Mathematical Problems of Computer Science 18, 1997, 40–45.

The Binary Hamming Rate-reliability-distortion Function*

E. A. Haroutunian and A. N. Haroutunian

Institute for Informatics and Automation Problems
of NAS of RA and YSU

Abstract

For a binary source and Hamming distortion measure the rate-reliability-distortion function is derived and analyzed.

1 Introduction

Shannon defined [1,2] rate-distortion function (RDF) as a minimal rate that can be achieved for source data transmission with distortion smaller than a required level. The properties of RDF can be found in the book of I. Csiszár and J. Körner [3] and in the work of Ahlswede [4].

Haroutunian and Mekoush introduced [5] the rate-reliability-distortion function (RRDF).

The discrete stationary memoryless source $\{X_k\}_{k=1}^{\infty}$ is a sequence of independent copies of a random variable X with finite alphabet $\mathcal{X}$ and probability distribution $P^* = \{P^*(x), x \in \mathcal{X}\}$. The probability of source message $\mathbf{x} = (x_1, ..., x_n) \in \mathcal{X}^n$ will be $P^*(\mathbf{x}) = \prod_{i=1}^{n} P^*(x_i)$. The reproduction alphabet is the finite set $\mathcal{U}$. Let $d : X \times U \to [0, \infty)$ be the single-letter distortion measure. The distortion between source message $\mathbf{x}$ and reconstructed vector $\mathbf{u} = (u_1, ..., u_n) \in \mathcal{U}^n$ is

$$d(x, u) = n^{-1} \sum_{i=1}^{n} d(x_i, u_i).$$

Block encoding and decoding of length n are respectively

$$f : \mathcal{X}^n \to \{1, 2, ..., L(n)\}, \ F : \{1, 2, ..., L(n)\} \to \mathcal{U}^n.$$

Let $\mathcal{A} = \{\mathbf{x} : F(f(\mathbf{x})) = \mathbf{u}, d(\mathbf{x}, \mathbf{u}) \leq \Delta\}$. The probability of error is

$$e(f, F, \Delta, n) = 1 - P^{*n}(\mathcal{A}).$$

In the paper all logarithms and exponents are to the base 2.

Definition: A number $R \geq 0$ is a (E, Δ)–achievable rate ($E > 0$, $\Delta \geq 0$) if for any $\epsilon > 0$ and sufficiently large n there exists a code (f, F) such that

$$n^{-1} \log L(n) \leq R + \epsilon$$

*This work was supported by INTAS under Grant 94-469.

and

$$e(f, F, \Delta, n) \leq \exp\{-nE\}.$$

For fixed $E > 0$ and $\Delta \geq 0$ denote

$$R(E, \Delta) = \min\{R: \ R \text{ is } (E, \Delta) - \text{achievable}\}.$$

Note that the rate-reliability-distortion function $R(E, \Delta)$ is a generalization of rate-distortion function $R(\Delta)$, which is limit point of $R(E, \Delta)$:

$$\lim_{E \to 0} R(E, \Delta) = R(\Delta).$$

RRDF is the minimal rate at which the messages of a source can be encoded when the probability of exceeding given distortion level Δ is less than or equal to $\exp\{-nE\}$, where n is the blocklength and E is called reliability.

Let $P = \{P(x), \, x \in \mathcal{X}\}$ be some probability distribution (PD) and

$$Q = \{Q(u \mid x), x \in \mathcal{X}, u \in \mathcal{U}\}$$

be a conditional PD. Denote

$$\alpha(E) = \{P : D(P \parallel P^*) \leq E\},$$

where $D(P \parallel P^*)$ is divergence

$$D(P \parallel P^*) = \sum_{x \in \mathcal{X}} P(x) \log \frac{P(x)}{P^*(x)}.$$

The mathematical expectation of distortion with respect to distributions P and Q is

$$E_{P,Q} d(X, U) = \sum_{x,u} P(x) Q(u \mid x) d(x, u).$$

Shannon's classic result states that $R(\Delta)$ is of a form

$$R(\Delta) = \min_{Q: E_{P,Q} d(X,U) \leq \Delta} I_{P,Q}(X \wedge U), \tag{1}$$

where mutual information is

$$I_{P,Q}(X \wedge U) = \sum_{x,u} P(x) Q(u \mid x) \log \frac{Q(u \mid x)}{\sum_x P(x) Q(u \mid x)}.$$

The following general result was stated in [5].

