ComEX

Introducing the perception-distortion tradeoff into the rate-distortion theory of general information sources*

Ryutaroh Matsumoto[a]

Department of Information and Communication Engineering, Nagoya University,
Furo-cho, Chitane-ku, Nagoya, Aichi 464–8603, Japan
a) *ryutaroh.matsumoto@nagoya-u.jp*

Abstract: Blau and Michaeli recently introduced a novel concept for inverse problems of signal processing, that is, the perception-distortion tradeoff. We introduce their tradeoff into the rate distortion theory of lossy source coding in information theory, and clarify the tradeoff among information rate, distortion and perception for general information sources.
Keywords: perception-distortion tradeoff, rate-distortion theory, data compression
Classification: Fundamental Theories for Communications

References

[1] Y. Blau and T. Michaeli, "The perception-distortion tradeoff," Proc. 2018 IEEE/ CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, Utah, USA, pp. 6228–6237, June 2018.
[2] T. S. Han, *Information-Spectrum Methods in Information Theory*, Springer, 2002. DOI:10.1007/978-3-662-12066-8

1   Introduction

An inverse problem of signal processing is to reconstruct the original information from its degraded version. It is not limited to image processing, but it often arises in the image processing. When a natural image is reconstructed, the reconstructed image sometimes does not look natural while it is close to the original image by a reasonable metric, for example mean squared error. When the reconstructed information is close to the original, it is often believed that it should also look natural.

Blau and Michaeli [1] questioned this unproven belief. In their research [1], they mathematically formulated the *naturalness* of the reconstructed information by a distance of the probability distributions of the reconstructed information and the original information. The reasoning behind this is that the perceptual quality of a

---

*The result in this letter will be presented at the 41st Symposium on Information Theory and its Applications, Fukushima, Japan, 18–21 December 2018.

reconstruction method is often evaluated by how often a human observer can distinguish an output of the reconstruction method from natural ones. Such a subjective evaluation can mathematically be modeled as a hypothesis testing [1]. A reconstructed image is more easily distinguished as the variational distance $\sigma(P_R, P_N)$ increases [1], where P_R is the probability distribution of the reconstructed information and P_N is that of the natural one. They regard the perceptional quality of reconstruction as a distance between P_R and P_N. The distance between the reconstructed information and the original information is conventionally called as distortion. They discovered that there exists a tradeoff between perceptional quality and distortion, and named it as the *perception-distortion tradeoff*.

Claude Shannon [2, Chapter 5] initiated the rate-distortion theory in 1950's. It clarifies the tradeoff between information rate and distortion in the lossy source coding (lossy data compression). The rate-distortion theory has served as a theoretical foundation of image coding for past several decades, as drawing a rate-distortion curve is a common practice in research articles of image coding. Since distortion and perceptional quality are now considered two different things, it is natural to consider a tradeoff among information rate, distortion and perceptional quality. Blau and Michaeli [1] briefly mentioned the rate-distortion theory, but they did not clarify the tradeoff among the three.

The purpose of this letter is to mathematically define the tradeoff for general information sources, and to express the tradeoff in terms of information spectral quantities introduced by Han and Verdú [2]. It should be noted that the tradeoff among the three quantities can be regarded as a combination of lossy source coding problem [2, Chapter 5] and random number generation problem [2, Chapter 2], both of which will be used to derive the tradeoff.

Since the length limitation is strict in this journal, citations to the original papers are replaced by those to the textbook [2], and the mathematical proof is a bit compressed. The author begs readers' kind understanding. The base of log is an arbitrarily fixed real number >1 unless otherwise stated.

