A LAGRANGIAN FORMULATION OF HIGH RATE QUANTIZATION

Joyce Shih, Anuradha K. Aiyer and Robert M. Gray

Information Systems Laboratory, Department of Electrical Engineering, Stanford, CA 94305
{joyces,anuaiyer,rmgray}@stanford.edu

ABSTRACT

The asymptotic optimal performance of variable-rate vector quantizers of fixed dimension and large rate was first developed in a rigorous fashion by Paul Zador. Subsequent design algorithms for such compression codes used a Lagrangian formulation in order to generalize Lloyd's classic quantizer optimization algorithm to variable rate codes. This formulation has been subsequently adopted in a variety of practical systems including rate-optimized streaming video. We describe a Lagrangian formulation of Zador's variable-rate quantization results and apply it to estimate Zador's constant using the generalized Lloyd algorithm.

1. INTRODUCTION

There are two primary approaches to the theoretical analysis of systems for analog-to-digital conversion and signal compression: Shannon's rate-distortion theory and Zador's high-rate quantization theory. Shannon's theory provides an information theoretic approach characterizing the optimal achievable performance of systems with fixed rate and large dimension or delay; Zador's high-rate quantization theory characterizes the optimal achievable performance of systems with fixed dimension and large rate. An account of the history and status of the two approaches (as of 1998) along with numerous references may be found in [1]. Zador's approach arguably makes more sense when the delay or dimension is small or moderate, but the allowed bit-rate is not, e.g., in systems requiring small delay and high quality. Our focus here is on variable-rate systems as they provide better distortion/bit-rate tradeoffs than fixed-rate systems.

A shortcoming of Zador's original formulation of the optimal performance of high-rate entropy-constrained vector quantizers [2] is that it does not directly lend itself to the comparison with codes designed or optimized using the generalized Lloyd algorithm for entropy-constrained quantizers, which uses a Lagrangian formulation of the quantization problem. Our goal is to provide a new statement of Zador's result using the Lagrangian formulation and to take advantage of the formulation by applying the Lloyd clustering algorithm to provide numerical estimates of Zador's constant. Preliminary results are presented. The theory may prove useful, for example, in deriving theoretical performance bounds and performance approximations for rate-controlled coding in the high-rate regime.

This work was partially supported by the Stanford University Research Experience for Undergraduates (REU) Program and by the National Science Foundation under grants NSF:MIP-9706284-001 and CCR-0073050.

2. VECTOR QUANTIZATION

A k-dimensional vector quantizer q is characterized by

- an encoder $\alpha : \Re^k \to \mathcal{I}$ mapping k-dimensional Euclidean space into the integers $\mathcal{I}$
- a reproduction decoder $\beta : \mathcal{I} \to \Re^k$ mapping indices into reproduction vectors
- an index coder $\psi : \mathcal{I} \to \{0,1\}^*$ mapping indices into variable-length binary sequences. ψ is assumed to be invertible and uniquely decodable.
- the overall quantization operation $q(x) = \beta(\alpha(x))$

The encoder is described by a partition $\mathcal{S} = \{S_i\}$ with $S_i = \{x : \alpha(x) = i\}$. The decoder is described by the reproduction codebook $\mathcal{C} = \{\beta(i);\ i \in \mathcal{I}\}$.

To measure performance we assume a distortion measure $d(x,y) \geq 0$, which for the sake of simplicity we assume to be the mean-squared error, $d(x,\beta(i)) = ||x - y_i||^2 = \sum_{l=0}^{k-1} |x_l - \beta(i)_l|^2$. Many asymptotic quantization results extend to weighted quadratic measures of the form $d(x,\beta(i)) = (x-\beta(i))^t B_x (x-\beta(i))$ for positive definite weighting matrices B_x and to other norms (see, e.g., [1]). The instantaneous rate of an index is given by $\ell(\psi(i))$, the length of the binary vector $\psi(i)$. If x is a realization of a random vector X described by a probability density function (pdf) f, then the average distortion and average rate of a quantizer are given by

$$\begin{aligned} D_f(q) &= \sum_i \int_{S_i} dx\, f(x) ||x - y_i||^2 \\ R_f(q) &= \sum_i P_f(S_i) \ell(i) \end{aligned}$$

