Rates of Convergence in Adaptive Universal Vector Quantization

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 — **We consider the problem of adaptive universal quantization. By *adaptive* quantization we mean quantization for which the delay associated with encoding the jth sample in a sequence of length n is bounded for all $n > j$. We here demonstrate the existence of an adaptive universal quantization algorithm for which any weighted sum of the rate and the expected mean square error converges almost surely and in expectation as $O(\sqrt{\log\log n/\log n})$ to the corresponding weighted sum of the rate and the distortion-rate function at that rate.**

I. Introduction

A data compression technique that asymptotically (in the length of the data sequence) achieves the rate-distortion bound on all sources within some class is called "universal." The two basic approaches to universal compression are known as two pass, off-line, or batch codes, and one pass, on-line, incremental, or adaptive codes. Batch algorithms require two passes through the data: the first to choose a code and the second to describe the data using the chosen code. Provably universal two pass noiseless codes and quantizers have been studied in Davisson[1], Ziv[2], Linder et al.[3], and Chou et al.[4]. Adaptive algorithms traverse the data only once, simultaneously developing and using a model for the data. While codes such as the Ziv-Lempel technique have solved the adaptive universal noiseless coding problem (albeit at a slower than optimal rate of convergence), attempts to generalize these algorithms to provably universal adaptive quantization techniques have until recently been unsuccessful. In [5], Zhang and Wei present an adaptive universal quantization technique for independent, identically distributed (iid) samples from finite, discrete alphabet sources. Assuming access to a true random number generator, Zhang and Wei examine their algorithm's asymptotic properties using a nontraditional measure of rate of convergence. We here consider an adaptive universal quantization algorithm for stationary, real-valued sources. The algorithm, introduced in [6], does not require a random number generator. We examine the asymptotic behavior of the proposed algorithm for iid, real-valued sources of bounded support and compare the results to those for batch quantizers as discussed in [3].

II. The Algorithm

We here consider the use of fixed rate codebooks of constant dimension k. Sequences of variable rate codebooks with increasing dimension will be considered in a later work.

Given the bound $\pm B$ on the distribution's support, the algorithm proceeds as follows.

1. Encode vector $X(1)$ using an initial codebook. Set $j = 1$.
2. Update the codebook using the vector(s) just coded.
3. Quantize each codevector component to width $B/2^{j-1}$ bins.
4. Describe the new codebook to the decoder.
5. Encode vectors $X(2^{2(j-1)}+1), \ldots, X(2^{2j})$.
6. Increment j by 1 and go to 2.

As the number of training samples increases, the quality of the resulting codebook improves and the accuracy to which we quantize its codewords increases. Note that the delay in encoding and decoding is simply the vector dimension k.

III. Universality and Rates of Convergence

Let $\{X_i\}$ be independent samples from a stationary random process in $\mathbb{R}^k$ with process measure P_θ, $\theta \in \Lambda$, and let $D^A(X^n)$ and $R^A(X^n)$ be the per symbol distortion and rate achieved by the above described adaptive algorithm on a sequence $\{X(1), \ldots, X(n)\}$. Define the per letter distortion and rate redundancies for the given algorithm as $\Delta_D(A,\theta) = D^A(X^n) - D_\theta(R)$ and $\Delta_R(A,\theta) = R^A(X^n) - R$ respectively, where $D_\theta(R)$ is the Shannon distortion-rate function associated with the source θ. Using E_θ to represent the expectation with respect to P_θ, and noticing that the dimension k must grow with n in order to achieve asymptotic optimality, the main results of this paper are as follows.

Theorem 1 *For memoryless, real-valued sources of bounded support, there exists an adaptive universal vector quantizer such that for all $\lambda > 0$,*

$$\Delta_D(A,\theta) + \lambda\Delta_R(A,\theta) = O\left(\sqrt{\frac{\log\log n}{\log n}}\right)$$

almost surely and in expectation.

This theorem is derived using Linder et al.'s [3] Vapnik-Chervonenkis bound on the probability that the expected distortion of a codebook designed for a sample of m vectors will differ greatly from the expected distortion of the optimal codebook of the same size and dimension. This is identical to the result for two pass algorithms in [3].

References

[1] L. D. Davisson, "Universal noiseless coding," *IEEE Transactions on Information Theory*, 19(6):783–795, November 1973.

[2] J. Ziv, "Coding of sources with unknown statistics—Part II: Distortion relative to a fidelity criterion," *IEEE Transactions on Information Theory*, 18(3):389–394, May 1972.

[3] T. Linder, G. Lugosi, and K. Zeger, "Rates of covergence in the source coding theorem, in empirical quantizer design, and in universal lossy source coding," *IEEE Transactions on Information Theory*, 1993. To appear.

[4] P.A. Chou, M. Effros, and R.M. Gray, "A vector quantization approach to universal noiseless coding and quantization," In preparation.

[5] Z. Zhang and V. K. Wei, "An on-line universal lossy data compression algorithm via continuous codebook refinement," May 1993. Preprint.

[6] M. Effros, P. A. Chou, and R. M. Gray, "Adaptive almost sure universal quantization," In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, Adelaide, Australia, April 1994.

[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.

0 - 7803-2015-8/94/$4.00 ©1994 IEEE