Theorem 1: For any $E > 0$, $\Delta \geq 0$

$$R(E, \Delta) = \max_{P \in \alpha(E)} \min_{Q: E_{P,Q} d(X,U) \leq \Delta} I_{P,Q}(X \wedge U). \tag{2}$$

2 Two Properties

It is well known that RDF is a non-increasing and convex function of Δ. Now let us establish the same properties for RRDF.

Property 1: $R(E,\Delta)$ is a non-increasing and convex function of Δ.

Proof: The first part is evident. Let for fixed E the points (Δ_1, R_1) and (Δ_2, R_2) belong to the curve $R(E,\Delta)$ and $\Delta_1 \leq \Delta_2$. We shall prove that for every λ, $0 < \lambda \leq 1$,

$$R(E,\Delta_\lambda) = R(E, \lambda\Delta_1 + (1-\lambda)\Delta_2) \leq \lambda R(E_1,\Delta_1) + (1-\lambda)R(E_2,\Delta_2).$$

Denote for fixed distribution P

$$R(P,\Delta) = \min_{Q: E_{P,Q} d(X,U) \leq \Delta} I_{P,Q}(X \wedge U).$$

The function $R(P,\Delta)$, as a RDF for P, is convex of Δ. Then

$$R(E,\Delta_\lambda) = \max_{P \in \alpha(E)} R(P, \lambda\Delta_1 + (1-\lambda)\Delta_2) \leq$$

$$\max_{P \in \alpha(E)} (\lambda R(P,\Delta_1) + (1-\lambda)R(P,\Delta_2)) \overset{(a)}{\leq} \lambda \max_{P \in \alpha(E)} R(P,\Delta_1) + (1-\lambda) \max_{P \in \alpha(E)} R(P,\Delta_2) =$$

$$= \lambda R(E_1,\Delta_1) + (1-\lambda)R(E_2,\Delta_2),$$

where (a) follows from the facts that $R(P,\Delta_1) \geq R(P,\Delta_2)$, because $\Delta_1 \leq \Delta_2$, and the point $R_\lambda = \lambda R(P,\Delta_1) + (1-\lambda)R(P,\Delta_2)$ divides the segment $[R(P,\Delta_1); R(P,\Delta_2)]$ by proportion $\lambda/1-\lambda$. Concequently, R_λ is not larger than the point that divides the segment $[\max_{P \in \alpha(E)} R(P,\Delta_1); \max_{P \in \alpha(E)} R(P,\Delta_2)]$ by the same proportion.

It is natural to examine the properties of RRDF with respect to reliability E. It is expected that higher reliability will result larger transmission rate, which asserts the following property.

Property 2: $R(E,\Delta)$ is a non-decreasing function of E.

3 Function $R_{\mathrm{BH}}(E,\Delta)$

In this article we specify $R(E,\Delta)$ for an important class of sources. For binary source $\mathcal{X} = \{0,1\}$ with generic PD $P^* = \{p^*, 1-p^*\}$ and Hamming distance

$$d(x,u) = \begin{cases} 0, & \text{if } x = u, \\ 1 & \text{if } x \neq u, \end{cases}$$

we denote RDF and RRDF by $R_{\mathrm{BH}}(\Delta)$ and $R_{\mathrm{BH}}(E,\Delta)$, respectively.

Let $H_{P^*}(X)$ and $H_\Delta(X)$, $\Delta \geq 0$, are binary entropy functions

$$H_{P^*}(X) = -p^* \log p^* - (1-p^*)\log(1-p^*),$$

$$H_\Delta(X) = -\Delta \log \Delta - (1-\Delta)\log(1-\Delta).$$

Similar notations of entropies are in force for other PD. It is known (see [6]) that

$$R_{BH}(\Delta) = \begin{cases} H_{P^*}(X) - H_\Delta(X), & 0 \le \Delta \le \min\{p^*, 1-p^*\}, \\ 0, & \Delta > \min\{p^*, 1-p^*\}. \end{cases}$$

