Variable-Rate Source Coding Theorems for Stationary Nonergodic Sources

M. Effros,[†] P.A. Chou,[‡] and R. M. Gray [†1]

[†]Information Systems Laboratory, Stanford, CA 94305-4055 USA
[‡]Xerox Palo Alto Research Center, 3333 Coyote Road, Palo Alto, CA 94304 USA

Abstract — **The source coding theorem and its converse imply that the optimal performance theoretically achievable by a fixed- or variable-rate block quantizer on a stationary ergodic source equals the distortion-rate function. While a fixed-rate block code cannot achieve arbitrarily closely the distortion-rate function on an arbitrary stationary nonergodic source, we show for the case of Polish alphabets that a variable-rate block code can. We also show that the distortion-rate function of a stationary nonergodic source has a decomposition as the average over points of equal slope on the distortion-rate functions of the source's stationary ergodic components. These results extend earlier finite alphabet results.**

I. Introduction

In [1], Shields et al. show that for any stationary nonergodic finite alphabet source, the distortion-rate function $D(R)$ equals the infimum of the average of the distortion-rate functions of the source's stationary ergodic components, where the average is taken over points on the component distortion-rate functions whose rates average to at most R. The achievability of this bound by variable-rate block codes is shown in [2].

We extend these variable rate quantization results from finite alphabets to complete separable metric spaces, or Polish alphabets. We employ a simplified variable-rate and variable-distortion using a Lagrangian formulation.

II. Results

Let $(A^\infty, \mathcal{B}^\infty, \mu, T)$ be a stationary dynamical system with Polish alphabet A. That is, let A be a complete separable metric space, let $\mathcal{B}$ be the Borel σ-algebra generated by the open sets of A, let A^∞ be the set of one-sided sequences $x = (x_1, x_2, \ldots)$ from A, let $\mathcal{B}^\infty$ be the σ-algebra of subsets of A^∞ generated by finite-dimensional rectangles with components in $\mathcal{B}$, let T be the left shift operator on A^∞, and let μ be a measure on the measurable space $(A^\infty, \mathcal{B}^\infty)$, stationary with respect to T.

Now let $\rho(x_1, y_1) < \infty$ be a real-valued nonnegative distortion measure for $x_1 \in A$, $y_1 \in \hat{A}$, where $\hat{A}$ is an abstract reproduction alphabet. Assume that $\rho(x_1, y_1)$ is continuous in x_1 for each $y_1 \in \hat{A}$ and that there exists a reference letter y_1^* such that $E_\mu \rho(X_1, y_1^*) < \infty$. Define $\rho(x^N, y^N) = \sum_{i=1}^N \rho(x_i, y_i)$.

Finally let Q be a variable-rate block quantizer with blocklength N. That is, let Q be a map from A^N onto some finite or countable set of codewords $\{y^N\} \subseteq \hat{A}^N$ composing a codebook $\mathcal{C} = \{(y^N, |y^N|)\}$ in which each codeword y^N has an associated variable-length binary description, with length denoted $|y^N|$. The description lengths must satisfy the Kraft inequality $\sum_{y^N \in \mathcal{C}} 2^{-|y^N|} \leq 1$.

[1]This material is based upon work partially supported by an AT&T Ph.D. Scholarship, by a grant from the Center for Telecommunications at Stanford, and by an NSF Graduate Fellowship.

The optimal performance theoretically achievable by any variable-rate block quantizer is the operational distortion-rate function $\delta^{\text{vr}}(R, \mu) = \inf_N \delta_N^{\text{vr}}(R, \mu)$, where $\delta_N^{\text{vr}}(R, \mu)$ is the Nth order operational distortion-rate function

$$\delta_N^{\text{vr}}(R, \mu) = \inf_Q \left\{ \frac{1}{N} E_\mu \rho(X^N, Q(X^N)) : \frac{1}{N} E_\mu |Q(X^N)| \leq R \right\}.$$

Here, the infimum is taken over all variable-rate block quantizers Q with blocklength N. We contrast this with the optimal performance theoretically achievable by fixed-rate block quantizers, $\delta^{\text{fr}}(R, \mu) = \inf_N \delta_N^{\text{fr}}(R, \mu)$, in which $\delta_N^{\text{fr}}(R, \mu)$ is defined as $\delta_N^{\text{vr}}(R, \mu)$ but with the infimum taken over all fixed-rate block quantizers with blocklength N.

