When Optimal Entropy-Constrained Quantizers have only a Finite Number of Codewords

Philip A. Chou
Microsoft Corporation
One Microsoft Way
Redmond, WA 98052-6399
Email pachou@microsoft.com

Bradley J. Betts
Information Systems Laboratory
Stanford University
Stanford, CA 94305-9510
Email betts@isl.stanford.edu

Abstract — **An entropy-constrained quantizer Q is optimal if it minimizes the expected distortion $Ed(X, Q(X))$ subject to a constraint on the output entropy $H(Q(X))$. In general, such an optimal entropy-constrained quantizer may have a countably infinite number of codewords. In this short paper, we show that if the tails of the distribution of X are sufficiently light (with respect to the distortion measure), then the optimal entropy-constrained quantizer has only a finite number of codewords. In particular, for the squared error distortion measure, if the tails of the distribution of X are lighter than the tails of a Gaussian distribution, then the optimal entropy-constrained quantizer has only a finite number of codewords.**

I. Introduction

It has long been known that entropy-coded lattice quantizers are nearly optimal in the following sense: they can achieve entropies that are within a fraction of a bit of the minimum possible entropy among all quantizers with a given expected distortion [1]. This is convenient for practical implementations, which can take advantage of fast lattice quantization algorithms, but it raises the spectre of having to estimate and store a probability distribution over a countably infinite set of codewords. Practical algorithms invariably restrict the codewords of the lattice quantizer to a bounded "granular" region, with little apparent loss.

In related work [2] designing locally optimal entropy-constrained vector quantizers (ECVQs) from training data, it has been repeatedly observed that the number of codewords in a locally optimal ECVQ is bounded by a number that depends on the source and the target entropy. That is, the number of codewords does not increase even if the ECVQ design algorithm is initialized with a greater number of codewords, or if a greater number of training vectors is made available. In some sense, there is a natural number of codewords for a given source at a given rate.

The above experiences suggest that optimal entropy-constrained quantizers may not necessarily have an infinite number of codewords. While this is clear for sources with bounded support, the question under consideration in this paper is whether or not optimal entropy-constrained quantizers always have an infinite number of codewords when the source has an unbounded region of support.

II. Results

Let X be a real random vector, and let $d(x,y) = ||x-y||^r$ be the rth power distortion measure, where $||\cdot||$ is the usual Euclidean norm, and $r > 0$. Let Q be a quantizer whose expected distortion $D(Q) = Ed(X, Q(X))$ and output entropy $R(Q) = H(Q(X))$ achieves a point on the lower convex hull of the operational distortion-rate function $\hat{D}(R) = \inf_Q\{D(Q) : R(Q) \leq R\}$. Q is optimal in the sense that it has the lowest possible distortion among all quantizers with the same or lower entropy. Furthermore, since it lies on the convex hull, there exists a positive Lagrange multiplier $\lambda > 0$ such that Q minimizes $D(Q) + \lambda R(Q)$.

Theorem. Suppose there exists a Q that minimizes $Ed(X, Q(X)) + \lambda H(Q(X))$ for some $\lambda > 0$. If for some $\epsilon > 0$ $P\{||X|| \geq t\} = o\left(2^{-(1+\epsilon)t^r/\lambda}\right)$, i.e., $P\{||X|| \geq t\}$ goes to zero faster than $2^{-(1+\epsilon)t^r/\lambda}$, then Q has only a finite number of codewords.

Proof (Idea). Let y_i be the codewords of Q, and let B_i be the decision regions of Q, such that $Q(x) = y_i$ whenever $x \in B_i$. Also, let $\alpha(x) = i$ whenever $x \in B_i$. Let $p(i)$ be the probability that $\alpha(X) = i$ or equivalently that $X \in B_i$. It is known that a necessary condition for Q to minimize $D(Q) + \lambda R(Q) = E[d(X, Q(X)) - \lambda \log p(\alpha(X))]$, is that for almost all x, $\alpha(x) = \arg\min_i d(x, y_i) - \lambda \log p(i)$. This is the tool that enables us to show that if the tail of the distribution of X is sufficiently light, then Q has only a finite number of codewords. The basic idea is that if the tail is light, then as would-be codewords y_i and decision regions B_i move outwards from the origin, for any $x \in B_i$, the penalized distortion $d(x, y_i) - \lambda \log p(i)$ of assigning x to y_i would grow larger at a rate faster than the penalized distortion $d(x, y_0) - \lambda \log p(0)$ of assigning x to some existing codeword y_0 near the origin. Hence any such $x \in B_i$ would be assigned instead to y_0, which would be a contradiction. □

As an application of the Theorem, consider a Gaussian source with mean 0 and variance σ^2. According to Feller [3], $P\{||X|| > t\} \leq e^{-t^2/2\sigma^2}$, which is $o(2^{-(1+\epsilon)t^2/\lambda})$ whenever $\lambda \geq 2(1+\epsilon)\sigma^2 \ln 2$. Thus at least for large λ (low rates), the optimal entropy-constrained quantizers, with respect to the squared error distortion measure, have only a finite number of codewords.

It is tempting to conjecture that for the squared error distortion measure, the Gaussian distribution is a breakpoint. We have shown that optimal entropy-constrained quantizers for distributions with tails lighter than a Gaussian (including distributions with finite support) have only a finite number of codewords. However, whether or not the converse holds is still an open problem.

References

[1] A. Gersho. Asymptotically optimal block quantization. *IEEE Trans. Information Theory*, 25(4):373–380, July 1979.

[2] P.A. Chou, T. Lookabaugh, and R.M. Gray. Entropy-constrained vector quantization. *IEEE Trans. Acoustics, Speech, and Signal Processing*, 37(1):31–42, January 1989.

[3] W. Feller. *An Introduction to Probability Theory and Its Applications*, volume 1. John Wiley & Sons, third edition, 1950.

