Rate-Reliability-Distortion Function for Source with Side Information Known to Encoder and Decoder

Mariam Haroutunian
IIAP NAS RA
Yerevan, Armenia
e-mail: armar@sci.am

Parandzem Hakobyan
IIAP NAS RA
Yerevan, Armenia
e-mail: par_h@iiap.sci.am

Jemma Santrosyan
Vanadzor State University
Vanadzor, Armenia
e-mail: j.santrosian@gmail.com

***Abstract*—In this paper, we consider the source model with side information available at the encoder and decoder. The Rate-reliability-distortion function is stated by constructing the lower and upper bounds.**

***Keywords*—Rate-reliability-distortion function, data compression, lossy source coding.**

I. Introduction

Source coding with side information is a foundational topic in information theory with applications in data compression, distributed systems, and secure communications.

Source coding with side information refers to scenarios, where the encoder or decoder has access to additional correlated information, called side information, which can assist in compressing the source more efficiently. The study of this topic originated with Shannon [1] and was expanded significantly by Slepian and Wolf [2], Wyner [3], Wyner and Ziv [4], and many others.

The models of the source with side information have applications in various fields, including:

- Sensor Networks: Efficient compression using correlation between sensors,
- Multimedia: Video compression using past frames or channel feedback as side information,
- Wireless Communications: Cooperative and relay networks,
- Privacy-Preserving Compression: Secure data transmission and storage,
- Machine Learning: Federated learning and distributed inference.

The central concept in lossy source coding is the **rate-distortion function**, which characterizes the trade-off between the rate at which information is transmitted (measured in bits per symbol) and the distortion (or loss of fidelity) that occurs during the lossy compression of a source. It is interpreted as follows: for higher distortion, the rate is less, hence one can compress more, losing more accuracy; for lower distortion, the rate is higher, hence one needs more bits to preserve the quality.

The next characteristic of the source is the **rate-reliability-distortion function**, the coding rate as a function of given distortion level and error exponent or reliability E. This function has been studied for various source models [5].

Here, we study the rate-reliability-distortion function for the source model with side information available at the encoder and decoder.

II. Notations and Definitions

We consider the system shown in Fig.1. The Discrete Memoryless Source (DMS) with state s is defined as a sequence $\{(X_i, S_i)\}_{i=1}^{\infty}$ of jointly distributed with distribution $P^*(x,s)$ random variables X and S taking values in finite sets $\mathcal{X}$ and $\mathcal{S}$. The finite set $\hat{\mathcal{X}}$, different in general from $\mathcal{X}$, is the reproduction alphabet at the receiver.

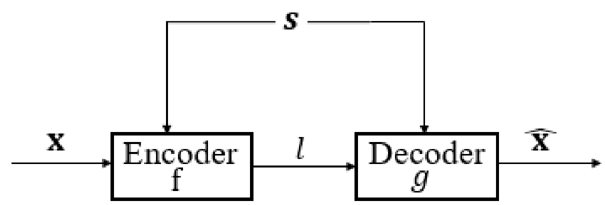

Fig. 1. The model of DMS with side information available at the encoder and decoder

We are given a distortion measure

$$d : \mathcal{X} \times \hat{\mathcal{X}} \to [0; \infty)$$

between source and reconstruction messages. The distortion measure for N-length sequences is the average of the components' distortions

$$d(\mathbf{x}, \hat{\mathbf{x}}) = \frac{1}{N} \sum_{n=1}^{N} d(x, \hat{x}).$$

The task of this system is to ensure the restoration of the source messages at the receiver within a given distortion level Δ and with a small error probability in the case when the state sequence is available to both the encoder and the decoder. A code (f_N, g_N) for this case is defined by a pair of mappings: a coding

$$f_N : \mathcal{X}^N \times \mathcal{S}^N \to \{1, 2, ..., L(N)\},$$

and decoding

$$g_N : \{1, 2, ..., L(N)\} \times \mathcal{S}^N \to \hat{\mathcal{X}},$$

https://doi.org/10.51408/csit2025_38

where $L(N)$ is the code volume. The code rate is

$$R(f_N, g_N) = \frac{1}{N} \log L(N).$$

Throughout this paper, all log-s and exp-s are of base 2.

