Delay-Constrained Source Coding for a Peak Distortion Measure

Cheng Chang and Anant Sahai

***Abstract*— We consider the problem of lossy source coding under a peak distortion measure in which source symbols are revealed to the encoder in real time and need to be reconstructed by the decoder within a fixed end-to-end delay. Following the lossless case in [3], we investigate the tradeoff between end-to-end delay and the probability of distortion violation. As in the lossless case, the delay-constrained error (distortion-violation) exponent is generally much higher than the fixed-block coding case.**

I. Introduction

The core issue we are interested in is the impact of "causality" on lossy source coding. In [9], the rate distortion performance for a strictly causal decoder is studied, and it is shown that the optimal performance can be obtained by time-sharing between memoryless codes. Thus, it is in general strictly worse than the performance of classical fixed-block source coding that allows arbitrarily large delay. The large deviation performance of the zero delay decoder problem is studied in [8].

Allowing some finite end-to-end delay, [11] shows that the average block coding rate distortion performance can still be approached exponentially with delay.

In this paper, we consider a coding system for a streaming source, drawn iid from a distribution p_x on finite alphabet $\mathcal{X}$. The encoder, mapping source symbols into bits at fixed rate R, is strictly causal and the decoder has to reconstruct the source symbols (under a peak distortion constraint) within a fixed end-to-end latency constraint. The system is illustrated in Figure 1.

Generalizing our previous work in [3], [2] on end-to-end delay performance of lossless source coding, we have derived the fixed-delay distortion-violation exponent for lossy source coding under a peak distortion constraint. A "focusing" type bound is derived that is quite similar to its lossless source coding counterpart. As shown in the appendix, the technical reason for the similarity is that the length of optimal variable-length codes, or equivalently the rate distortion functions, are concave in the empirical distribution for both lossless source coding and lossy source coding under peak distortion constraint. This is not the case for rate distortion functions under average distortion measures [4].

C. Chang and A. Sahai are with the Department of Electrical Engineering and Computer Science, University of California at Berkeley. Email: cchang@eecs.berkeley.edu, sahai@eecs.berkeley.edu

A. Source coding for streaming data with an end-to-end delay constraint

In [3], [2], we studied the special case of lossless coding. We showed that the error exponent with fixed end-to-end delay is *much higher* than its fixed block-length counterpart. The delay exponent also turned out to be related to the buffer overflow exponent studied by Jelinek in [6].

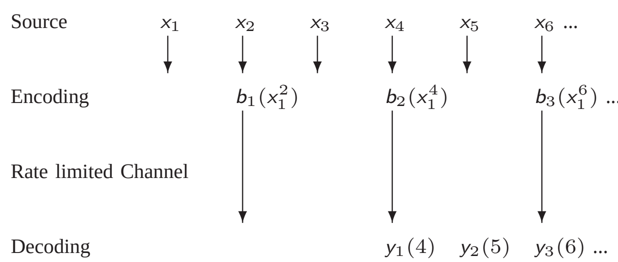


Fig. 1. Time line for delay constrained source coding: rate $R = \frac{1}{2}$, delay $\Delta = 3$. $y_i(i + \Delta)$ is the reconstruction of x_i at time $i + \Delta$

In this paper, we relax the lossless coding constraint to allow some distortion on a per-symbol basis. This is different[1] from the time-averaged distortion studied in [8], [9], [11].

B. Rate distortion under a peak distortion constraint

[4] introduced peak distortion measures:

$$d(x_1^N, y_1^N) \triangleq \max_{1 \le i \le n} d(x_i, y_i) \quad (1)$$

and the corresponding rate distortion theorem:

Proposition 1: The rate-distortion function $R(D)$ for peak distortion:

$$R(p_x, D) \triangleq \min_{W \in \mathcal{W}_D} I(p_x, W) \quad (2)$$

where $\mathcal{W}_\mathcal{D}$ is the set of all transition matrices that satisfy the peak distortion constraint, i.e. $\mathcal{W}_D = \{W : W(y|x) = 0, \text{if } d(x,y) > D\}$. To have $\Pr(d(x_1^N, y_1^N) >$

[1]The difference is for our case, we can not relax the distortion from one symbol to another.