2 Preliminaries

The following definitions are borrowed from Han's textbook [2]. Let

$$\mathbf{X} = \{X^n = (X_1^{(n)}, \ldots, X_n^{(n)})\}_{n=1}^{\infty}$$

be a general information source, where the alphabet of the random variable X^n is the n-th Cartesian product $\mathcal{X}^n$ of some finite alphabet $\mathcal{X}$. For a sequence of real-valued random variables $Z_1, Z_2, \ldots$ we define

$$\text{p-}\limsup_{n\to\infty} Z_n = \inf\{\alpha \mid \lim_{n\to\infty} \Pr[Z_n > \alpha] = 0\}.$$

For two general information sources $\mathbf{X}$ and $\mathbf{Y}$ we define

$$\bar{I}(\mathbf{X}; \mathbf{Y}) = \text{p-}\limsup_{n\to\infty} \frac{1}{n} \log \frac{P_{X^n Y^n}(X^n, Y^n)}{P_{X^n}(X^n) P_{Y^n}(Y^n)},$$

and

$$F_{\mathbf{X}}(R) = \limsup_{n\to\infty} \Pr\left[\frac{1}{n} \log \frac{1}{P_{X^n}(X^n)} \geq R\right].$$

For two distributions P and Q on an alphabet $\mathcal{X}$, we define the variational distance $\sigma(P, Q)$ as $\sum_{x \in \mathcal{X}} |P(x) - Q(x)|/2$. In the rate-distortion theory, we usually assume a reconstruction alphabet different from a source alphabet. In order to consider the distribution similarity of reconstruction, in this letter we assume $\mathcal{X}^n$ as both source and reconstruction alphabets.

An encoder of length n is a mapping $f_n : \mathcal{X}^n \to \{1, \ldots, M_n\}$, and the corresponding decoder of length n is a mapping $g_n : \{1, \ldots, M_n\} \to \mathcal{X}^n$. $\delta_n : \mathcal{X}^n \times \mathcal{X}^n \to [0, \infty)$ is a general distortion function with the assumption $\delta_n(x^n, x^n) = 0$ for all n and $x^n \in \mathcal{X}^n$.

Definition 1 *A triple (R, D, S) is said to be achievable if there exists a sequence of encoder and decoder $(f_n,\ g_n)$ such that*

$$\limsup_{n \to \infty} \frac{\log M_n}{n} \leq R, \tag{1}$$

$$p\text{-}\limsup_{n \to \infty} \frac{1}{n} \delta_n(X^n, g_n(f_n(X^n))) \leq D, \tag{2}$$

$$\limsup_{n \to \infty} \sigma(P_{g_n(f_n(X^n))}, P_{X^n}) \leq S. \tag{3}$$

Define the function $R(D, S)$ by

$$R(D, S) = \inf\{R \mid (R, D, S) \text{ is achievable}\}.$$

Theorem 2

$$R(D, S) = \max\left\{ \inf_{\mathbf{Y}} \overline{I}(\mathbf{X}; \mathbf{Y}), \inf\{R \mid F_{\mathbf{X}}(R) \leq S\} \right\}$$

where the infimum is taken with respect to all general information sources $\mathbf{Y}$ satisfying

$$p\text{-}\limsup_{n \to \infty} \frac{1}{n} \delta_n(X^n, Y^n) \leq D. \tag{4}$$

Proof: Let a pair of encoder f_n and decoder g_n satisfies Eqs. (1)–(3). Then by [2, Theorem 5.4.1] we have

$$R \geq \inf_{\mathbf{Y}} \overline{I}(\mathbf{X}; \mathbf{Y}), \tag{5}$$

where $\mathbf{Y}$ satisfies Eq. (4). On the other hand, the decoder g_n can be viewed as a random number generator to X^n from the alphabet $\{1, \ldots, M_n\}$. By [2, Converse part of the proof of Theorem 2.4.1] we have

$$R \geq \inf\{R \mid F_{\mathbf{X}}(R) \leq S\}. \tag{6}$$

This complete the converse part of the proof.