We consider the memoryless source, which means that for N-length vector pairs $\mathbf{x} = (x_1, x_2, ..., x_N) \in \mathcal{X}^N$ and $\mathbf{s} = (s_1, s_2, ..., s_N) \in \mathcal{S}^N$

$$P^{*N}(\mathbf{x}, \mathbf{s}) = \prod_{n=1}^{N} P^*(x, s).$$

We will use the following distributions:

$$P \triangleq \{P(x, s) = P_0(s)P_1(x|s), x \in \mathcal{X}, s \in \mathcal{S}\}$$

and the conditional distribution Q

$$Q \triangleq \{Q(\hat{x}|x, s), x \in \mathcal{X}, \hat{x} \in \mathcal{S}, s \in \mathcal{S}\}.$$

For the formulation of the result, we remind the following definitions [6].

The **entropy** of RV S is

$$H_{P_0}(S) \triangleq -\sum_{s} P_0(s) \log P_0(s).$$

The **conditional entropy** of RV X relative to the random variable S is

$$H_P(X|S) \triangleq -\sum_{s,x} P(x, s) \log P_1(x|s).$$

The **conditional mutual information** of RV X and $\hat{X}$ relative to the random variable S is

$$I_{P,Q}(X; \hat{X}|S) \triangleq$$

$$\sum_{x \in \mathcal{X},\ \hat{x} \in X, s \in \mathcal{S}} P(x, s)Q(\hat{x}|x, s) \log \frac{Q(\hat{x}|x, s)}{PQ(\hat{x}|s)},$$

where

$$PQ(\hat{x}|s) = \sum_{x \in \mathcal{X}} P_1(x|s)Q(\hat{x}|x, s).$$

The following property will be used below:

$$I_{P,Q}(X; \hat{X}|S) = H_P(X|S) - H_{P,Q}(X|\hat{X}, S).$$

The **Kullback-Leibler divergence** between the distributions P and P^* is defined as follows:

$$D(P||P^*) \triangleq \sum_{x \in \mathcal{X}, s \in \mathcal{S}} P(x, s) \log \frac{P(x, s)}{P^*(x, s)}.$$

The set of satisfactorily transmitted vectors for the given $\mathbf{s}$, which are reconstructed within the distortion constraint $\Delta \geq 0$ is as follows:

$$\mathcal{A}(\mathbf{s}) = \{\mathbf{x} : g_N(f_N(\mathbf{x}, \mathbf{s}), \mathbf{s}) = \hat{\mathbf{x}},\ d(\mathbf{x}, \hat{\mathbf{x}}) \leq \Delta\}.$$

The **error probability** of the code (f_N, g_N) is defined as:

$$e(f_N, g_N, P^*, \Delta) = 1 - \min_{\mathbf{s} \in \mathcal{S}^N} P^{*N}(\mathcal{A}(\mathbf{s})).$$

$R \geq 0$ is called (E, Δ)**-achievable rate** for given P^*, $E > 0$ and $\Delta \geq 0$, if for every $\epsilon > 0, \delta > 0$, there exists a code (f_N, g_N) such that

$$\frac{1}{N} \log L(N) \leq R + \epsilon$$

and the error probability is exponentially small

$$e(f_N, g_N, P^*, \Delta) \leq \exp\{-N(E - \delta)\}.$$

The minimum (E, Δ)-achievable rate for given PD P^* is called the **rate-reliability-distortion function** and denoted by $R(E, \Delta, P^*)$.