$D) = 0$, we can implement a universal variable length prefix-free code with code length $l_D(x_1^N)$ where

$$l_D(x_1^N) = n(R(p_{x_1^N}, D) + \delta_N) \quad (3)$$

where $p_{x_1^N}$ is the empirical distribution of x_1^N, and δ_N goes to 0 as N goes to infinity.

This is a simple corollary of the type covering lemma [4]. The problem is only interesting if the target peak distortion D is higher than $\underline{D}$ and lower than $\overline{D}$, where

$$\underline{D} \triangleq \max_{x\in\mathcal{X}} \min_{y\in\mathcal{Y}} d(x,y) \text{ and } \overline{D} \triangleq \min_{y\in\mathcal{Y}} \max_{x\in\mathcal{X}} d(x,y)$$

Note that both $\underline{D}$ and $\overline{D}$ only depend on the distortion measure $d(\cdot,\cdot)$, not the source distribution p_{x}.

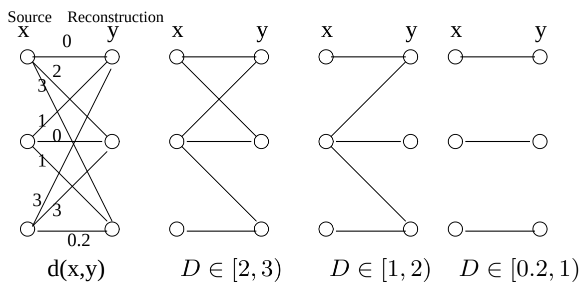


Fig. 2. $d(x,y)$ and valid reconstructions under different peak distortion constraints D. (x,y) is linked if $d(x,y) \le D$. $\underline{D} = 0.2$ and $\overline{D} = 3$. For $D \in [0.2, 1)$, this is a lossless source coding problem.

The rate distortion function is in general non-concave in the source distribution p as pointed out in [7]. But for peak distortion, $R(p,D)$ *is* concave in p for a fixed distortion constraint D. The proof is in the appendix.

Lemma 1: $R(p,D)$ is concave in p for fixed D.

As a simple corollary of the block-coding error exponents for average distortion from [7], we have the following result.

Proposition 2: Block coding error exponent under peak distortion:

$$\liminf_{n\to\infty} -\frac{1}{N}\log_2 \Pr(d(\mathsf{x}_1^N, \mathsf{y}_1^N) > D) = E_D^b(R)$$

$$\text{where } E_D^b(R) \triangleq \min_{q_{\mathsf{x}}: R(q_{\mathsf{x}}, D) > R} D(q_{\mathsf{x}} \| p_{\mathsf{x}}) \quad (4)$$

where y_1^N is the reconstruction of x_1^N using an optimal rate R code.

For lossless source coding, if $R > \log_2|\mathcal{X}|$, the error probability is 0 and the error exponent is infinite. Similarly, for lossy source coding under peak distortion, the error exponent is infinite whenever

$$R > \overline{R}_D \triangleq \sup_{q_{\mathsf{x}}} R(q_{\mathsf{x}}, D)$$

where $\overline{R}_D$ only depends on $d(\cdot,\cdot)$ and D.

II. Main Result

We investigate the relation between delay Δ and the probability of distortion violation $\Pr(d(\mathsf{x}_i, \mathsf{y}_i(i+\Delta)) > D)$, where $\mathsf{y}_i(i+\Delta)$ is the reconstruction of x_i at time $i+\Delta$ and D is the peak distortion constraint.

Definition 1: A rate R sequential source code shown in Figure 1 achieves error (distortion violation) exponent $E_D(R)$ with delay if for all $\epsilon > 0$, there exists $K < \infty$, s.t. $\forall i, \Delta > 0$

$$\Pr(d(\mathsf{x}_i, \mathsf{y}_i(i+\Delta)) > D) \le K 2^{-\Delta(E_D(R)-\epsilon)}$$

Theorem 1: Consider fixed rate source coding of iid streaming data $\mathsf{x}_i \sim p_{\mathsf{x}}$, with a non-negative peak distortion measure d. For $D \in (\underline{D}, \overline{D})$, and rates $R \in (R(p_{\mathsf{x}}, D), \overline{R}_D)$, the following error exponent with delay is optimal and achievable.

$$E_D(R) \triangleq \inf_{\alpha>0} \frac{1}{\alpha} E_D^b((\alpha+1)R) \quad (5)$$

where $E_D^b(R)$ is the block coding error exponent under peak distortion constraint, as defined in Proposition 2.

