Quantization with Lagrangian Distortion Measures

Robert M. Gray
Information Systems Laboratory, Department of Electrical Engineering
Stanford, CA 94305
Email:rmgray@stanford.edu

***Abstract*— The Lagrangian formulations of fixed-rate and variable-rate vector quantization can be combined into a single formulation involving constraints on both entropy and log codebook size. The approach leads to a Lloyd design algorithm and to results on high-rate or high-resolution quantization. It provides both a unified view of the traditional results and an approach to studying the implications of Gersho's conjecure, including asymptotic quantizer point density functions. We here describe the approach, preliminary results, and open problems.**

I. Introduction

The theory of quantization derives largely from Lloyd's work [9], which formalized the optimal performance and found asymptotic approximations to the optimal performance for high-rate scalar quantization. The ideas were initially extended to vector quantizers by Zador [10]. Zador considered two separate cases of code constraints: total number of codewords (fixed-rate codes) and quantizer output entropy (variable-rate or entropy-constrained codes). The fixed-rate results of Zador were generalized and simplified by Bucklew and Wise [1] and Graf and Luschgy [5]. The entropy-constrained results of Zador were generalized in [6] using the Lagrangian formulation of [3]. Extensions to combined constraints on entropy and codebook size are presented without proof, but with heuristic arguments based on Gersho's conjecture. Details are contained in a paper in progress. See also [7], [8].

A *quantizer* or *vector quantizer* q on $\Re^k$ can be described by the following mappings and sets: an *encoder* $\alpha : \Re^k \to \mathcal{I}$, where $\mathcal{I} = \{0, 1, 2, \ldots\}$ is an index set, an associated partition $\mathcal{S} = \{S_i : i \in \mathcal{I}\}$, $S_i \subset \Re^k$, such that $\alpha(x) = i$ if $x \in S_i$, a *decoder* $\beta : \mathcal{I} \to \Re^k$, an associated reproduction codebook $\mathcal{C} = \{\beta(i) : i \in \mathcal{I}\}$ of size $N(q) = |\mathcal{C}|$, and a *length function* $\{\ell(i) : i \in \mathcal{I}\}$ which is *admissible* in the sense that $\sum_{i\in\mathcal{I}} e^{-\ell(i)} \leq 1$. Let q denote both the collection of mappings and the overall mapping $q(x) = \beta(\alpha(x))$.

The *distortion* $d(x, \hat{x})$ between an input x and a quantized version $\hat{x} = \beta(\alpha(x))$ is assumed here to be the squared error distortion $||x - \hat{x}||^2$. If X is a random vector with density f (assumed to be absolutely continuous with respect to Lebesgue measure), the *average distortion* is defined as $D_f(q) = E_f d(X, \beta(\alpha(X)))$, where E_f denotes expectation with respect to f. The average *rate* is defined by $R_f(q) = (1-\eta)E_f \ell(\alpha(X)) + \eta \ln N(q)$. Define the average Lagrangian distortion

$$\rho(f, \lambda, \eta, q) = E_f\big(d(X, \beta(\alpha(X))) + \lambda[(1-\eta)\ell(\alpha(X)) + \eta \ln N(q)]\big), \quad (1)$$

where $\lambda > 0$ and $\eta \in [0, 1]$. The corresponding optimization is to characterize

$$\rho(f, \lambda, \eta) = \inf_q \rho(f, \lambda, \eta, q). \quad (2)$$

In this context the high-rate results for the traditional cases of variable-rate coding ($\eta = 0$) and fixed-rate coding ($\eta = 1$) can be considered as special cases of the limiting behavior as $\lambda \to 0$ of

$$\theta(f, \lambda, \eta) = \frac{\rho(f, \lambda, \eta)}{\lambda} + \frac{k}{2} \ln \lambda. \quad (3)$$

The Lagrangian form of the variable-rate result [6] is that under suitable conditions

$$\lim_{\lambda\to 0} \theta(f, \lambda, 0) = h(f) + \theta_k, \quad (4)$$

where $h(f)$ is the differential entropy of the pdf f and $\theta_k \triangleq \inf_{\lambda>0} \theta(u, \lambda, 0)$, and u is the uniform pdf on a unit cube. The Lagrangian form of the fixed-rate result [8] is

$$\lim_{\lambda\to 0} \theta(f, \lambda, 1) = \psi_k + \ln ||f||^{k/2}_{k/(k+2)}, \quad (5)$$

where $\psi_k \triangleq \inf_{\lambda>0} \theta(u, \lambda, 1)$. The proofs of these results tend to be tedious, but they largely follow Zador's original approach:

Step 1 Prove the result for a uniform density on a cube.
Step 2 Prove the result for a pdf that is piecewise constant on disjoint cubes of equal volume.

Step 3 Prove the result for a general pdf on a cube by approximating it by a piecewise constant pdf on small cubes.

Step 4 Extend the result for a general pdf on the cube to general pdfs on $\Re^k$ by limiting arguments.

The uniform density u on $[0,1)^k$ plays a fundamental role both as Step 1 and as a simple example sufficient for developing the properties of the Zador constants related to θ_k and ψ_k.

