

Quantum Rate-Distortion Coding

Howard Barnum

August 9, 2001

Hampshire College and ISIS
Amherst, MA 01002 USA
hbarnum@hampshire.edu

1 Introduction

The fidelity criteria introduced in noisy and noiseless coding theorems may seem excessively stringent. The classical criterion, for example, requires that the probability of an error in the entire block approach zero as the block length goes to infinity. A code with a constant nonzero error rate per symbol would fail this test miserably (error probability would go to one in the large block limit), but could still be perfectly acceptable as long as the error rate was sufficiently small. (Most, if not all, noisy channel coding protocols used with real-world communications channels are examples.) Similarly in the quantum mechanical case, we might be willing—taking an i.i.d. source for simplicity in the example— to tolerate a constant rate of bad EPR pairs in the entanglement- transmission case, or a finite deviation (“distortion”) of the average pure state fidelity of each transmission from one. A theory which tells us, given an “error rate” or level of distortion which we have decided we can tolerate, whether a given channel (noisy or noiseless) can achieve that error rate, would be decidedly useful. This is rate-distortion theory.

One might think one could get by with substantially less resources if one accepts the less ambitious fidelity criterion of requiring a constant distortion rate. Classical rate-distortion theory tells us that there is no great savings in allowing small average distortion rather than asymptotically perfect transmission. Thus rate-distortion theory helps establish the relevance

of theoretical results like the asymptotic block-coding versions of noiseless and noisy channel coding, to real-world schemes.

2 A quantum version of rate-distortion

Let us use as our measure of distortion either one minus the entanglement fidelity (for the entanglement transmission problem) or one minus the average pure-state fidelity (for pure-state ensemble transmission problems). This must be evaluated for single transmissions, and then averaged over the block of n transmissions. I will confine myself to i.i.d. sources, with marginal density operator ρ , at least initially. Thus $\rho^{(n)} \equiv \rho^{\otimes n}$. The channel will be taken to be noiseless; then a $(n, 2^{nR})$ *rate-distortion code* consists of a map $\mathcal{E}^{(n)}$ from n copies of the source space to n copies of a channel space of dimension 2^{nR} , followed by a decoding $\mathcal{D}^{(n)}$ from n channels to n source spaces. The average distortion for an i.i.d. source can then be defined as:

$$D_e(\mathcal{E}^{(n)}, \mathcal{D}^{(n)}) \equiv \sum_{i=1}^n \frac{1}{n} (1 - F_e(\rho, \mathcal{T}_i^{(n)})), \quad (1)$$

where \mathcal{T}_i is the “marginal operation” on the i -th copy of the source space induced by the overall operation $\mathcal{D}^{(n)} \circ \mathcal{E}^{(n)}$. More formally,

$$\mathcal{T}_i^{(n)}(\sigma) \equiv \text{tr}_{Q_1, \dots, Q_{i-1}, Q_{i+1}, \dots, Q_n} [(\mathcal{D}^{(n)} \circ \mathcal{E}^{(n)})(\rho \otimes \rho \cdots \otimes \rho \otimes \sigma \otimes \rho \cdots \otimes \rho)], \quad (2)$$

where the σ in the input density operator is in the i -th position. (It is easily checked that this defines a tracepreserving operation.) The same definition, but with $\overline{F}(E, \mathcal{T}_i^{(n)})$, as the fidelity criterion, defines the average pure-state distortion \overline{D} .

R is said to be the *rate* of a rate-distortion code. To avoid confusion, I note here that the rate of a rate-distortion code has a significance roughly inverse to that of the rate of information transmission through a noisy channel. (The terminology is already well-established in classical information theory.) The rate in rate-distortion is the rate at which the source is described, that is, the number of qubits, or the log of the number of Hilbert space dimensions, used to encode the source, per source emission. Thus the goal of rate-distortion theory is to achieve low rates, i.e. to encode the source into as few qubits as possible per source emission.

A rate-distortion pair (R, D) is *achievable* for a given source iff there is a sequence of $(n, 2^{nR})$ rate-distortion codes $(\mathcal{E}^{(n)}, \mathcal{D}^{(n)})$ such that

$$\lim_{n \rightarrow \infty} D(\mathcal{E}^{(n)}, \mathcal{D}^{(n)}) \leq D. \quad (3)$$

Here D is whatever average distortion measure is used, e.g. \bar{D} or D_e . The *rate-distortion feasible set* for a source is the closure of the set of achievable rate-distortion pairs. The *rate-distortion function* $R(D)$ is defined by

$$R(D) \equiv \inf R | (R, D) \text{ is achievable.} \quad (4)$$

The *rate-distortion frontier* is the graph of the rate distortion function; the *distortion-rate function* is the inverse of the rate-distortion function.