A. Numerical Examples

Consider $p_{\mathsf{x}} = \{0.1, 0.7, 0.2\}$ and a distortion measure on $\mathcal{X} \times \mathcal{Y}$ as shown in Figure 2. We plot the rate distortion $R - D$ curve and the error exponents in Figure 3. We have a higher delay error exponent than block coding, just as in the lossless case of [3].

III. Proofs

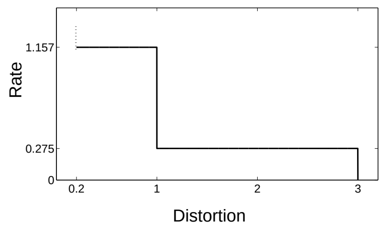
In this section, we show that the error exponent in Theorem 1 is both achievable asymptotically with delay and that no better exponents are possible.

A. Converse

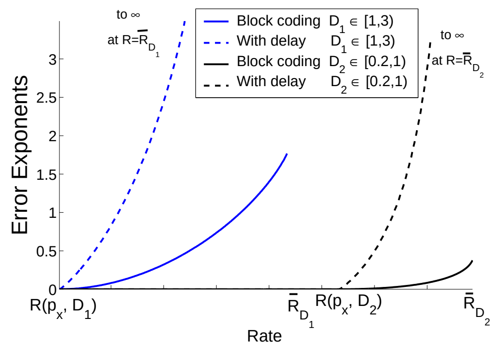
The proof of the converse is similar to the upper bound argument in [3] for lossless source coding with delay constraints. To bound the best possible delay exponent, a genie-aided encoder/decoder pair is used to translate the block-coding bounds in Proposition 2 to the fixed delay context. The arguments are analogous to the "focusing bound" derivation in [10] for the case of channel coding with feedback.

Proof: For simplicity of exposition, we ignore integer effects arising from the finite nature of Δ, R, etc. For every $\alpha > 0$ and delay Δ, consider a fixed-rate code running until time $\frac{\Delta}{\alpha} + \Delta$. By this time, the decoder will have committed to estimates for the source symbols up to time $i = \frac{\Delta}{\alpha}$. The total number of bits used during this period is $(\frac{\Delta}{\alpha} + \Delta)R$.

Now consider a genie that gives the encoder access to the first i source symbols at the beginning of time,



(a) $R - D$ curve under peak distortion constraint is a staircase function



(b) Error exponents of delay constrained and block source coding under different peak distortion constrains

Fig. 3. For $D_1 \in [1, 3)$, $R(p_x, D_1) = 0.275$ and $\overline{R}_{D_1} = 0.997$. For $D_2 \in [0.2, 1)$, the problem degenerates to lossless coding, so $R(p_x, D_1) = H(p_x) = 1.157$ and $\overline{R}_{D_1} = \log_2(3) = 1.585$

rather than forcing the encoder to get the source symbols gradually. Simultaneously, loosen the requirements on the decoder by only demanding correct estimates for the first i source symbols by the time $\frac{\Delta}{\alpha} + \Delta$. In effect, the deadline for decoding the *past* source symbols is extended to the deadline of the i-th symbol itself.

Any lower-bound to the distortion-violation probability of the new problem is clearly also a bound for the original problem. Furthermore, the new problem is just a fixed-length block-coding problem requiring the encoding of i source symbols into $(\frac{\Delta}{\alpha} + \Delta)R$ bits. The rate per symbol is

$$\left((\frac{\Delta}{\alpha} + \Delta)R\right)\frac{1}{i} = \left((\frac{\Delta}{\alpha} + \Delta)R\right)\frac{\alpha}{\Delta} = (\alpha + 1)R$$

Proposition 2 implies that the probability of distortion violation is at least exponential in $iE^b_D((\alpha + 1)R)$. Since $i = \frac{\Delta}{\alpha}$, this translates into a distortion-violation exponent of at most $\frac{E^b_D((\alpha+1)R)}{\alpha}$ with parameter Δ.

