Fixed Rate Universal Lossy Source Coding for Memoryless Sources and Rates of Convergence

Tamás Linder[1] Gábor Lugosi[2] Kenneth Zeger[3]

[1]Dept. of Telecommunications, Technical University of Budapest, 1521 Stoczek u. 2, Budapest, Hungary.
[2]Dept. of Mathematics, Faculty of Elect. Eng., Technical University of Budapest, 1521 Stoczek u. 2, Budapest, Hungary.
[3]Coordinated Science Laboratory, Dept. of Elect. and Comp. Engineering, University of Illinois, Urbana, IL 61801

Abstract — **A fixed rate universal lossy source coding scheme is introduced for i.i.d. sources. It is shown that as the sample size n grows the per letter distortion obtained using this universal quantizer converges to Shannon's distortion-rate function $D(R)$ in expectation at a rate $O(\log n/n)$ for finite alphabet sources and $O(\sqrt{\log n/n})$ both in expectation and almost surely for bounded real valued sources.**

I. Introduction

Much is known about lossless universal source coding, while much less is known about its lossy counterpart in terms of performance bounds. Fixed rate universal lossy source coders for abstract alphabets were first shown to exist by Ziv [1] and later pursued by other researchers. Nearly all of the known results have been existence proofs or algorithm descriptions. For the lossless case and for i.i.d. or finite-memory sources exact lower bounds are known of the form $c\log n/n$ for the rate redundancy, and these bounds are achievable (see e.g. [2] or [3]). Very little is known, however, about how fast the average distortion converges to $D(R)$.

The only known results on rates of convergence for lossy universal codes appear in [4] and [5]. In [4], Yu and Speed give a variable-rate universal lossy coding scheme for finite alphabet i.i.d. sources that has a convergence rate of $O(\log n/n)$ for the expected codelength provided certain technical conditions are satisfied. In [5], Ziv's fixed-rate scheme, universal over the class of infinite alphabet stationary sources, is shown to converge almost surely at a rate of $O(\sqrt{\log\log n/\log n})$ on i.i.d. inputs.

In the present paper we introduce a new fixed-rate lossy source coding algorithm that is universal over the the class of i.i.d. sources. The average distortion in our scheme is shown to converge to $D(R)$ at a rate of $O(\log n/n)$ for finite alphabet sources (in expectation) and $O(\sqrt{\log n/n})$ for real-valued bounded sources (both in expectation and almost surely). For the infinite alphabet case the distortion criterion is in the mean square sense, while in the finite alphabet case the distortion function is not constrained. By restricting the class of sources we are able to obtain a much faster rate of convergence than in [5]. The main advantages of our new scheme is that it is fixed-rate and has a fast rate of convergence for both finite and infinite alphabet sources.

II. Results

Let $\mathcal{A}$ and $\mathcal{B}$ be the finite source and reproduction alphabets respectively, and suppose that a single letter distortion measure $d:\mathcal{A}\times\mathcal{B}\to[0,d_{\max}]$ is given.

[1]Supported in part by NSF Grant No. NCR-92-96231.

Theorem 1 *For all $R>0$ there exists a sequence of codes $\{g_n\}_{n=1}^{\infty}$ of fixed rate at most R, such that for any i.i.d source $\{X_n\}$ on $\mathcal{A}$ with $D(R)>0$,*

$$\frac{1}{n}\mathbf{E}d(X_1^n, g_n(X_1^n)) \le D(R)+|D'(R)|(|\mathcal{A}|+1/2+o(1))\left(\frac{\log n}{n}\right)$$

Theorem 2 *For any $R>0$ there exists a sequence of codes $\{g_n\}_{n=1}^{\infty}$ of fixed rate at most R, such that for any i.i.d. real source $\{X_n\}$ of bounded support with $D(R)>0$,*

$$\frac{1}{n}\mathbf{E}\|X_1^n - g_n(X_1^n)\|^2 \le D(R)+c\sqrt{\frac{\log n}{n}}$$

for some c depending on R and the source distribution. For a given rate R, the constant c is uniformly upper bounded for the class of i.i.d. sources of the same bounded support.

In fact, more is true. The following theorem says that the $O(\sqrt{\log n/n})$ rate of convergence also holds for the sample distortion with probability one.

Theorem 3 *For any $R>0$ there exists a sequence of codes $\{g_n\}_{n=1}^{\infty}$ of fixed rate at most R, such that for any i.i.d. real source $\{X_n\}$ of bounded support with $D(R)>0$,*

$$\frac{1}{n}\|X_1^n - g(X_1^n)\|^2 - D(R) = O\left(\sqrt{\frac{\log n}{n}}\right)$$

almost surely.

References

[1] J. Ziv, "Coding of sources with unknown statistics- Part II: Distortion relative to a fidelity criterion," *IEEE Trans. Inform. Theory,* vol. IT-18, pp. 389-394, May 1972.

[2] R. E. Krichevsky and V. K. Trofimov, "The performance of universal encoding." *IEEE Trans. Inform. Theory,* vol. IT-27, pp. 199-207 , Mar. 1981.

[3] J. Rissanen, "Stochastic complexity and modeling," *Annals of Statistics,* vol. 14, pp. 1080-1100, 1986.

[4] B. Yu and T. P. Speed, "A rate of convergence result for a universal D-semifaithful code," *IEEE Trans. Inform. Theory,* vol. IT-39, pp. 813-820 , May 1993.

[5] T. Linder, G. Lugosi, and K. Zeger, "Rates of Convergence in the Source Coding Theorem, in Empirical Quantizer Design, and in Universal Lossy Source Coding," *IEEE Trans. Inform. Theory,* (to appear).

[6] R. Pilc, "The transmission distortion of a source as a function of the encoding codelength," *Bell System Technical Journal* vol. 47, pp. 827-885, 1968.

0-7803-2015-8/94/$4.00 ©1994 IEEE