The form of the Lagrangian approach suggests a conjecture for the asymptotic behavior consistent with the high-rate Lagrangian fixed- and variable-rate cases:

$$\boxed{\lim_{\lambda\to 0}\left(\frac{\rho(f,\lambda,\eta)}{\lambda}+\frac{k}{2}\ln\lambda\right)=\theta(f,\eta)\,,} \tag{6}$$

where $\theta(f,\eta)$ is finite. Furthermore, the particular form of the traditional cases suggests a second conjecture—that the asymptotically optimal performance $\theta(f,\eta)$ can be expressed as a sum of two terms, the first involving an infimum for the uniform density on a unit cube and the second depending on the specific pdf:

$$\theta(f,\eta)=\theta_k(\eta)+h(f,\eta), \quad \text{where} \tag{7}$$

$$\theta_k(\eta)\triangleq\inf_{\lambda>0}\left(\frac{\rho(u,\lambda,\eta)}{\lambda}+\frac{k}{2}\ln\lambda\right). \tag{8}$$

From the traditional variable-rate and fixed-rate cases, $\theta_k(0)=\theta_k$ and $h(f,0)=h(f)$, whereas $\theta_k(1)=\psi_1$ and $h(f,1)=\ln||f||_{k/(k+2)}^{k/2}$.

II. Lloyd optimality conditions

For $\lambda>0$ and $\eta\in(0,1)$, the Lloyd optimality properties become

- For a given decoder β and length function ℓ, the optimal encoder satisfies $\alpha(x)=\operatorname{argmin}_i\left(d(x,\beta(i))+\lambda\eta\ell(i)\right)$.
- For a given encoder α, the optimal decoder satisfies $\beta(i)=\operatorname{argmin}_y E(d(X,y|\alpha(X)=i)$ if the minimum exists.
- For a given encoder α, the optimal codelength is $\ell(i)=-\ln\Pr(\alpha(X)=i)$. Therefore $E_f\ell(\alpha(X))=H_f(q)$, the Shannon entropy of the quantizer output.
- A necessary condition for optimality of a fixed-rate quantizer q with codebook $\mathcal{C}$ is that there be no subcodebook $\mathcal{C}'\subset\mathcal{C}$ for which

$$\begin{aligned}&D_f(\mathcal{C}')+\lambda\left[(1-\eta)H_f(\mathcal{C}')+\eta\ln|\mathcal{C}'|\right]\\ &<\ D_f(\mathcal{C})+\lambda\left[(1-\eta)H_f(\mathcal{C})+\eta\ln|\mathcal{C}|\right],\end{aligned} \tag{9}$$

where $H_f(\mathcal{C})$ and $H_f(\mathcal{C}')$ denote the entropies of the partitions corresponding to the optimal encoding for codebook $\mathcal{C}$ and $\mathcal{C}'$.

Given a codebook $\mathcal{C}$ (or decoder β) or partition $\mathcal{S}$ (or encoder α), the Lloyd properties determine the remaining components, so optimizing over quantizers is equivalent to optimizing over codebooks or partitions.

III. Asymptotically optimal quantizers

It is convenient to consider the quantity defined by normalizing the average Lagrangian distortion by λ and adding a weighted $\ln\lambda$ term. Define

$$\begin{aligned}\theta(f,\lambda,\eta,\mathcal{S})&=\frac{\rho(f,\lambda,\eta,\mathcal{S})}{\lambda}+\frac{k}{2}\ln\lambda\\ &=\frac{D_f(\mathcal{S})}{\lambda}+[(1-\eta)H_f(\mathcal{S})+\eta\ln N(\mathcal{S})]+\frac{k}{2}\ln\lambda\end{aligned}$$

and the related quantities $\theta(f,\lambda,\eta)=\inf_{\mathcal{S}}\theta(f,\lambda,\eta,\mathcal{S})$, $\overline{\theta}(f,\eta)=\limsup_{\lambda\to0}\theta(f,\lambda,\eta)$, and $\underline{\theta}(f,\eta)=\liminf_{\lambda\to0}\theta(f,\lambda,\eta)$. The main conjecture (6) is true if and only if $\underline{\theta}(f,\eta)=\overline{\theta}(f,\eta)$, in which case the common value is denoted $\theta(f,\eta)$.

Lemma 1: $\theta(f,\lambda,\eta,\mathcal{S})$ and $\theta(f,\lambda,\eta)$ are monotonic nondecreasing, concave, and continuous functions of η.

Corollary 1: Suppose that for a pdf f the conjecture (6) holds for all $\eta\in[0,1]$ and hence

$$\theta(f,\eta)=\lim_{\lambda\to0}\theta(f,\lambda,\eta)$$

exists for $\eta\in[0,1]$. Then $\theta(f,\eta)$ is a monotone nondecreasing concave function of η and it is continuous except possibly at the origin. Furthermore, $\theta'(f,\eta)=d\theta(f,\eta)/d\eta$ exists and is finite for all $\eta\in(0,1)$ except possibly on a set of Lebesgue measure 0.