Since this is true for all $\alpha > 0$, we have a bound on the distortion violation exponent with fixed delay Δ:

$$\inf_{\alpha>0} \frac{1}{\alpha} E^b_D((\alpha + 1)R) \quad (6)$$

The α that minimizes (6) tells how much of the past $(\frac{\Delta}{\alpha})$ is involved in the dominant error event. ■

B. Achievability

We prove achievability by giving a universal coding scheme illustrated in Figure 4.

A block-length N is chosen that is much smaller than the target end-to-end delays[2], while still being large enough. For a discrete memoryless source $\mathcal{X}$, distortion measure $d(\cdot, \cdot)$, peak distortion constraint D, and large block-length N, we use the universal variable length prefix-free code in Proposition 1 to encode the i'th block $\vec{x}_i = x^{iN}_{(i-1)N+1} \in \mathcal{X}^N$. The code length $l_D(\vec{x}_i)$ is shown in (3),

$$NR(p_{\vec{x}_i}, D) \leq l_D(\vec{x}_i) \leq N(R(p_{\vec{x}_i}, D) + \delta_N) \quad (7)$$

The overhead δ_N is negligible for large N, since δ_N goes to 0 as N goes to infinity. The binary sequence describing the source is fed into a FIFO buffer described in Figure 4. The buffer is drained at a fixed rate R to obtain the encoded bits.[3] The decoder uses the bits it has received so far to get the reconstructions. If the relevant bits have not arrived by the time the reconstruction is due, it just guesses and we presume that a distortion-violation will occur.

As the following proposition indicates, the coding scheme is delay universal, i.e. the distortion-violation probability goes to 0 with exponent $E_D(R)$ for all source symbols and for all delays Δ big enough.

Proposition 3: For the iid source $\sim p_x$, peak distortion constraint D, and large N, using the universal real-time code described above, for all $\epsilon > 0$, there exists $K < \infty$, s.t. for all t, Δ:

$$\Pr(d(\vec{x}_t, \vec{y}_t((t + \Delta)N)) > D) \leq K2^{-\Delta N(E_D(R)-\epsilon)}$$

where $\vec{y}_t((t + \Delta)N)$ is the estimate of $\vec{x}_t$ at time $(t + \Delta)N$.

Before proving Proposition 3, we state the following lemma (proved in the appendix) bounding the probability of atypical source behavior.

Lemma 2: (Source atypicality) For all $\epsilon > 0$, block length N large enough, there exists $K < \infty$, s.t. for all n, for all $r < \overline{R}_D$:

$$\Pr\left(\sum_{i=1}^{n} l_D(\vec{x}_i) > nNr\right) \leq K2^{-nN(E^b_D(r)-\epsilon)} \quad (8)$$

[2]We are interested in the performance with asymptotically large delays Δ.

[3]Notice that if the buffer is empty, the output of the encoder buffer can be gibberish binary bits. The decoder simply discards these meaningless bits because it is aware that the buffer is empty.

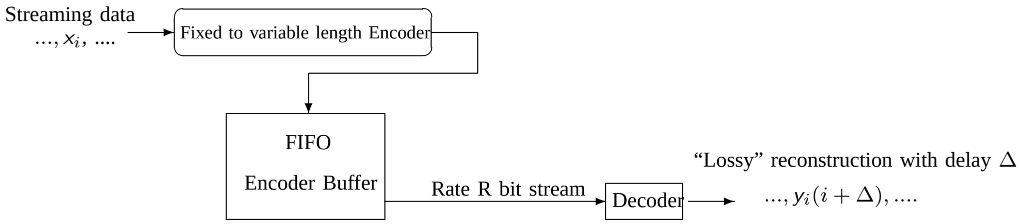


Fig. 4. A delay optimal lossy source coding system.

Now we are ready to prove Proposition 3.