Now we assert

Theorem 2: For every $E > 0$ and $\Delta \ge 0$

$$R_{BH}(E,\Delta) = \begin{cases} H_{P_E}(X) - H_\Delta(X), & p^* \notin [\alpha_1, \alpha_2],\ 0 \le \Delta \le \max\limits_{P\in\alpha(E)} \min\{p, 1-p\}, \\ 1 - H_\Delta(X), & p^* \in [\alpha_1, \alpha_2],\ \Delta \le \max\limits_{P\in\alpha(E)} \min\{p, 1-p\}, \\ 0, & \Delta > \max\limits_{P\in\alpha(E)} \min\{p, 1-p\}, \end{cases} \tag{3}$$

where

$$[\alpha_1, \alpha_2] = \left[\frac{\exp\{E\} - \sqrt{\exp\{2E\} - 1}}{\exp\{E+1\}}; \frac{\exp\{E\} + \sqrt{\exp\{2E\} - 1}}{\exp\{E+1\}}\right]$$

and

$$D(P_E \parallel P^*) = E.$$

Proof: From (1) and (2) we can derive

$$R_{BH}(\Delta) = \begin{cases} \max\limits_{P\in\alpha(E)} (H_P(X) - H_\Delta(X)), & 0 \le \Delta \le \max\limits_{P\in\alpha(E)} \min\{p, 1-p\}, \\ 0, & \Delta > \max\limits_{P\in\alpha(E)} \min\{p, 1-p\}. \end{cases}$$

Let $0 \le \Delta \le \max\limits_{P\in\alpha(E)} \min\{p, 1-p\}$. Our task is simplification of

$$\max_{P\in\alpha(E)} (H_P(X) - H_\Delta(X)) = \max_{P\in\alpha(E)} H_P(X) - H_\Delta(X).$$

Note that if PD $\{1/2, 1/2\} \in \alpha(E)$, then

$$\max_{P\in\alpha(E)} H_P(X) = 1,$$

which takes place, when

$$D(P \parallel P^*) = p \log \frac{p}{p^*} + (1-p) \log \frac{1-p}{1-p^*} \le E.$$

The last inequality reduces to

$$\frac{1}{2}(\log \frac{1}{2p^*} + \log \frac{1}{2(1-p^*)}) \le E,$$

$$\log \frac{1}{4p^*(1-p^*)} \le 2E,$$

$$p^*(1-p^*) \geqslant 2^{-2(E+1)} \tag{4}$$

Therefore, if the last inequality holds, then

$$\max_{P:D(P\|P^*)\le E} H_P(X) = 1.$$

The condition (4) may be rewritten as a quadratic inequality

$$p^{*2} - p^* + 2^{-2(E+1)} \le 0,$$

which is equivalent to

$$p^* \in [\alpha_1, \alpha_2] = \left[\frac{\exp\{E\} - \sqrt{\exp\{2E\} - 1}}{\exp\{E+1\}}; \frac{\exp\{E\} + \sqrt{\exp\{2E\} - 1}}{\exp\{E+1\}}\right]$$

It means that the value of $R_{\mathrm{BH}}(E,\Delta)$ is constant and equals $1 - H_\Delta(X)$ for all generic PDs from $[\alpha_1, \alpha_2]$, which with $E \to 0$ tends to the segment $[0;1]$.
Now let us consider the case, when $p^* \notin [\alpha_1, \alpha_2]$. We show that

$$\max_{P\in\alpha(E)} H_P(X) = H_{P_E}(X), \text{ where } P_E = \{p_E, 1-p_E\} \text{ and } D(P_E \parallel P^*) = E, \tag{5}$$

assuming p_E the nearest 1/2 value which results from equation $D(P_E \parallel P^*) = E$.
The reference (5) will be true by the following argument.
Lemma: The function

$$D(P \parallel P^*) = p \log \frac{p}{p^*} + (1-p)\log\frac{1-p}{1-p^*}$$

is a monoton function of p for P from $\alpha(E)$.
Proof: Let $P_1 = \{p_1, 1-p_1\}$ and $P_2 = \{p_2, 1-p_2\}$ are some binary PD and $p^* < p_1 \le p_2$. It is required to prove the inequality $D(P_2 \parallel P^*) \ge D(P_1 \parallel P^*)$. The set $\alpha(E)$ is convex by P, that is if $P' \in \alpha(E)$ and $P'' \in \alpha(E)$, then $D(\lambda P' + (1-\lambda)P'' \parallel P^*) \in \alpha(E)$, because

$$D(\lambda P' + (1-\lambda)P'' \parallel P^*) \le \lambda D(P' \parallel P^*) + (1-\lambda)D(P'' \parallel P^*) \le \lambda E + (1-\lambda)E = E. \tag{6}$$