The classic approach to describing optimal performance is the distortion-rate approach (the dual of Shannon's rate-distortion approach). For $R > 0$, define the operational distortion-rate function as $\delta_f(R) = \inf_{q:R_f(q)\leq R} D_f(q)$. Zador proved that under certain technical conditions on the pdf f,

$$\lim_{R\to\infty} 2^{\frac{2}{k}R}\delta_f(R) = b_{2,k}2^{\frac{2}{k}h(f)} \tag{1}$$

where $b_{2,k}$ is Zador's constant, which depends only on k and not f, and $h(f) \stackrel{\Delta}{=} -\int dx\, f(x)\log f(x)$ is the differential entropy. When entropies appear as an exponent of 2, they are assumed to have units of bits. Otherwise their units are nats. The constant $b_{2,k}$ is known only for $k = 1$ and upper and lower bounds are known for general k that converge as $k \to \infty$. If a pdf f is such that (1) holds, we say it has the *traditional Zador property*.

The optimality properties and code improvement algorithm for variable rate codes are better stated in terms of a Lagrangian formulation of the problem. Toward this end, define for $\lambda > 0$ a Lagrangian distortion

$$\rho_\lambda(x, i) = d(x, y_i) + \lambda r(i)$$

and average distortion

$$\begin{aligned}\rho_(f, \lambda, q) &= E_f\left(d(X, q(X)) + \lambda E_f r(\alpha(X))\right)\\ &= D_f(q) + \lambda R_f(q)\end{aligned}$$

and optimal performance

$$\rho(f, \lambda) = \inf_q \rho(f, \lambda, q).$$

Intuitively, each λ yields a distortion-rate pair on the operational distortion-rate function curve. Small λ means high rate, large λ means small rate. Thus one would expect that the high rate traditional Zador formulation should translate into a small λ result.

The Lagrangian formulation yields generalized Lloyd conditions for optimality (i.e., necessary conditions for optimal codes) which can be successively applied to yield a sequence of codes with diminishing Lagrangian distortion. These are (see, e.g., [1]):

Centroid condition: The optimal decoder β for given encoder α and index coder ψ, is determined by

$$\beta(i) = \operatorname*{argmin}_y E_f[d(X, y)|\alpha(X) = i].$$

For the squared error, this is $\beta(i) = E_f[X|\alpha(X) = i]$.

Minimum distortion property: The optimal encoder α, given decoder β and index coder ψ, is determined by $\alpha(x) = \operatorname{argmin}_i\left(d(x, y_i) + \lambda\ell(i)\right)$

Kraft/Huffman property: The optimal index coder ψ, given an encoder α and decoder β, is an optimal lossless code for $\alpha(X)$ or, equivalently, $q(X)$ (e.g., a Huffman code with lengths satisfying the Kraft inequality).

The Kraft/Huffman property ensures that

$$H_f(q(X)) \leq E_f[\ell(\alpha(X))] < H_f(q(X)) + 1$$

where the Shannon entropy is defined as usual as

$$H_f(q(X)) = -\sum_i P_f(S_i)\log P_f(S_i)$$

Hence it is common practice to make the approximation that

$$\begin{aligned}\ell(i) &\approx -\log P_f(\alpha(X) = i)\\ R_f(q) &\approx E_f\ell(\alpha(X)) = H_f(q(X))\end{aligned}$$

in both theory and practice, yielding entropy-constrained vector quantization (ECVQ). This is in fact done in Zador's original derivation and it replaces the Kraft/Huffman property in the Lloyd conditions by a simple computation of index entropy.

3. TRADITIONAL VS. LAGRANGIAN ASYMPTOTIC PERFORMANCE

We say that a pdf f has the *Lagrange-Zador property* if the following limit exists:

$$\lim_{\lambda\to 0}\left(\frac{\rho(f, \lambda)}{\lambda} + \frac{k}{2}\ln\lambda\right) - h(f) = \theta_k \tag{2}$$

where θ_k depends only on the dimension and not on the pdf.