Proof: At time $(t+\Delta)N$, the decoder cannot decode $\vec{x}_t$ within peak distortion D only if the binary strings describing $\vec{x}_t$ are *not* all out of the buffer. Since the encoding buffer is FIFO, this means that the number of outgoing bits from some time t_1, where $t_1 \leq tN$ to $(t+\Delta)N$ is less than the number of the bits in the buffer at time t_1 plus the number of incoming bits from time t_1 to time tN. Suppose the buffer is last empty at time $tN - nN$ where $0 \leq n \leq t$, given this condition, the peak distortion is not satisfied only if $\sum_{i=0}^{n-1} l_D(\vec{x}_{t-i}) > (n+\Delta)NR$. Write $l_{D,\max}$ as the longest possible code length. $l_{D,max} \leq |\mathcal{X}|\log_2(N+1) + N\log_2|\mathcal{X}|$. Then $\Pr(\sum_{i=0}^{n-1} l_D(\vec{x}_{t-i}) > (n+\Delta)NR) > 0$ only if $n > \frac{(n+\Delta)NR}{l_{D,max}} > \frac{\Delta NR}{l_{D,max}} \triangleq \beta\Delta$. So

$$
\begin{aligned}
&\Pr\left(d(\vec{x}_t, \vec{y}_t((t+\Delta)N)) > D\right) \\
\leq \quad & \sum_{n=\beta\Delta}^{t} \Pr(\sum_{i=0}^{n-1} l(\vec{x}_{t-i}) > (n+\Delta)NR) \qquad (9) \\
\leq_{(a)} \quad & \sum_{n=\beta\Delta}^{t} K_1 2^{-nN(E_D^b(\frac{(n+\Delta)NR}{nN})-\epsilon_1)} \\
\leq_{(b)} \quad & \sum_{n=\gamma\Delta}^{\infty} K_2 2^{-nN(E_D^b(R)-\epsilon_2)} \\
& + \sum_{n=\beta\Delta}^{\gamma\Delta} K_2 2^{-\Delta N(\min_{\alpha>1}\{\frac{E_D^b(\alpha R)}{\alpha-1}\}-\epsilon_2)} \\
\leq_{(c)} \quad & K_3 2^{-\gamma\Delta N(E_D^b(R)-\epsilon_2)} \\
& + |\gamma\Delta - \beta\Delta| K_3 2^{-\Delta N(E_D(R)-\epsilon_2)} \\
\leq_{(d)} \quad & K 2^{-\Delta N(E_D(R)-\epsilon)}
\end{aligned}
$$

where $K_i's$ and $\epsilon_i's$ are properly chosen real numbers. (a) is true because of Lemma 2. Define $\gamma \triangleq \frac{E_D(R)}{E_D^b(R)}$. In the first part of (b), we only need the fact that $E_D^b(R)$ is non decreasing with R. In the second part of (b), we write $\alpha = \frac{n+\Delta}{n}$ and take the α to minimize the error exponents. The first term of (c) comes from the sum of a convergent geometric series and the second is by the definition of $E_D(R)$. (d) is by the definition of γ. ■

Combining (6) and Proposition 3, we establish the desired results summarized in Theorem 1.

IV. Future Work

Both the converse (focusing bound) and the achievability analysis can be adapted for average distortion measures. However there is a gap between the two bounds on the error exponent due to the non-concavity of the rate-distortion function in the empirical distribution. Hence, the optimal end-to-end delay constrained error exponent for lossy source coding under average distortion constraints remain unknown.

Appendix

A. Proof of Lemma 1

Proof: To show that $R(p, D)$ is concave in p, it is enough to show that for any two distributions p_0 and p_1 and for any $\lambda \in [0, 1]$,

$$R(p_\lambda, D) \geq \lambda R(p_0, D) + (1-\lambda)R(p_1, D)$$

where $p_\lambda = \lambda p_0 + (1-\lambda)p_1$. Define:

$$W^* = \arg\min_{W \in \mathcal{W}_D} I(p_\lambda, W)$$

Then, from the definition of $R(p, D)$ we know that $R(p_\lambda, D)$

$$
\begin{aligned}
&= I(p_\lambda, W^*) \\
&\geq \lambda I(p_0, W^*) + (1-\lambda) I(p_1, W^*) \qquad (10) \\
&\geq \lambda \min_{W \in \mathcal{W}_D} I(p_0, W) + (1-\lambda) \min_{W \in \mathcal{W}_D} I(p_1, W) \\
&= \lambda R(p_0, D) + (1-\lambda) R(p_1, D)
\end{aligned}
$$