III. Main Result

Consider the following set of joint distributions P on $\mathcal{X} \times \mathcal{S}$:

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

Let $\mathcal{Q}(P, \Delta)$ be the set of all conditional PDs $Q_P(\hat{x}|x, s) = Q_P$, corresponding to the PD P, for which the following conditions hold:

$$\mathbf{E}d(X, \hat{X}) = \sum_{x,s,\hat{x}} P(x, s)Q_P(\hat{x}|x, s)d(x, \hat{x}) \leq \Delta. \qquad (1)$$

The main result is the following theorem.

Theorem 1: For given P^*, every $E > 0$ and $\Delta \geq 0$,

$$R(E, \Delta, P^*) = \max_{P \in \alpha(E, P^*)} \min_{Q_P \in \mathcal{Q}(P, \Delta)} I_{P,Q_P}(X; \hat{X}|S).$$

For the proof of Theorem 1, we will use the following modification of the Covering Lemma [5], which is based on the method of types (for definitions and properties, we referee to [7]).

Lemma 1: (Covering Lemma) Let for $\epsilon > 0$

$$J(P, Q) = \exp\{N(I_{P,Q}(X; \hat{X}|S) + \epsilon)\}.$$

Then, for every type P_0, state sequence $\mathbf{s} \in \mathcal{T}_{P_0}^N(S)$, conditional types P_1 and Q, there exists a collection of vectors

$$\{\hat{\mathbf{x}}_j \in \mathcal{T}_{P,Q}^N(\hat{X}|\mathbf{s}), j = 1, ..., J(P, Q)\},$$

such that the set

$$\{\mathcal{T}_{P,Q}^N(X|\hat{\mathbf{x}}_j, \mathbf{s}), j = 1, ..., J(P, Q)\},$$

covers $\mathcal{T}_P^N(X|\mathbf{s})$ for N large enough, that is

$$\mathcal{T}_P^N(X|\mathbf{s}) \subset \bigcup_{j=1}^{J(P,Q)} \mathcal{T}_{P,Q}^N(X|\hat{\mathbf{x}}_j, \mathbf{s}).$$

We omit the proof of Lemma 1, since it is obvious.

Proof of Theorem 1. The proof of the theorem consists of two parts.

First, we will show that

$$R(E, \Delta, P^*) \leq \max_{P \in \alpha(E, P^*)} \min_{Q_P \in \mathcal{Q}(P, \Delta)} I_{P,Q_P}(X; \hat{X}|S). \qquad (2)$$

Let us represent the set of all source messages of length N as follows:

$$\mathcal{X}^N \times \mathcal{S}^N = \bigcup_{P \in \mathcal{P}_N(\mathcal{X} \times \mathcal{S})} \mathcal{T}_P^N(X, S)$$

$$= \bigcup_{P_0 \in \mathcal{P}_N(\mathcal{S})} \bigcup_{P_1 \in \mathcal{P}_N(\mathcal{X}, P_0)} \mathcal{T}_P^N(X, S),$$

where $\mathcal{P}_N(S)$ is the set of possible types P_0 of vectors $\mathbf{s} \in \mathcal{S}^{\mathbf{N}}$, $\mathcal{P}_N(X \times S)$ is the set of possible types P of pairs $(\mathbf{x}, \mathbf{s}) \in \mathcal{X}^{\mathbf{N}} \times \mathcal{S}^{\mathbf{N}}$ and $\mathcal{P}_N(\mathcal{X}, P_0)$ is the set of all possible conditional types P_1 for $\mathbf{s}$ of type P_0.