Lemma 1 *A pdf f has the traditional Zador property if and only if it has the Lagrange-Zador property. The constants are related by*

$$\theta_k = \frac{k}{2}\ln\frac{2e}{k}b_{2,k} \tag{3}$$

Proof: We begin with some notation. For the Lagrangian formulation, define

$$\begin{aligned}\theta(f, \lambda, q) &= \frac{\rho(f, \lambda, q)}{\lambda} + \frac{k}{2}\ln\lambda - -h(f)\\ &= \frac{D_f(q)}{\lambda} + H_f(q(X)) - h(f) + \frac{k}{2}\ln\lambda\end{aligned}$$

$$\begin{aligned}\theta(f, \lambda) &= \inf_q \theta(f, \lambda, q)\\ \overline{\theta}(f) &= \limsup_{\lambda\to 0}\theta(f, \lambda)\\ \underline{\theta}(f) &= \liminf_{\lambda\to 0}\theta(f, \lambda)\end{aligned}$$

$$\begin{aligned}
\zeta(f,R,q) &= D_f(q)2^{\frac{2}{k}(R-h(f))} \\
\zeta(f,R) &= \inf_{q:H_f(q(X))\le R} \zeta(f,R,q) \\
\underline{\zeta}(f) &= \liminf_{R\to\infty} \zeta(f,R) \\
\overline{\zeta}(f) &= \limsup_{R\to\infty} \zeta(f,R)
\end{aligned}$$

The traditional form of Zador's property can now be described as $\underline{\zeta}(f) = \overline{\zeta}(f) = b_{2,k}$ and the Lagrangian form as $\underline{\theta}(f) = \overline{\theta}(f) = \theta_k$.

The connections between the limits follow from the equality

$$\begin{aligned}
\theta(f,\lambda,q) &= \frac{k}{2}\left[\frac{2D_f(q)}{k\lambda} - -\ln\frac{2D_f(q)}{k\lambda} - -1\right] \\
&\quad + \frac{k}{2}\ln\left(\frac{2e}{k}D_f(q)2^{\frac{2}{k}(H_f(q(X))-h(f))}\right).
\end{aligned}$$

The term in the square brackets is nonnegative since $\ln r \le r-1$.

If $\underline{\theta}$ is finite, we can choose $\lambda_n \to 0$ as $n\to\infty$ so that $\theta(f,\lambda_n) \to \underline{\theta}(f)$ and hence a sequence of quantizers q_n exists such that $\theta(f,\lambda_n,q_n) \to \underline{\theta}(f)$ and hence $\lambda_n\theta(f,\lambda_n,q_n) \to 0$. Thus

$$D_f(q_n) + \lambda_n(H_f(q_n(X)) - h(f)) + \frac{k}{2}\lambda_n \ln\lambda_n \to 0.$$

Since the rightmost term goes to zero and the differential entropy term goes to 0, this also means that

$$D_f(q_n) \to 0. \tag{4}$$

Define $\lambda_n^* = 2D_f(q_n)/k$ and observe that

$$\begin{aligned}
\theta(f,\lambda_n^*,q_n) &= \frac{k}{2}\ln\left(\frac{2e}{k}D_f(q_n)2^{\frac{2}{k}(H_f(q_n(X))-h(f))}\right) \\
&\le \theta(f,\lambda_n,q_n) \qquad (5)
\end{aligned}$$

.

The divergence inequality can be used to prove that for all f, λ, and q that $\theta(f,\lambda,q) \ge -k\ln\pi$. Since $D_f(q_n) \to 0$, this necessarily implies that $H_f(q_n(X)) \to \infty$. Thus

$$\begin{aligned}
&\underline{\theta}(f) \\
&= \lim_{n\to\infty} \theta(f,\lambda_n,q_n) \\
&\ge \liminf_{n\to\infty} \frac{k}{2}\ln\left(\frac{2e}{k}D_f(q_n)2^{\frac{2}{k}(H_f(q_n(X))-h(f))}\right) \\
&\ge \liminf_{n\to\infty} \frac{k}{2}\ln\left(\frac{2e}{k}\delta_f(H_f(q_n(X)))2^{\frac{2}{k}(H_f(q_n(X))-h(f))}\right) \\
&= \frac{k}{2}\ln\left(\frac{2e}{k}\liminf_{n\to\infty}\delta_f(H_f(q_n(X)))2^{\frac{2}{k}(H_f(q_n(X))-h(f))}\right) \\
&\ge \frac{k}{2}\ln\frac{2e}{k}\underline{\zeta}(f)
\end{aligned}$$