(10) is true because $I(p, W)$ is concave in p for fixed W and $p_\lambda = \lambda p_0 + (1-\lambda)p_1$. The rest is from the definitions. □

B. Proof of Lemma 2

Proof: We only need to show the case for $r > R(p_x, D)$. By Cramér's theorem [5], for all $\epsilon_1 > 0$, there exists K_1, such that :

$$
\begin{aligned}
\Pr(\sum_{i=1}^{n} l_D(\vec{x}_i) > nNr) &= \Pr(\frac{1}{n}\sum_{i=1}^{n} l_D(\vec{x}_i) > Nr) \\
&\leq K_1 2^{-n(\inf_{z>Nr} I(z)-\epsilon_1)}
\end{aligned}
$$

where the rate function $I(z)$ is [5]:

$$I(z) = \sup_{\rho \geq 0}\{\rho z - \log_2(\sum_{(\vec{x} \in \mathcal{X}^N} p_x(\vec{x}) 2^{\rho l_D(\vec{x})})\} \quad (11)$$

It is clear that $I(z)$ is monotonically increasing with z and $I(z)$ is continuous. Thus

$$\inf_{z > Nr} I(z) = I(Nr) \quad (12)$$

Using the upper bound on $l_D(\vec{x})$ in (7):

$$\begin{aligned} & \log_2(\sum_{\vec{x} \in \mathcal{X}^N} p_x(\vec{x}) 2^{\rho l_D(\vec{x})}) \\ \leq\ & \log_2(\sum_{q_x \in \mathcal{T}^N} 2^{-ND(q_x \| p_x)} 2^{\rho(\delta_N + NR(q_x, D))}) \\ \leq\ & \log_2(2^{N\epsilon_N} 2^{-N \min_{q_x}\{D(q_x\|p_x) - \rho R(q_x, D) - \rho\delta_N\}}) \\ =\ & N\big(-\min_{q_x}\{D(q_x\|p_x) - \rho R(q_x, D) - \rho\delta_N\} + \epsilon_N\big) \end{aligned}$$

where $\mathcal{T}^N$ is the set of all types of $\mathcal{X}^N$,and $2^{N\epsilon_N}$ is the number of types in $\mathcal{X}^N$, $0 < \epsilon_N \leq \frac{|\mathcal{X}| \log_2(N+1)}{N}$, and so ϵ_N goes to 0 as N goes to infinity.

Substitute the above inequalities into $I(Nr)$ in (11):

$$\begin{aligned} I(Nr) \quad \geq \quad & N\big(\sup_{\rho \geq 0}\{\min_{q_x} \rho(r - R(q_x, D) - \delta_N) \\ & + D(q_x\|p_x)\} - \epsilon_N\big) \end{aligned} \quad (13)$$

We next show that $I(Nr) \geq N(E_D^b(r) + \epsilon_N)$ where ϵ goes to 0 as N goes to infinity. We show the existence of a saddle point of the min-max for a function

$$f(q_x, \rho) = \rho(r - R(q_x, D) - \delta_N) + D(q_x\|p_x)$$

Obviously, for fixed q_x, $f(q_x, \rho)$ is a linear function of ρ, thus concave. Also for fixed $\rho \geq 0$, $f(q_x, \rho)$ is a convex function of q_x, because both $-R(q_x, D)$ and $D(q_x\|p_x)$ are convex in q_x. Write

$$g(u) = \min_{q_x} \sup_{\rho \geq 0}(f(q_x, \rho) + \rho u)$$

Showing that $g(u)$ is finite around $u = 0$ establishes the existence of the saddle point as shown in Exercise 5.25 [1].

$$\begin{aligned} & \min_{q_x} \sup_{\rho \geq 0} f(q, \rho) + \rho u \\ =_{(a)}\ & \min_{q_x} \sup_{\rho \geq 0} \rho(r - R(q_x, D) - \delta_N + u) + D(q_x\|p_x) \\ \leq_{(b)}\ & \min_{q_x: R(q_x, D) \geq r - \delta_N + u} \ \sup_{\rho \geq 0} \\ & \quad \rho(r - R(q_x, D) - \delta_N + u) + D(q_x\|p_x) \\ \leq_{(c)}\ & \min_{q_x: R(q_x, D) \geq r - \delta_N + u} D(q_x\|p_x) \quad <_{(d)} \infty \end{aligned} \quad (14)$$