We can represent $P_1 = \lambda P^* + (1-\lambda)P_2$ $(0 < \lambda < 1)$ and as in (6), write down

$$D(P_1 \parallel P^*) \le D(\lambda P^* + (1-\lambda)P_2 \parallel P^*) \le \lambda D(P \parallel P^*) + (1-\lambda)D(P_2 \parallel P^*) \le$$

$$\le (1-\lambda)D(P_2 \parallel P^*) \le D(P_2 \parallel P^*).$$

Therefore, lemma is proved and (5) yields, which gives us (3).
Theorem 3: $R_{\mathrm{BH}}(E,\Delta)$ is concave function of E.
Proof: First note that

$$\lim_{E\to 0} R_{\mathrm{BH}}(E,\Delta) = \lim_{E\to 0} \left(\max_{P\in\alpha(E)} H_P(X) - H_\Delta(X)\right) =$$

$$= H_{P^*}(X) - H_\Delta(X) = R_{\mathrm{BH}}(\Delta)$$

and for fixed $\Delta \ge 0$ and P^* there exists a value E_0 such that if $E \ge E_0$, then $R_{\mathrm{BH}}(E,\Delta)$ is constant and equals $1 - H_\Delta(X)$. Since $1 - H_\Delta(X)$ is the maximal value of binary Hamming RRDF, it remains to prove the concavity of $R_{\mathrm{BH}}(E,\Delta)$ at the interval $(0; E_0]$.

Let $0 < E_1 < E_2$ and

$$R(E_1, \Delta) = H_{P_{E_1}}(X) - H_\Delta(X), \; R(E_2, \Delta) = H_{P_{E_2}}(X) - H_\Delta(X)$$

where $D(P_{E_1} \parallel P^*) = E_1$, $D(P_{E_2} \parallel P^*) = E_2$.
For a $0 < \lambda < 1$ we have

$$R(\lambda E_1 - (1-\lambda)E_2, \Delta) = \max_{P \in \alpha(\lambda E_1 - (1-\lambda)E_2)} H_P(X) - H_\Delta(X) =$$

$$= H_{P_{\lambda E_1 - (1-\lambda)E_2}}(X) - H_\Delta(X) \overset{(a)}{\geq} H_{\lambda P_{E_1} + (1-\lambda)P_{E_2}}(X) - H_\Delta(X) \overset{(b)}{\geq}$$

$$\overset{(b)}{\geq} \lambda H_{P_{E_1}}(X) + (1-\lambda)H_{P_{E_2}}(X) - H_\Delta(X) =$$

$$= \lambda H_{P_{E_1}}(X) + (1-\lambda)H_{P_{E_2}}(X) - \lambda H_\Delta(X) - (1-\lambda)H_\Delta(X) =$$

$$= \lambda(H_{P_{E_1}}(X) - H_\Delta(X)) + (1-\lambda)(H_{P_{E_2}}(X) - H_\Delta(X)) =$$

$$= \lambda R(E_1, \Delta) + (1-\lambda)R(E_2, \Delta),$$

where (a) follows from inequality

$$D(\lambda P_{E_1} + (1-\lambda)P_{E_2} \parallel P^*) \leq \lambda D(P_{E_1} \parallel P^*) + (1-\lambda)D(P_{E_2} \parallel P^*) =$$

$$= \lambda E_1 + (1-\lambda)E_2.$$

and (b) follows from the concavity of entropy. So the Theorem 3 is proved.

References

[1] C. E. Shannon, "Coding theorems for a discrete source with a fidelity criterion", IRE Nat. Conv. Rec., pt. 4, pp. 142- 163, Mar. 1959.

[2] T. Berger, Rate Distortion Theory: A Mathematical Basis for Data Compression, Englewood Cliffs, NJ: Prentice-Hall, 1971.

[3] I. Csiszár and J. Körner, Information theory: Coding Theorems for Discrete Memoryless Systems, New York: Academic Press, 1981.

[4] R. Ahlswede, "Extremal properties of rate-distortion functions", IEEE Trans. on Infotm. Theory, vol. 36, no. 1, pp. 166–171, January 1990.

[5] E. A. Haroutunian and B. Mekoush, "Estimates of optimal rates of codes with given error probability exponent for certain sources", (in Russian), Abstr. of sixth Intern. Symposium on Inform. Theory, Tashkent, vol. 1, pp 22- 23, 1984.

[6] T. Cover, J. Thomas, Elements of Information Theory, New York, Wiley, 1991.