The Shannon distortion-rate function is defined similarly, as $D(R, \mu) = \inf_N D_N(R, \mu)$, where $D_N(R, \mu)$ is the Nth order distortion-rate function

$$D_N(R, \mu) = \inf_\nu \left\{ \frac{1}{N} E_{\mu\nu} \rho(X^N, Y^N) : \frac{1}{N} I_{\mu\nu}(X^N; Y^N) \leq R \right\}.$$

Here, ν is a conditional probability or test channel from A^N to $\hat{A}^N$ defining, with μ, a joint probability or hookup $\mu\nu$ on X^N and Y^N, and I is the mutual information.

It is well-known that both $D(R, \mu)$ and $\delta^{\text{fr}}(R, \mu)$ are convex in R [3]; $\delta^{\text{vr}}(R, \mu)$ is likewise convex in R, by a timesharing argument. Hence $\delta^{\text{vr}}(R, \mu)$ and $D(R, \mu)$ can be characterized by their support functionals [4, p. 135] the weighted operational distortion-rate function $\ell(\lambda, \mu) = \inf_R \left[\delta^{\text{vr}}(R, \mu) + \lambda R\right]$ and the weighted Shannon distortion-rate function $L(\lambda, \mu) = \inf_R [D(R, \mu) + \lambda R]$.

The source coding theorem and its converse imply that when μ is ergodic, $\delta^{\text{vr}}(R, \mu) = \delta^{\text{fr}}(R, \mu) = D(R, \mu)$ for all $R \geq 0$ (and hence $\ell(\lambda, \mu) = L(\lambda, \mu)$ for all $\lambda \geq 0$). When μ is nonergodic, let $\{\mu_x : x \in A^\infty\}$ denote the ergodic decomposition of μ. The ergodic decomposition exists since A Polish implies $(A, \mathcal{B})$ standard [5, Theorem 3.3.1], and hence $(A^\infty, \mathcal{B}^\infty)$ is standard [5, Lemma 2.4.1] which gives the desired property by [5, Theorem 7.4.1]. The main results of this paper are that under the conditions given above

Theorem 1 $\ell(\lambda, \mu) = \int \ell(\lambda, \mu_x) d\mu(x)$ $\forall \lambda \geq 0$,

Theorem 2 $L(\lambda, \mu) = \int L(\lambda, \mu_x) d\mu(x)$ $\forall \lambda \geq 0$,

Theorem 3 $\ell(\lambda, \mu) = L(\lambda, \mu)$ $\forall \lambda \geq 0$, and hence $\delta^{\text{vr}}(R, \mu) = D(R, \mu)$ $\forall R \geq 0$.

References

[1] P. C. Shields, D. L. Neuhoff, L. D. Davisson, and F. Ledrappier, "The distortion-rate function for nonergodic sources," *The Annals of Probability*, 6(1):138–143, 1978.

[2] A. Leon-Garcia, L. D. Davisson, and D. L. Neuhoff, "New results on coding of stationary nonergodic sources," *IEEE Transactions on Information Theory*, 25(2):137–144, March 1979.

[3] R. M. Gray, *Entropy and Information Theory*, Springer-Verlag, New York, 1990.

[4] D. G. Luenberger, *Optimization by Vector Space Methods*, John Wiley and Sons, New York, 1969.

[5] R. M. Gray, *Probability, Random Processes, and Ergodic Properties*, Springer-Verlag, New York, 1988.