Finite-state Rate-Distortion for Individual Sequences

Dharmendra S. Modha
IBM Almaden Research Center
650 Harry Road
San Jose, CA 95120
e-mail: dmodha@us.ibm.com

Daniela Pucci de Farias
Department of Mechanical Engineering
MIT
Cambridge, MA 02139
e-mail: pucci@mit.edu

Abstract — **We introduce a class of lossy finite-state machines for lossy compression of an individual sequence drawn from a finite alphabet at a fixed distortion, and define a fundamental quantity *finite-state rate-distortion* that is an asymptotically attainable lower bound on the compression rate of any lossy finite-state machine. For Hamming distortion, we obtain a universal lower bound on the finite-state rate-distortion of any individual sequence.**

I. Introduction

How compressible is a given individual sequence? Since Shannon's entropy is defined with respect to the underlying distribution of a stationary random process, it cannot be used to define the compressibility of a fixed given sequence. Lempel and Ziv [1] studied the compressibility of a given individual sequence with respect to a class of generalized finite-state information-lossless encoders. For a sequence x, they defined a quantity $\rho(x)$ termed its *finite-state compressibility* and demonstrated that it is an asymptotical lower bound on the compression rate of any finite-state information-lossless encoder.

How compressible is a given individual sequence when a fixed per-letter distortion is allowed? At least two such (closely related) definitions for lossy compressibility of an individual sequence have been given by Ziv [2] and Yang and Kieffer [3]. Let D denote the allowed per-letter distortion with respect to some distortion measure d. The essence of these definitions is to consider all sequences that are in the D-ball of the source sequence, and define the rate-distortion function for the individual source sequence as the infimum of some notion of lossless complexity, for example, $\rho(\cdot)$, over the D-ball.

We offer another definition of rate-distortion function for an individual source sequence with respect to a class of finite-state lossy machines, namely, *finite-state rate-distortion.* In essence, we do not consider all sequences that are in the D-ball of the source sequence, but only those that are in D-ball of the source sequence <u>and</u> can be obtained via a finite-state encoder acting on the source sequence. Hence, our definition may exclude some sequences from the D-ball of the source sequence. Nonetheless, finite-state rate-distortion is identical to rate-distortion for individual sequences in [2, 3]. Our definition provides a deeper understanding of the lossy encoding process that may be useful in constructing practical encoders that asymptotically attain this lower bound. As a key advantage, for Hamming distortion, we establish a universal lower bound on finite-state rate-distortion of an individual sequence. If the individual sequence is drawn from a stationary, ergodic source, then our lower bound coincides with the classical Shannon lower bound.

II. Definitions

Let B and $\hat{B}$ denote finite *source* and *reproduction* alphabets. Let $d : B \times \hat{B} \to [0, \infty)$ denote a bounded, non-negative distortion function that satisfies $\max_{b\in B} \min_{\hat{b}\in\hat{B}} d(b, \hat{b}) = 0$. For $x \in B^\infty$ and $y \in \hat{B}^\infty$, write $d_\infty(x, y) = \limsup_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} d(x_i, y_i)$.

III. Statement of Main Results

Let $\mathbb{M}(s, B)$ denote the set of encoders with at most s states that map sequences in B^∞ to $\{0, 1\}^\infty$. Let $\mathbb{M}^i(s, B) \subset \mathbb{M}(s, B)$ denote the set of *information lossless* machines. Given an individual sequence $x \in B^\infty$, following [1], define *finite-state compressibility* of x as

$$\rho(x, B) = \lim_{s\to\infty} \limsup_{n\to\infty} \min_{\mu\in\mathbb{M}^i(s,B)} \frac{L(x_1^n, \mu)}{n},$$

where $L(x_1^n, \mu)$ is the number of bits output by the machine μ upon seeing the first n bits of the input sequence.

Fix a distortion D. For $s \in \mathbb{N}$, we define a class of *D-lossy machines*, say, $\hat{\mathbb{M}}(s, D, B, \hat{B})$, that have at most s states and map sequences in B^∞ to sequences in $\hat{B}^\infty$ such that for every encoder $\hat{E} \in \hat{\mathbb{M}}(s, D, B, \hat{B})$

$$\sup_{x\in B^\infty} d_\infty(x, \hat{E}x) \le D.$$

We refer to $\hat{E}x$ as the *lossy sequence.*

Definition 1 *For a fixed source sequence $x \in B^\infty$, define the* finite-state rate-distortion *as:*

$$\pi(x, D) = \lim_{s\to\infty} \min_{\hat{E}\in\hat{\mathbb{M}}(s,D,B,\hat{B})} \rho(\hat{E}x, \hat{B}).$$

Let $R(D|x) \equiv \inf\{\rho(y, \hat{B}) : y \in \hat{B}^\infty, d_\infty(x, y) \le D\}$ be as in [2, 3] where it is shown that $R(D|x)$ is an asymptotically attainable lower bound.

Lemma 1 *For every individual sequence x, $\pi(x, D) = R(D|x)$, and, if x is drawn from a stationary, ergodic process then $\pi(x, D) = R(D)$ almost surely.*

Theorem 1 *If $B = \hat{B}$ and d is Hamming distortion, then*

$$\pi(x, D) \ge \rho(x, B) - h_{|B|}(D),$$

where $h_\alpha(\delta) = -\delta \log \delta - (1 - \delta) \log(1 - \delta) + \delta \log(\alpha - 1)$.

References

[1] J. Ziv and A. Lempel, "Compression of individual sequences via variable-rate coding," *IEEE Trans. Inform. Theory*, vol. 24, no. 5, pp. 530–536, 1978.

[2] J. Ziv, "Distortion-rate theory for individual sequences," *IEEE Trans. Inform. Theory*, vol. 26, no. 2, pp. 137–143, 1980.

[3] E.-H. Yang and J. C. Kieffer, "Simple universal lossy data compression schemes derived from the Lempel-Ziv algorithm," *IEEE Trans. Inform. Theory*, vol. 42, no. 1, pp. 239–245, 1996.