of quantum formalism rather than an external postulate. In fact, we have revealed a remarkable alternative: either the Jaynes principle can be derived as a *theorem* for quantum information theory, or its meaning for this field is not so profound as one

could expect. Indeed, if the Jaynes compression did not work well in general, then it would mean that the Jaynes inference scheme fails to play the most natural role that can be found for it within quantum information theory.

Our results suggest also a general question concerning quantum information processing at incomplete data. Namely, note that the scheme we used here (the Jaynes compression) consisted of two basic stages

(i) the estimated form of state is produced by means of the Jaynes principle.

(ii) the compression protocol is chosen as if the *actual* density matrix of the ensemble were *equal* to the Jaynes state.

Suppose now that we have some different task than QIC (e.g. we need to distill entanglement). Then the question is whether the above approach will work in this general case. We then would have the following steps.

(i) the estimated form of state is produced by means of an inference scheme

(ii) the suitable protocol is chosen as if the *actual* density matrix were *equal* to inferred one.

The inference scheme cannot be in general the Jaynes one but it must rather depend on the kind of task. Indeed, it was shown [6] that in the case of entanglement processing the Jaynes inference fails as it can produce *inseparable* (entangled) state although there exist *separable* (disentangled) ones consistent with data. There is an open question, whether the proposed general approach provides faithful and optimal information processing.

Acknowledgements The authors would like to thank Richard Jozsa for helpful comments, stimulating discussion and simplifying the proof for two-observable case. They are also grateful to Armen Allahverdyan and Chris Fuchs for valuable comments. M. H. and P. H. would like to acknowledge the support by Foundation for Polish Science.

Note added: Quite recently, the general solution of the problem of compression of quantum information at incomplete data has been provided in the paper "*Universal compression of quantum information*" by Richard Jozsa and the present authors [19]. It follows that the equality $R_{\min} = S_J$ is true in general. The result is obtained *via* different approach than the one presented here.

References

[1] C. Bennett, G. Brassard, C. Crepeau, R. Jozsa, A. Peres, W. K. Wootters: *Phys. Rev. Lett.* **70** (1993) 1895

[2] D. Bouwmeester, J.-W. Pan, K. Mattle, M. Elbl, H. Weinfurter, A. Zeilinger: *Nature (London)* **390** (1997) 575 ; D. Boschi, S. Brance, F. de Martini, L. Hardy, S. Popescu: *Phys. Rev. Lett.* **80** (1998) 1121

[3] C. H. Bennett, G. Brassard, S. Popescu, B. Schumacher, J. Smolin, W. K. Wootters: *Phys. Rev. Lett.* **76** (1996) 722; C. H. Bennett, D. P. Di Vincenzo, J. Smolin, W. K. Wootters: *Phys. Rev. A* **54** (1997) 3814; M. Horodecki, P. Horodecki, R. Horodecki: *Phys. Rev. Lett.* **78** (1997) 574.

[4] B. Schumacher: *Phys. Rev. A* **51** (1995) 2738

[5] R. Jozsa, B. Schumacher: *J. Mod. Opt.* **41** (1994) 2343

[6] R. Horodecki, M. Horodecki, P. Horodecki: *Los Alamos e-print quant-ph/9709010*

[7] E. Jaynes: *Phys. Rev.* **108** (1957) 171; *ibid* **108** (1957) 620; Am. J. Phys. **31** (1963) 66

[8] By complete set of observables one means the maximal set of linearly independent observables (where the trivial observable represented by identity operator is excluded). A set which does not fulfil the above conditions is called incomplete one.

[9] Throughout this paper we will use the base-2 logarithm rather than natural one.

[10] See e.g. V. Bužek, G. Drobný, G. Adam, R. Derka, P. L. Knight: *J. Mod. Opt.* **44** (1997) 2607; for an extensive presentation of the use of Jaynes principle see W. T. Grandy, *Am. J. Phys* **65** (1997) 466

[11] A. Wehrl: *Rev. Mod. Phys.* **50** (1978) 221

[12] H. Barnum, Ch. Fuchs, R. Jozsa, B. Schumacher: *Phys. Rev. A* **54** (1996) 4707

[13] A. Uhlmann: *Rep. Math. Phys.* **9** (1976) 273; R. Jozsa: *J. Mod. Opt.* **41** (1994) 2315

[14] R. Jozsa (unpublished); Hoi-Kwong Lo: *Opt. Commun.* **119** (1995) 552; A. E. Allahverdyan, D. B. Saakian: *Los Alamos e-print quant-ph/9702034*

[15] M. Horodecki: *Phys. Rev. A* **57** (1998) 3348; H. Barnum, C. Caves, Ch. Fuchs, R. Jozsa, B. Schumacher (unpublished).

[16] The original formulation of lemma in Ref. [5] is slightly different: the projection P is supposed to project onto a subset of eigenvectors of $\varrho^{\otimes N}$. However, to prove the lemma one does not need to make such an assumption.

[17] K. Kraus: *States, Effects and Operations: Fundamental Notions of Quantum Theory* (Wiley, New York, 1991)

[18] R. Jozsa: private communication.

[19] R. Jozsa, M. Horodecki, P. Horodecki, R. Horodecki: *Los Alamos e-print quant-ph/9805017*