(a) is a definition. (b) is true because $R(p_x, D) < r < \overline{R}_D$, thus for very small δ_N and u, $R(p_x, D) < r - \delta_N + u < \overline{R}_D$. Thus there exists a distribution q_x, s.t. $R(q_x, D) \geq r - \delta_N + u$. (c) is because $R(q_x, D) \geq r - \delta_N + u$ and $\rho \geq 0$. (d) is true because we might as well assume that $p_x(x) > 0$ for all $x \in \mathcal{X}$, and $r - \delta_N + u < \overline{R}_D$. Thus we proved the existence of the saddle point of $f(q, \rho)$.

$$\sup_{\rho \geq 0}\{\min_q f(q, \rho)\} = \min_q\{\sup_{\rho \geq 0} f(q, \rho)\} \quad (15)$$

Note that if $R(q_x, D) < r - \delta_N$, ρ can be chosen to be arbitrarily large to make $\rho(r - R(q_x, D) - \delta_N) + D(q_x\|p_x)$ arbitrarily large. Thus the q_x to minimize $\sup_\rho \rho(r - R(q_x, D) - \delta_N) + D(q_x\|p_x)$ satisfies $r - R(q_x, D) - \delta_N \geq 0$. So

$$\begin{aligned} & \min_{q_x}\{\sup_{\rho \geq 0} \rho(r - R(q_x, D) - \delta_N) + D(q_{xy}\|p_{xy})\} \\ =_{(a)}\ & \min_{q_x: R(q_x, D) \geq r - \delta_N} \ \sup_{\rho \geq 0} \\ & \quad \{\rho(r - R(q_x, D) - \delta_N) + D(q_x\|p_x)\} \\ =_{(b)}\ & \min_{q_x: R(q_x, D) \geq r - \delta_N} \{D(q_x\|p_x)\} \\ =_{(c)}\ & E_D^b(r - \delta_N) \end{aligned} \quad (16)$$

(a) follows from the argument above. (b) is because $r - R(q_x, D) - \delta_N \leq 0$ and $\rho \geq 0$, and thus $\rho = 0$ maximizes $\rho(r - R(q_x, D) - \delta_N)$. (c) is by definition in (4). Combining (13) (15) and (16), letting N be sufficiently big so that δ_N is sufficiently small, and noticing that $E_D^b(r)$ is continuous on r, we get the desired bound in (8). □

References

[1] Stephen Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.
[2] Cheng Chang and Anant Sahai. The error exponent with delay for lossless source coding. *Information Theory Workshop, Punta del Este, Uruguay*, March 2006.
[3] Cheng Chang and Anant Sahai. Price of ignorance. *IEEE Trans. Inform. Theory*, submitted.
[4] Imre Csiszár and János Körner. *Information Theory*. Akadémiai Kiadó, Budapest, 1986.
[5] Amir Dembo and Ofer Zeitouni. *Large Deviations Techniques and Applications*. Springer, 1998.
[6] Frederick Jelinek. Buffer overflow in variable length coding of fixed rate sources. *IEEE Transactions on Information Theory*, 14:490–501, 1968.
[7] Katalin Marton. Error exponent for source coding with a fidelity criterion. *IEEE Transactions on Information Theory*, 20(2):197–199, 1974.
[8] Neri Merhav and Ioannis Kontoyiannis. Source coding exponents for zero-delay coding with finite memory. *IEEE Transactions on Information Theory*, 49(3):609–625, 2003.
[9] David L. Neuhoff and R. Kent Gilbert. Causal source codes. *IEEE Transactions on Information Theory*, 28(5):701–713, 1982.
[10] Anant Sahai. Why block length and delay are not the same thing. *IEEE Trans. Inform. Theory*, submitted, arXiv: cs.IT/0610139.
[11] D. Tse, R.G. Gallager, and J.N. Tsitsiklis. Optimal buffer control for variable-rate lossy compression. *31st Allerton Conference*, 1993.