For each $\delta > 0$ and N large enough, the estimation of the probability of occurrence of a source of types beyond $\alpha(E + \delta, P^*)$ can be estimated using the properties of the types and the definition of the set $\alpha(E, P^*)$:

$$\begin{aligned}
&P^{*N}\left(\bigcup_{P \notin \alpha(E+\delta, P^*)} \mathcal{T}_P^N(X, S)\right) \\
=& \sum_{P \notin \alpha(E+\delta, P^*)} P^{*N}\left(\mathcal{T}_P^N(X, S)\right) \\
\leq& (N+1)^{|\mathcal{X}||\mathcal{S}|} \exp\left\{-N \min_{P \notin \alpha(E+\delta, P^*)} D(P||P*)\right\} \\
\leq& \exp\{-NE - N\delta + |\mathcal{X}||\mathcal{S}| \log(N+1)\} \\
\leq& \exp\{-N(E + \delta/2)\}.
\end{aligned}$$

Consequently, it is sufficient to show the existence of a code with required error probability $e(f_N, g_N, P^*, \Delta)$ for the vectors with type P from $\alpha(E + \delta, P^*)$. For each $\Delta \geq 0$, let us pick some types P_0, P_1 such that $P \in \alpha(E + \delta, P^*)$ and some $Q_P \in \mathcal{Q}(P, \Delta)$. Let for each $\mathbf{s} \in \mathcal{T}_{P_0}^N(S)$

$$C(P, Q_P, j) = \mathcal{T}_{P,Q_P}^N(X|\hat{\mathbf{x}}_j, \mathbf{s}) - \bigcup_{j' < j} \mathcal{T}_{P,Q_P}^N(X|\hat{\mathbf{x}}_{j'}, \mathbf{s}),$$

$$j = \overline{1, J(P, Q_P)}.$$

We define a code (f_N, g_N) for each vector $\mathbf{s}$ with the encoding:

$$f_N(\mathbf{x}|\mathbf{s}) = \begin{cases} j, & \text{when } \mathbf{x} \in C(P, Q_P, j),\ P \in \alpha(E + \delta, P^*), \\ j_0, & \text{when } \mathbf{x} \in \mathcal{T}_P^N(X|\mathbf{s}),\ P \notin \alpha(E + \delta, P^*), \end{cases}$$

and the decoding

$$g_N(j|\mathbf{s}) = \hat{\mathbf{x}}_j, \qquad g_N(j_0|\mathbf{s}) = \hat{\mathbf{x}}_0,$$

where the number j_0 and the reconstruction vector $\hat{\mathbf{x}}_0$ are fixed. Obviously, with such code, an error occurs only when the number j_0 is sent.

According to the definition of the code and the inequality (1), for $P \in \alpha(E + \delta, P^*)$ and $Q_P \in \mathcal{Q}(P, \Delta)$ we have:

$$\begin{aligned}
d(\mathbf{x}, \hat{\mathbf{x}}_{\mathbf{j}}) &= \sum_{x,s,\hat{x}} P(x, s) Q_P(\hat{x}|x, s) d(x, \hat{x}) \\
&= \mathbf{E}_{P,Q_P} d(X, \hat{X}) \leq \Delta, \quad j = \overline{1, J(P, Q_P)}.
\end{aligned}$$

According to Lemma 1, the number of vectors $\hat{\mathbf{x}}$ for each $\mathbf{s}$, type P and corresponding conditional type $Q_P \in \mathcal{Q}(P, \Delta)$ is:

$$L_{P,Q_P}(N) = \exp\{N(I_{P,Q}(X; \hat{X}|S) + \epsilon)\}.$$

Then, taking into account that the number of types has a polynomial estimate, for the transmitter rate $L(N)$, we find the following estimation:

$$\begin{aligned}
L(N) \leq& \sum_{P \in \alpha(E+\delta, P^*)} \min_{Q_P \in \mathcal{Q}(P,\Delta)} L_{P,Q_P}(N) \leq (N+1)^{|\mathcal{X}||\mathcal{S}|} \\
&\times \max_{P \in \alpha(E+\delta, P^*)} \min_{Q_P \in \mathcal{Q}(P,\Delta)} \exp\left\{N(I_{P,Q}(X; \hat{X}|S) + \epsilon)\right\}.
\end{aligned}$$

Hence, the corresponding limit for the transmission rate is:

$$\begin{aligned}
&\frac{1}{N} \log L(N) - \epsilon - \frac{1}{N}|\mathcal{X}||\mathcal{S}| \log(N+1) \leq \\
&\leq \max_{P \in \alpha(E+\delta, P^*)} \min_{Q_P \in \mathcal{Q}(P,\Delta)} I_{P,Q}(X; \hat{X}|S). \qquad (3)
\end{aligned}$$

Taking into account the arbitrariness of ϵ and δ and the continuity of the information expression (3), we get (2).

