Optimal compression of quantum information for one qubit source at incomplete data

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

Pawel Horodecki***
Faculty of Applied Physics and Mathematics
Technical University of Gdańsk, 80–952 Gdańsk, Poland

We consider the problem of optimal processing of quantum information at incomplete experimental data characterizing the quantum source. We show that the Jaynes principle puts bound for maximal compression rate. We then prove that for one-qubit quantum source the principle offers a very simple scheme for optimal compression of quantum information.

Pacs Numbers: 03.65.Bz

The techniques of 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 \( a_i \) of some incomplete set of observables \( \{ A_i \} \) [8]

\[
\overline{a}_i = \langle A_i \rangle = \text{Tr}_\rho A_i.
\] (1)

According to the principle the most probable (or representative) state \( \rho \) maximizes von Neumann entropy

\[
S(\rho) = -\text{Tr} \rho \ln \rho
\] (2)

under the constraint (1).

In spite of great number [9] of applications of Jaynes principle its status as well as interpretation still remain unclear [10]. 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 [10]. 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. In fact, it was pointed out that the Jaynes inference is not universal, as it cannot be used in the case of entanglement processing [6] (hence also in the case of incomplete information about parameters of noisy quantum channels). 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 Letter 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 Letter is to investigate the connection between the Jaynes principle and the problem of compression of quantum information produced by the quantum source characterized by incomplete data. We show that the entropy of the Jaynes 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. This sheds new light on the status of the Jaynes principle. The latter, from the point of view of quantum information theory seems to be a consequence of quantum formalism rather than an external postulate.

To begin with let us outline the problem of QIC [4,5,11]. Suppose we have a source generating state \( \rho_i \) (called message) with probability \( p_i \). The task is to transmit the states \( \rho_i \) 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 \( \rho_i \), 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 emitted by the source.

Fidelity defined by means of mean values between states [12]

\[
F(\rho_{in}, \rho_{out}) = \left[ \text{Tr} \sqrt{\rho_{in} \rho_{out} \sqrt{\rho_{in}}^*} \right]^2.
\] (3)
If the input state is pure ($\varrho_{in} = |a_{in}\rangle\langle a_{in}|$) then the fidelity takes the form

$$F(\varrho_{in}, \varrho_{out}) = \langle a_{in}|\varrho_{out}|a_{in}\rangle.$$  \hfill (4)

In this case $F$ can be interpreted as probability that the output state $\varrho_{out}$ passes the test of being the state $\varrho_{in}$. The overall scheme of compression-decompression protocol is the following

$$\varrho_{i_1} \otimes \ldots \otimes \varrho_{i_N} \quad \text{Alice’s compression} \quad \Lambda_A \quad \tilde{\varrho}_{i_1, \ldots, i_N}$$

$$\rightarrow \quad \text{by means of qubits} \quad \tilde{\varrho}_{i_1, \ldots, i_N} \quad \text{Bob’s decompression} \quad \Lambda_B \quad \varrho_{out}$$

with the condition

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

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

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. [11] showed that any possible compression protocol cannot compress the signal better than the SJ protocol. Thus for ensemble of pure states we have

$$R_{\text{min}} = S(\varrho).$$  \hfill (6)

For ensemble of mixed states the problem is more complicated and in general remains still open [13,14]. As a matter of fact, for some cases one can compress the signal better that indicated by $S(\varrho)$. For our purposes it is enough to know that the ensemble of the mixed states can always be compressed to the value of the von Neumann entropy of ensemble by means of the SJ protocol [13] (although the compression may not be optimal).

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 $\ket{0}\bra{0}$ where $\ket{0}$ 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

$$p = \mathrm{Tr} \varrho \otimes N P^\perp,$$  \hfill (7)

asymptotically vanishes then the condition of faithful transmission (6) 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) [15]. Moreover, the eigenvalues of $\varrho$ can be divided into two parts: an amount of approximately $2^{N S(\varrho)}$ typical eigenvalues carrying almost all “weight” of the matrix $\varrho$ and the remaining eigenvalues (atypical) the sum of which vanishes for large $N$. The subspace $\mathcal{H}$ 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 $\dim \mathcal{H} = \text{the number of typical eigenvalues} \approx 2^{NS(\varrho)}$).

Consider now the case of incomplete data. Namely, suppose that Alice (who is to compress the signal states) knows neither the states $\varrho_i$ 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. Now 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 $a_i$.

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 entropy of the Jaynes state.
\( R_{\text{min}}(\{A;\bar{a}_i\}) \geq S_J. \) \hspace{1cm} (10)

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 to the mentioned result of Barnum et al. [11], any protocol which compresses the signal to the value less than the von Neumann entropy of the Jaynes state 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 entropy of the Jaynes state? Below we will show that in the case of one-qubit source the answer is "yes". The bound \[ \text{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.}

Suppose that Alice has measured only one (nondegenerate [16]) 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

\[ A = \lambda_1|v\rangle\langle v| + \lambda_2|w\rangle\langle w|, \] \hspace{1cm} (11)

where \( \lambda_i \) are eigenvalues. Let the density matrix \( \varrho \) of input ensemble write in the eigenbasis \( \{v, w\} \) of the observable \( A \) as

\[ \varrho = \varrho_{11}|v\rangle\langle v| + \varrho_{12}|v\rangle\langle w| + \varrho_{21}|w\rangle\langle v| + \varrho_{22}|w\rangle\langle w|. \] \hspace{1cm} (12)