Summarizing,

$$\underline{\theta}(f) \ge \frac{k}{2}\ln\frac{2e}{k}\underline{\zeta}(f). \tag{6}$$

Now suppose that Zador's traditional result holds, hence $\underline{\zeta}(f) = \overline{\zeta}(f) = b_{2,k}$ and for any sequence $R_n \to \infty$ there is a sequence of quantizers q_n with $H_f(q_n(X)) \le R_n$ for which $\zeta(f,R_n,q_n) \to b_{2,k}$ so that

$$D_f(q_n)2^{\frac{2}{k}(R_n-h(f))} \to b_{2,k}.$$

Choose $\lambda_n \to 0$ such that $\theta(f,\lambda_n) \to \overline{\theta}(f)$. For this sequence λ_n, define

$$R_n = h(f) + \frac{k}{2}\log\frac{2b_{2,k}}{k\lambda_n}$$

and construct q_n as above for this R_n. Then

$$\begin{aligned}
&\frac{k}{2}\ln\frac{2e}{k}b_{2,k} \\
&\quad= \lim_{n\to\infty}\frac{k}{2}\ln\frac{2e}{k}D_f(q_n)2^{\frac{2}{k}(R_n-h(f))} \\
&\quad\ge \liminf_{n\to\infty}\left(\frac{k}{2}\ln\frac{2e}{k}D_f(q_n)2^{\frac{2}{k}(H_f(q_n(X))-h(f))}\right) \\
&\quad= \liminf_{n\to\infty}\Bigg(\theta(f,\lambda_n,q_n) - - \\
&\qquad \frac{k}{2}\left[\frac{2D_f(q_n)}{k\lambda_n} - -\ln\frac{2D_f(q_n)}{k\lambda_n} - -1\right]\Bigg)
\end{aligned}$$

Since $\theta(f,\lambda_n,q_n) \ge \theta(f,\lambda_n)$ and the ratios go to 1 (based on the choice of R_n), it follows that

$$\frac{k}{2}\ln\frac{2e}{k}b_{2,k} \ge \liminf_{n\to\infty}\theta(f,\lambda_n) = \overline{\theta}(f),$$

which proves that if the traditional Zador limit holds, then so does the Lagrangian form with

$$\theta_k = \frac{k}{2}\ln\frac{2e}{k}b_{2,k}. \tag{7}$$

Suppose instead that the Lagrangian form of Zador's theorem holds, so that $\underline{\theta}(f) = \overline{\theta}(f) = \theta_k$. Choose $R_n \to \infty$ and q_n so that $\zeta(f,R_n,q_n) \to \underline{\zeta}(f)$ and $H_f(q_n(X)) \le R_n$. Then

$$\begin{aligned}
\zeta(f,R_n,q_n) &= D_f(q_n)2^{\frac{2}{k}(R_n-h(f))} \\
&\ge D_f(q_n)2^{\frac{2}{k}(H_f(q_n(X))-h(f))}
\end{aligned}$$

and hence

$$\begin{aligned}
\frac{k}{2}\ln\frac{2e}{k}\underline{\zeta}(f) &= \lim_{n\to\infty}\frac{k}{2}\ln\frac{2e}{k}\zeta(f,R_n,q_n) \\
&\ge \liminf_{n\to\infty}\frac{k}{2}\ln\frac{2e}{k}D_f(q_n)2^{\frac{2}{k}(H_f(q_n(X))-h(f))} \\
&= \liminf_{n\to\infty}\theta(f,\lambda_n^*,q_n) \\
&\ge \liminf_{n\to\infty}\theta(f,\lambda_n^*) = \theta_k
\end{aligned}$$

K	Sphere Lower Bound	Actual Value	Simulation Value	Cube Upper Bound	Zador Upper Bound
1	0.08333	0.08333	0.08323	0.08333	0.5
2	0.07958		0.07918	0.08333	0.5
3	0.07697		0.07900	0.08333	0.1157
4	0.07503		0.07776	0.08333	0.09974
∞	0.05854	0.05854			

Table 1. Values and Bounds for Quantization Coefficients $b_{2,k}$

using (5) with $\lambda_n^* = 2D_f(q_n)/k$ and (4). With (6) this proves that the existence of the Lagrangian limit implies that of the traditional Zador limit.