Now, we pass to the inverse part, we shall prove that:

$$R(E, \Delta, P^*) \geq \max_{P \in \alpha(E, P^*)} \min_{Q_P \in \mathcal{Q}(P,\Delta)} I_{P,Q_P}(X; \hat{X}|S). \quad (4)$$

Let $\epsilon > 0$ be fixed. Consider a code (f_N, g_N) for each blocklength N with (E, Δ)-achievable rate R. We must show that for some $Q_P \in \mathcal{Q}(P, \Delta)$ the following inequalities hold for N large enough:

$$\frac{1}{N} \log L(N) + \epsilon \geq \max_{P \in \alpha(E, P^*)} I_{P,Q_P}(X; \hat{X}|S). \qquad (5)$$

Let for each state sequence $\mathbf{s}$, $\mathcal{A}'(\mathbf{s})$ be the complement of the set $\mathcal{A}(\mathbf{s})$. The following statement is true:

$$\left|\mathcal{A}(\mathbf{s}) \bigcap \mathcal{T}_P^N(X|\mathbf{s})\right| = \left|\mathcal{T}_P^N(X|\mathbf{s})\right| - \left|\mathcal{A}'(\mathbf{s}) \bigcap \mathcal{T}_P^N(X|\mathbf{s})\right|.$$

For $P \in \alpha(E - \epsilon, P^*)$

$$\begin{aligned}
&\left|\mathcal{A}'(\mathbf{s}) \bigcap \mathcal{T}_P^N(X|\mathbf{s})\right| = \frac{P^{*N}(\mathcal{A}'(\mathbf{s}) \bigcap \mathcal{T}_P^N(X|\mathbf{s}))}{P^{*N}(\mathbf{x}|\mathbf{s})} \\
\leq& \exp\{N(H_P(X|S) + D(P||P^*))\} \exp\{-N(E - \epsilon)\} \\
\leq& \exp\{N(H_P(X|S) - \epsilon)\}.
\end{aligned}$$

Hence,

$$\begin{aligned}
&\left|\mathcal{A}(\mathbf{s}) \bigcap \mathcal{T}_P^N(X|\mathbf{s})\right| \\
&\geq (N+1)^{-|\mathcal{X}||\mathcal{S}|} \exp\{NH_P(X|S)\} \\
&- \exp\{N(H_P(X|S) - \epsilon)\} \\
&= \exp\{N(H_P(X|S) - \epsilon)\} \left(\frac{\exp\{N\epsilon\}}{(N+1)^{|\mathcal{X}||\mathcal{S}|}} - 1\right) \qquad (6) \\
&\geq \exp\{N(H_P(X|S) - \epsilon)\}.
\end{aligned}$$

For each $\mathbf{x} \in \mathcal{A}(\mathbf{s}) \bigcap \mathcal{T}_P^N(X|\mathbf{s})$ corresponds a unique vector $\hat{\mathbf{x}}$ such that

$$\hat{\mathbf{x}} = g_N(f_N(\mathbf{x}, \mathbf{s}), \mathbf{s}) \text{ and } \hat{\mathbf{x}} \in \mathcal{T}_{P,Q}^N(\hat{X}|\mathbf{x}, \mathbf{s}).$$