If we assume that the coherent information continues to play the role, in quantum information theory, of the mutual information in classical information theory, then we are led to define quantum analogues of the information rate-distortion function.

The *entanglement information rate-distortion function* $R^I(D)$ for a source is defined by:

$$R^I(D) \equiv \min_{\mathcal{A} | d(\mathcal{A}) \leq D} I_c(\rho, \mathcal{A}). \quad (5)$$

One may conjecture that, as in the classical case, the information rate-distortion function just defined is equal to the information-disturbance function defined above, and thus that $R^I(D)$ tells us the lowest rate at which we can use channel qubits to end a quantum source with entanglement distortion no greater than D . We might worry that peculiarly quantum features such as the superadditivity of the coherent information or the failure of data pipelining require some modifications to the straightforward quantum analogue of the classical result, as they do in the case of noisy channel coding. In what follows, I will derive a lower bound on the required description rate; perhaps this bound is not tight due to the peculiarly quantum effects just discussed, although the fact that general encodings are used in deriving the bound makes me doubt that the failure of data pipelining is relevant. I will not discuss achievability. I expect the techniques required for noisy channel coding may help in showing achievability, although rate-distortion may be more difficult as we cannot rely on bounds that only become tight for fidelities near one; the saving grace may be that the “noise”-like element is only

truncation to a smaller space, and this is likely to be much easier to deal with than a general channel operation.

The proof I will give uses two lemmas. First, we need the convexity of the information rate-distortion function:

Lemma 1 $R^I(D)$ is a nonincreasing, convex function of D ; that is,

$$D_1 < D_2 \rightarrow R^I(D_1) \geq R^I(D_2), \text{ and} \quad (6)$$

$$R^I(\lambda D_1 + (1 - \lambda)D_2) \leq \lambda R^I(D_1) + (1 - \lambda)R^I(D_2) \quad (7)$$

where $0 \leq \lambda \leq 1$.

Proof: Nondecrease: As D increases, the domain of the minimization in the definition of $R^I(D)$ becomes larger (or at least no smaller); therefore, $R^I(D)$ does not increase.

Convexity: Let (R_1, D_1) and (R_2, D_2) be points on the information rate-distortion curve, and let \mathcal{E}_1 and \mathcal{E}_2 be operations achieving the minimum in the definition of $R^I(D)$ for $D = D_1$ and $D = D_2$ respectively. Consider the operation $\mathcal{E}_\lambda \equiv \lambda\mathcal{E}_1 + (1 - \lambda)\mathcal{E}_2$. Since the entanglement disturbance is linear in the operation, this operation has disturbance $D_\lambda \equiv D(\mathcal{E}_\lambda) = \lambda D(\mathcal{E}_1) + (1 - \lambda)D(\mathcal{E}_2)$. Since $R^I(D_\lambda)$ is the minimum of the coherent information over operations, $R^I(D_\lambda) \leq I_c(\rho, \mathcal{E}_\lambda)$. And since the coherent information is convex in the operation, this is less than $\lambda I_c(\rho, \mathcal{E}_1) + (1 - \lambda)I_c(\rho, \mathcal{E}_2) = \lambda R^I(D_1) + (1 - \lambda)R^I(D_2)$. ■

Notice that the only property of the disturbance that was used in this proof was the linearity of the disturbance in the operation; hence it applies to any quantum rate-distortion function defined using a disturbance measure with this property, in particular to the information rate-distortion function using average pure-state fidelity.

The second lemma we need is that the coherent information for a process on a composite state is greater than or equal to the total of the “marginal coherent informations” for the reductions of the process and the initial state to the subsystems.

Lemma 2

$$I_c(\rho^{(n)}, \mathcal{E}^{(n)}) \geq \sum_i I_c(\rho_i, \mathcal{E}_i^{(n)}).$$

Here the definition of the reduced operation $\mathcal{E}_i^{(n)}$ is the same as that of $\mathcal{T}_i^{(n)}$ in (2), except that $\mathcal{D}^{(n)}$ is omitted on the RHS. ρ_i is of course the marginal density operator of the i -th system. *Proof:* The lemma obviously follows from the two-system case:

$$I_c(\rho^{(2)}, \mathcal{E}^{(2)}) \geq I_c(\rho_1, \mathcal{E}_1^{(2)}) + I_c(\rho_2, \mathcal{E}_2^{(2)}). \quad (8)$$