4. RESULTS

As described earlier, the optimality properties for variable rate quantizers provide a code design algorithm for variable rate vector quantizers. The empirical distribution based on a training sequence was used to estimate the expectations and probabilities. Using the approximation that the entropy of the indices yields the average length of the noiseless encoded indices, the resulting code is an entropy-constrained VQ and the algorithm minimizes the Lagrangian functional $J_\lambda(\alpha, \beta) = E(d(X, \beta(\alpha(X)))) + \lambda H(\alpha(X))$. Since the Zador constant is independent of the distribution, simulations were performed on the simplest possible nontrivial distribution, a uniform density on the k-dimensional unit cube. For this case $h(f) = 0$ and hence $J_\lambda(\alpha, \beta) = \theta(f, \lambda, q)$ so that optimizing J for small λ provides an estimate of θ_k and hence of $b_{2,k}$. For a decreasing sequence λ_n, the algorithm was run until a stopping criterion was met, where we used $(J_{old} - J)/J > .005$. The reproduction codebook was initiated with a randomly selected uniform reproduction codebook. The results are reported in the table for dimensions 1 through 4 along with the known results for dimension 1 and known upper and lower bounds. Simulations are currently running for higher dimensions. The number of training vectors and the codebook size varied from test to test. For dimension 1,2, and 3, the codebook size was 1024; for dimension 4, the codebook size was $50,000$. The preliminary results are summarized in Table 1 and simulations are continuing.

For dimension 1, the test was run five times with $50,000$ training vectors, one time with $100,000$ training vectors, and three times with $250,000$ training vectors. The preliminary results show an average value of 0.08323 for $b_{2,1}$,which is a 0.1% deviation from the actual value. We ran a similar test for the second dimension, but this time we focused on $250,000$ training vectors because algorithm performance improves as codebook size increases [3]. The test was run once with $50,000$ training vectors, once with $100,000$ training vectors, once with $500,000$ training vectors, and then finally seven times with $250,000$ training vectors. The results show an average value of 0.079184 for $b_{2,2}$. This differs from Zador's constant for fixed rate coding by 1.3%, but it is not known if this is also the constant for the variable rate case (although it has been conjectured that the two constants are the same).

For dimension 3, eight simulations were run using $250,000$ training vectors and two simulations were run using $500,000$. The results show an average of 0.079 for $b_{2,3}$, which agrees with the known bounds. For dimension 4, five simulations were run with $500,000$ training vectors. For $b_{2,4}$, an average of 0.07776 was computed.

5. CONCLUSION

Zador's traditional formulation of asymptotic (high rate) variable rate optimal quantization performance has been reformulated in a Lagrangian form. Using the Lagrangian form, the generalized Lloyd ECVQ algorithm has been used to estimate Zador's constant, which is known only for dimensions 1 and asymptotically large dimension. Simulation studies are continuing for the higher dimensions and will be reported at the conference. The Lagrangian formulation is also being used to provide a new and more general proof of Zador's results [4].

6. REFERENCES

[1] R.M. Gray and D.L. Neuhoff, "Quantization," *IEEE Transactions on Information Theory*, Vol. 44, pp. 2325–2384, October 1998.

[2] P.L. Zador, "Topics in the asymptotic quantization of continuous random variables," Bell Laboratories Technical Memorandum, 1966.

[3] P.A. Chou, T. Lookabaugh, and R.M. Gray, "Entropy constrained vector quantization," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, Vol. ASSP–37, pp. 31–42, January 1989.

[4] R.M. Gray and J. Li, "On Zador's Entropy-Constrained Quantization Theorem," *Proceedings of the 2001 IEEE Data Compression Conference (DCC)*, March 2001.