Let us divide the set of all vectors $\left|\mathcal{A}(\mathbf{s})\bigcap\mathcal{T}_P^N(X|\mathbf{s})\right|$ into subsets by conditional types Q_P. The class having maximum cardinality for given P, we denote by

$$\left(\left|\mathcal{A}(\mathbf{s})\bigcap\mathcal{T}_P^N(X|\mathbf{s})\right|\right)_{Q_P}.$$

According to the number of conditional types Q, for sufficiently large N, we have:

$$\begin{aligned}&\left|\mathcal{A}(\mathbf{s})\bigcap\mathcal{T}_P^N(X|\mathbf{s})\right|\\ &\leq (N+1)^{|\mathcal{X}||\mathcal{S}|}\left(\left|\mathcal{A}(\mathbf{s})\bigcap\mathcal{T}_P^N(X|\mathbf{s})\right|\right)_{Q_P}\\ &\leq \exp\{N\epsilon/2\}\left(\left|\mathcal{A}(\mathbf{s})\bigcap\mathcal{T}_P^N(X|\mathbf{s})\right|\right)_{Q_P}. \end{aligned} \tag{7}$$

Let

$$\mathcal{D}(\mathbf{s})=\{\hat{\mathbf{x}}: g_N(f_N(\mathbf{x},\mathbf{s}),\mathbf{s})=\hat{\mathbf{x}}\ ,$$
$$\text{for some}\ \mathbf{x}\in\mathcal{A}(\mathbf{s})\bigcap\mathcal{T}_P^N(X|\mathbf{s})\bigcap\mathcal{T}_{P,Q_P}^N(X|\hat{\mathbf{x}},\mathbf{s})\Big\}.$$

From the definition of the code $|\mathcal{D}(\mathbf{s})|\leq L(N)$, then

$$\begin{aligned}&\left|\left(\mathcal{A}(\mathbf{s})\bigcap\mathcal{T}_P^N(X|\mathbf{s})\right)\right|_{Q_P}\\ &\leq \sum_{\hat{\mathbf{x}}\in\mathcal{D}(\mathbf{s})}\left|\mathcal{T}_{P,Q}^N(X|\hat{\mathbf{x}},\mathbf{s})\right| \\ &\leq L(N)\exp\{NH_{P,Q_P}(X|\hat{X},S)\}.\end{aligned} \tag{8}$$

From (6-8) follows

$$L(N)\geq\exp\{N(I_{P,Q_P}(X;\hat{X}|S)-\epsilon)\},$$

for each $P\in\alpha(E-\epsilon,P*)$ and some Q_P for which $\mathbf{E}_{P,Q_P}d(X,\hat{X})$. Theorem 1 is proved.

IV. Conclusion

In this paper, the source model with side information available at the encoder and decoder is studied. The analytical form of the Rate-reliability-distortion function is obtained by constructing the lower and the upper bounds.

References

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

[2] D. Slepyan and J. K. Wolf, "Noiseless coding of correlated information sources", *IEEE Transactions on Information Theory*, vol. 19, no.6, pp. 471 – 480, 1973.

[3] A. D. Wyner, "On source coding with side information at the decoder", *IEEE Transactions on Information Theory*, vol. 21, no. 5, pp. 294 – 300, 1975.

[4] A. D. Wyner and J.Ziv, "The rate distortion function for source coding with side information at the decoder", *IEEE Transactions on Information Theory*, vol. 22, no. 1, pp. 1 – 10, 1976.

[5] E. Haroutunian, M. Haroutunian and A. Harutyunyan, "Reliability Criteria in Information Theory and in Statistical Hypothesis Testing", *Foundations and Trends in Communications and Information Theory*, vol. 4, nos 2-3, pp. 97–263, 2007. doi: 10.1561/0100000008

[6] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, Second Edition. Wiley, New York, 2006.

[7] I. Csiszár, "Method of types", *IEEE Transactions on Information Theory*, vol. 44, no. 6, pp. 2505—2523, 1998.