If we model this in the usual way, by purifying Q_1 into R_1 and Q_2 into R_2 , adjoining an initially pure environment E and effecting the operation $\mathcal{E}^{(2)}$ by a unitary interaction $U^{Q_1 Q_2 E}$, this becomes:

$$S(\rho^{Q_1 Q_2}) - S(\rho^{R_1 Q_1 R_2 Q_2}) \geq S(\rho^{Q_1}) + S(\rho^{Q_2}) - S(\rho^{R_1 Q_1}) - S(\rho^{R_2 Q_2}), \quad (9)$$

which may be rewritten

$$S(\rho^{R_1 Q_1}) + S(\rho^{R_2 Q_2}) - S(\rho^{R_1 Q_1 R_2 Q_2}) \geq S(\rho^{Q_1}) + S(\rho^{Q_2}) - S(\rho^{Q_1 Q_2}). \quad (10)$$

The quantity appearing in this last form is the sum of the marginal entropies of two subsystems, minus the joint entropy of the composite system; it is a quantity which can be larger in quantum theory than it can in classical theory, due to entanglement. In this form, the inequality says that this excess of marginal over joint entropies is reduced if we ignore (trace over) parts of each of the subsystems. This follows from strong subadditivity, as we may show by rewriting it yet again as:

$$S(\rho^{R_1 Q_1 R_2 Q_2}) + S(\rho^{Q_1}) + S(\rho^{Q_2}) \leq S(\rho^{Q_1 Q_2}) + S(\rho^{R_1 Q_1}) + S(\rho^{R_2 Q_2}). \quad (11)$$

In this form, it follows from two applications of strong subadditivity (thanks to Michael Nielsen for this observation). We start with a case of strong subadditivity for the three systems R_1 , Q_1 , and $R_2 Q_2$:

$$S(\rho^{R_1 Q_1 R_2 Q_2}) + S(\rho^{Q_1}) \leq S(\rho^{R_1 Q_1}) + S(\rho^{Q_1 R_2 Q_2}). \quad (12)$$

Adding $S(\rho^{Q_2})$ to both sides gives:

$$S(\rho^{R_1 Q_1 R_2 Q_2}) + S(\rho^{Q_1}) + S(\rho^{Q_2}) \leq S(\rho^{R_1 Q_1}) + S(\rho^{Q_1 R_2 Q_2}) + S(\rho^{Q_2}). \quad (13)$$

The last two terms on the right hand side are then upper bounded by another application of strong subadditivity in the form $S(\rho^{Q_1 R_2 Q_2}) + S(\rho^{Q_2}) \leq S(\rho^{Q_1 Q_2}) + S(\rho^{R_2 Q_2})$, giving (11). \blacksquare

Theorem 1 *Let $(\mathcal{E}^{(n)}, \mathcal{D}^{(n)})$ be a $(2^{nR}, n)$ rate-distortion code with distortion D . Then $R \geq R^I(D)$.*

Proof: I give the proof as a chain of inequalities and equivalences, followed by notes justifying each inequality when possible.

$$nR \geq S(\rho^{(n)'}) \tag{14}$$

$$\geq S(\rho^{(n)'}) - S_e(\rho, \mathcal{E}^{(n)}) \equiv I_c(\rho^{(n)}, \mathcal{E}^{(n)}) \tag{15}$$

$$\geq I_c(\rho^{(n)}, \mathcal{D}^{(n)} \circ \mathcal{E}^{(n)}) \tag{16}$$

$$\geq \sum_i I_c(\rho_i, \mathcal{E}_i^{(n)}) \tag{17}$$

$$\geq \sum_i R^I(d(\rho, \mathcal{E}_i^{(n)})) \equiv n \sum_i \frac{1}{n} R^I(d(\rho, \mathcal{E}_i^{(n)})) \tag{18}$$

$$\geq nR^I\left(\sum_i \frac{1}{n} d(\rho, \mathcal{E}_i^{(n)})\right) \equiv nR^I(D). \tag{19}$$

(14) holds because nR is the log of the dimension of an n -block of channel Hilbert space, which constitutes an upper bound to the von Neumann entropy of a density operator on that space. (15) follows from the positivity of entropy exchange, (16) from the data processing inequality, (17) from Lemma 2, the superadditivity of coherent information compared to marginal coherent information, (18) follows from the definition of the entanglement information rate-distortion function, and (19) from Lemma (1), the convexity of the rate-distortion function. ■