Extremal Properties of Rate–Distortion Functions

Rudolf Ahlswede

Abstract

At a recent workshop (Hakone, Japan, July 1–2, 1988) I. Csiszár stated that inspite of strong efforts the following question had been waiting for an answer for many years. It is true that for fixed distortion level Δ the rate–distortion function $R(P,\Delta)$ has in the distribution P no local maxima with value different from the global maximum? We show that in general the answer is negative. However, the answer is positive for Hamming distortion measures. Moreover, there R is Schur–concave.

1 Basic Concepts and Auxiliary Results

Let $(X_t)_{t=1}^{\infty}$ be a discrete memoryless source (DMS), that is, a sequence of independent and identically distributed random variables (RV's) with values in a finite set $\mathcal{X}$. We are also given a finite reconstruction space $\hat{\mathcal{X}}$ and a per letter distortion measure

$$d : \mathcal{X} \times \hat{\mathcal{X}} \to \mathbb{R}_+. \tag{1.1}$$

For a function F defined on a product space $\mathcal{Y}^n$ we use the notation

$$\text{rate}(F) = \frac{1}{n}\log_2 \|F\|, \ \|F\| = \text{ the cardinality of the range of } F. \tag{1.2}$$

For an encoding function $f_n : \mathcal{X}^n \to \hat{\mathcal{X}}^n$ we consider the reproduction

$$\hat{X}^n = (\hat{X}_1, \dots, \hat{X}_n) = f_n(X^n) \tag{1.3}$$

and the average distortion $\text{dist}(f_n)$, defined by

$$\text{dist}(f_n) = \mathbb{E}\frac{1}{n}\sum_{t=1}^{n} d(X_t, \hat{X}_t). \tag{1.4}$$

ρ is said to be an ε–achievable rate for distortion level Δ, if for all large n there are encoding functions f_n with

$$\text{rate}(f_n) \le \rho, \ \text{dist}(f_n) \le \Delta + \varepsilon. \tag{1.5}$$

Thus

$$R(\Delta) = \inf \left(\bigcap_{\varepsilon > 0} \{\rho : \rho \text{ is } \varepsilon\text{–achievable for } \Delta\} \right) \tag{1.6}$$

is the smallest rate which for arbitrarily small ε is ε–achievable for Δ. This quantity depends on the generic distribution P_X of the source. Thus we get a function

$$R : \mathcal{P}(\mathcal{X}) \times \mathbb{R}_+ \to \mathbb{R}_+,\ \mathcal{P}(\mathcal{X}) = \text{ set of probability distribution (PD's) on } \mathcal{X}, \tag{1.7}$$

of two variables, which was introduced by Shannon [1] and is called rate–distortion function. Using time–sharing one readily verifies with the forgoing definitions that $R(P, \Delta)$ is convex in Δ for every $P \in \mathcal{P}(\mathcal{X})$. A remarkable characterization of R is due to Shannon [1].

Rate–Distortion Theorem. *Let $(X, \hat{X})$ denote a pair of RV's with values in $\mathcal{X} \times \hat{\mathcal{X}}$ and $I(X \wedge \hat{X})$ their mutual information. Then*

$$R(P, \Delta) = \min_{(X,\hat{X}) : P_X = P, \mathbb{E}d(X,\hat{X}) \le \Delta} I(X, \hat{X}) \textit{ for every } \Delta \in \mathbb{R}_+ \textit{ and } P \in \mathcal{P}(\mathcal{X}).$$

This formula has been used to derive several analytical properties of R, mainly as function of Δ. Whereas those properties can be found in many textbooks, the properties of R as function of P are not as well understood.

An exception seems to be the following basic result of Gallager (Theorem 9.5.1 and its extension on pages 467–469 in [2]), which he derives from Shannon's formula. For its formulation we need some notation.

Let the elements in $\mathcal{X}$ be labelled such that for fixed distribution Q

$$Q(0) \ge Q(1) \ge \cdots \ge Q(\alpha - 1), \sum_{i=0}^{\alpha-1} Q(i) = 1. \tag{1.8}$$

For the fixed distortion level Δ let m be the *smallest* non–negative integer with

$$m\ Q(m) + \sum_{k=m+1}^{\alpha-1} Q(K) \le \Delta \le (m-1)Q(m-1) + \sum_{k=m}^{\alpha-1} Q(k). \tag{1.9}$$

The conventional understanding of (1.9) is for $m = 0$

$$1 - Q(0) \le \Delta \le 1 \tag{1.10}$$

and for $m = \alpha$

$$0 \le \Delta \le (\alpha - 1)Q(\alpha - 1). \tag{1.11}$$