We start the direct part of the proof. Assume that a triple (R, D, S) satisfys Eqs. (5) and (6). Let M_n satisfy

$$\limsup_{n \to \infty} \frac{1}{n} \log M_n \leq R. \tag{7}$$

Let $f_n^{(1)}$ and $g_n^{(1)}$ be an encoder and a decoder constructed in [2, Lemma 1.3.1] with codebook $\{1, \ldots, M_n\}$. Let $f_n^{(2)}$ and $g_n^{(2)}$ be an encoder and a decoder constructed in [2, Theorem 5.4.1] with codebook $\{M_n + 1, \ldots, 2M_n\}$. Assume that we have a

source sequence $x^n \in \mathcal{X}^n$. If $-\log P_{X^n}(x^n) < \log M_n$ then let $f_n^{(1)}(x^n) \in \{1, \ldots, M_n\}$ be the codeword. If $-\log P_{X^n}(x^n) \geq \log M_n$ then let $f_n^{(2)}(x^n) \in \{M_n + 1, \ldots, 2M_n\}$ be the codeword. Let f_n be the above encoding process. At the receiver of a codeword $1 \leq m \leq 2M_n$, if $m \leq M_n$ then decode m by $g_n^{(1)}$, otherwise decode m by by $g_n^{(2)}$. Let the above decoding process as g_n.

If $f_n^{(1)}$ and $g_n^{(1)}$ are used then the source sequence x^n is reconstructed by a receiver without error by [2, Lemma 1.3.1] and we have $\delta_n(x^n, g_n(f_n(x^n))) = 0$. The probability ϵ_n of $f_n^{(1)}$ and $g_n^{(1)}$ *not* being used is

$$\epsilon_n \leq \Pr\left[\frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} \geq \frac{1}{n}\log M_n\right].$$

Combined with the assumption $R \geq \inf\{R \mid F_{\mathbf{X}}(R) \leq S\}$ and Eq. (7)

$$\limsup_{n\to\infty} \epsilon_n \leq S,$$

which implies Eq. (3).

On the other hand, $f_n^{(2)}$ and $g_n^{(2)}$ satisfy Eq. (2), so the combined encoder f_n and g_n also satisfies Eq. (2). The information rate of f_n is at most $R + \frac{\log 2}{n}$, which implies that Eq. (1) holds with the constructed f_n and g_n. This completes the direct part of the proof. $\square$

3 Example with a mixed information source

A typical example of non-ergordic general information source is a mixed information source [2, Section 1.4]. Since Theorem 2 is a bit abstract, we explicitly compute $R(D, S)$ for a mixed information source. Let $\mathcal{X} = \{0, 1\}$, and $\delta_n(x^n, y^n)$ be the Hamming distance between $x^n, y^n \in \mathcal{X}^n$. Consider two distributions P and Q on $\mathcal{X}$ defined by

$$P(0) = 1/2, P(1) = 1/2,$$
$$Q(0) = 1/4, Q(1) = 3/4.$$

For $x^n = (x_1, \ldots, x_n)$, in our mixed information source we have

$$\Pr[X^n = x^n] = \frac{1}{2}\prod_{i=1}^{n} P(x_i) + \frac{1}{2}\prod_{i=1}^{n} Q(x_i).$$

By [2, Theorem 5.8.1, Example 5.8.1 and Theorem 5.10.1] we see that

$$R \geq \inf_{\mathbf{Y}:\text{Eq. (4) holds}} \overline{I}(\mathbf{X}; \mathbf{Y})$$

if and only if

$$R \geq h(1/2) - h(D), \qquad (8)$$

where $h(u)$ is the binary entropy function $-u\log u - (1 - u)\log(1 - u)$.

On the other hand, by [2, Example 1.6.1], we have

$$F_{\mathbf{X}}(R) = \begin{cases} 1 & \text{if } R < h(1/4), \\ 1/2 & \text{if } h(1/4) \leq R < 1, \\ 0 & \text{if } 1 \leq R. \end{cases}$$

By the above formulas and assuming $\log = \log_2$, we can see

$$R(D, S) = \begin{cases} 1 & \text{if } S = 0, \\ \max\{h(1/4), 1 - h(D)\} & \text{if } 0 < S \le 1/2, \\ 1 - h(D) & \text{if } 1/2 < S. \end{cases}$$

Acknowledgments