The diagonal elements of \( \varrho \) can be expressed in terms of the mean value \( \bar{a} \) and eigenvalues \( \lambda_1, \lambda_2 \) as follows

\[ \varrho_{11} = \frac{\bar{a} - \lambda_2}{\lambda_1 - \lambda_2}, \quad \varrho_{22} = 1 - \varrho_{11} = \frac{\lambda_1 - \bar{a}}{\lambda_1 - \lambda_2}. \] \hspace{1cm} (13)

Note that density matrices satisfying the constraint \( \langle A \rangle = a \) can differ from each other only by off-diagonal elements. As one knows [10] discarding the off-diagonal elements cannot decrease entropy, so that the Jaynes state \( \varrho_J \) (which has maximal entropy) must be equal to

\[ \varrho_J = \varrho_{11}|v\rangle\langle v| + \varrho_{22}|w\rangle\langle w|. \] \hspace{1cm} (14)

Hence \( \varrho_{11} \) and \( \varrho_{22} \) are eigenvalues of \( \varrho_J \).

Compare now the density matrix \( \varrho_J^{\otimes N} \) 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

\[ \text{Tr}(\varrho_J^{\otimes N} P) = \text{Tr}(\varrho J^{\otimes N} P). \] \hspace{1cm} (15)

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 \( \varrho_J \). 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

\[ R_{\text{min}}(A; \bar{a}) = S_J. \] \hspace{1cm} (16)

The result incorporates the case of ensemble of mixed states as such ensemble can also be compressed by means of SJ protocol [13].

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

\[ \text{Tr}(\varrho A) = \bar{a}, \quad \text{Tr}(\varrho B) = \bar{b}. \] \hspace{1cm} (17)

Let us write the observables by means of Pauli matrices \( \{\sigma_i\} \):

\[ A = \alpha I + \bar{a} \sigma_x, \quad B = \beta I + \bar{b} \sigma_x \] \hspace{1cm} (18)

where \( \alpha, \beta, a, b \) are some real numbers, \( \bar{a}, \bar{b} \) and unit vectors in real three-dimensional space and \( \bar{a} \sigma_x = \sum_{i=1}^3 x_i \sigma_i \).

Consider first the case when the observables commute. Then we immediately get that \( \bar{a} \) is parallel to \( \bar{b} \) and it is easy to see that one of two mean values of (17) completely determines another. Hence the problem reduces to case of one-observable data.

Suppose then that \( A, B \) do not commute. Then both vectors \( \bar{a}, \bar{b} \) are not parallel, defining then some plane. It allows to represent the Bloch vector of any state \( \varrho \) satisfying the data in the three-dimensional basis: \( \hat{a}, \hat{c} = [\bar{b} - (\bar{a}, \bar{b})\bar{a}]\sqrt{1 - (\bar{a}, \bar{b})^2}, \hat{c} = \hat{a} \times \hat{c} \). It allows to write any matrix \( \varrho \) satisfying the data (17) in the form depending only one free parameter \( \gamma \):

\[ \varrho = \frac{1}{2} (I + r\hat{r} \sigma + \gamma \hat{c} \hat{\sigma}). \quad \gamma^2 \leq 1 - r^2 \] \hspace{1cm} (19)

The vector \( r\hat{r} \) is defined as

\[ r\hat{r} \equiv \frac{\bar{a} - \alpha}{\bar{a}} \hat{a} + \frac{1}{\sqrt{1 - (\bar{a}, \bar{b})^2}} \frac{\bar{b} - \beta}{\bar{b}} (\bar{a}, \bar{b})\hat{c} \] \hspace{1cm} (20)

and it contains all information given by mean values (17). Now we can use the rotation in real three-dimensional space \( O : \{\hat{r}, \hat{c} \times \hat{r}, \hat{c} \} \rightarrow \{\hat{z}, \hat{x}, \hat{y}\} \) and define the unitary operation \( U(\hat{O}) \) by means of well known homomorphism between \( SU(2) \) and \( O_3 \) groups which is constituted by the relation \( U(O)\kappa U(\hat{O}) = (\kappa \bar{\sigma}) \kappa \bar{\sigma} \) for any unit vector \( \kappa \). Finally, if we introduce the following one qubit basis in \( C^2 : v' = U(\hat{O})|0\rangle, w' = U(\hat{O})|1\rangle \)
then the matrix $\varrho$ takes the form and properties completely parallel to (12). In fact, then the diagonal elements are determined by (17) as they amount to $(1 \pm r)/2$ while off-diagonal elements $\gamma/2$ includes no information about the data, being then completely free. Thus the whole reasoning following the equation (13) can be performed to give

$$R_{\text{min}}(A; B; \bar{a}, \bar{b}) = S_f.$$  

(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 problem of two-observable data.

To summarize, we have shown that the Jaynes principle puts bound for maximal compression rate. Moreover, for one-qubit source the Jaynes principle 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 allow to hope that within the quantum information theory the Jaynes principle could be derived as a theorem. Moreover they suggest 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 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 scheme fails as it can produce inseparable state although there exist separable ones consistent with data. Then there is an open question, whether the above scheme provide faithful and optimal information processing.

The authors would like to thank Richard Jozsa for helpful comments and stimulating discussion. They are also grateful to Armen Allahverdyan and Chris Fuchs for valuable comments. M. H. and P. H. would like to acknowledge the financial support by Foundation for Polish Science.

[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.
[15] The original formulation of lemma in Ref. [5] is slightly different: the projection $P$ is supposed to project onto a subset of eigenvectors of $\rho^\otimes N$. However, to prove the lemma one does not need to make such an assumption.
[16] Of course degenerate observable in one-qubit case is proportional to identity hence it does not provide any information about the density matrix of the source.