If equation (6) holds for f and η, then for every sequence $\lambda_n\to0$ there exists a sequence of partitions $\mathcal{S}_n$ for which

$$\lim_{n\to\infty}\theta(f,\lambda_n,\eta,\mathcal{S}_n)=\theta(f,\eta), \tag{10}$$

in which case we say that the sequence $\mathcal{S}_n$ is (η,λ_n)-asymptotically optimal or simply (η,λ_n)-a.o. for the pdf f.

If $\mathcal{S}_n$ is (η,λ_n)-a.o. for a pdf f, then (12) can be used as in the fixed-rate and variable-rate cases to show that

$$\lim_{n\to\infty}\frac{k}{2}\ln\left(\frac{2e}{k}D_f(\mathcal{S}_n)e^{\frac{2}{k}[(1-\eta)H_f(\mathcal{S})+\eta\ln|\mathcal{S}|]}\right)=\theta(f,\eta). \tag{11}$$

A sequence of partitions $\mathcal{S}_n$ satisfying (11) will be called η-asymptotically optimal or briefly η-a.o.

for the pdf f. If a sequence is (η, λ_n)-a.o. for any sequence λ_n, then it is eta-a.o. Conversely if it is η-a.o., then it is (η, λ_n)-a.o. for some $\lambda_n \to 0$.

In the traditional cases the asymptotic behavior of distortion and rate can be teased apart from the linear combination of the two in (10) [7], [8]. These results can be extended to the combined constraint case using the following inequality based on the $\ln r \le r - 1$ inequality:

$$\theta(f, \lambda, \eta, \mathcal{S}) \ge \frac{k}{2} \ln \left(\frac{2e}{k} D_f(\mathcal{S}) e^{\frac{2}{k} [(1-\eta) H_f(\mathcal{S}) + \eta \ln |\mathcal{S}|]} \right) \quad (12)$$

with equality if and only if $\lambda = 2D_f(\mathcal{S})/k$. If $\mathcal{S}_n$ is η-a.o. for f, then

$$\lim_{n \to \infty} \frac{2D_f(\mathcal{S}_n)}{k\lambda_n} = 1 \quad (13)$$

$$\lim_{n \to \infty} \Big((1-\eta) H_f(\mathcal{S}_n) + \eta \ln |\mathcal{S}_n| + \frac{k}{2} \ln \lambda_n \Big) = \theta(f, \eta) - \frac{k}{2}. \quad (14)$$

If, in addition, we make the assumption of Corollary 1 that (6) holds for all $\eta \in [0, 1]$, then we can also separate out the behavior of $H_f(\mathcal{S}_n)$ and $\ln |\mathcal{S}_n)|$ in terms of λ_n:

$$\lim_{n \to \infty} (\ln |\mathcal{S}_n| - H_f(\mathcal{S}_n)) = \theta'(f, \eta). \quad (15)$$

Perhaps surprisingly, the two growing terms $\ln |\mathcal{S}_n|$ and $H_f(\mathcal{S}_n))$ differ only by a constant in the limit if $\theta'(f, \eta)$ is finite! Combining the previous results yields the following corollary which separates out the asymptotic behavior of distortion, entropy, codebook size, and the difference between the entropy and codebook size.

Lemma 2: Suppose that for a pdf f the conjecture (6) holds for all $\eta \in [0, 1]$, hence $\theta(f, \eta) = \lim_{\lambda \to 0} \theta(f, \lambda, \eta)$ exists for $\eta \in [0, 1]$. Then for almost all $\eta \in (0, 1)$, if $\mathcal{S}_n$ is (η, λ_n)-a.o. then

$$\lim_{n \to \infty} \frac{2D_f(\mathcal{S}_n)}{k\lambda_n} = 1$$

$$\lim_{n \to \infty} \Big(H_f(\mathcal{S}_n) + \frac{k}{2} \ln \lambda_n \Big) = \theta(f, \eta) - \eta\theta'(f, \eta) - \frac{k}{2}$$

$$\lim_{n \to \infty} \Big(\ln |S_n| + \frac{k}{2} \ln \lambda_n \Big) = \theta(f, \eta) + (1-\eta)\theta'(f, \eta) - \frac{k}{2}$$

$$\lim_{n \to \infty} \big(\ln |S_n| - H_f(\mathcal{S}_n) \big) = \theta'(f, \eta).$$

IV. Heuristic Derivation

Gersho's conjecture and the associated approximations can be used to derive the basic results for joint entropy and codebook size constrained quantization. This approach, although not rigorous, provides insight into the results and a consistency check with the rigorous development. An obvious modification suggests a solution to the general case.

Gersho's conjecture involves two assumptions regarding asymptotically optimal sequences of fixed-rate and variable-rate quantizers. First, it is assumed that there exists a quantizer point density function $\Lambda(x)$ such that a sequence of optimal codes with N codewords, $N = 1, 2, \ldots$ will satisfy

$$\lim_{N \to \infty} \frac{1}{N} \times (\text{ \# reproduction vectors in a set } S) = \int_S \Lambda(x)\, dx; \quad \text{all } S, \quad (16)$$

where $\int_{\Re^k} \Lambda(x)\, dx = 1$. Second, Gersho assumed that If $f_X(x)$ is smooth and R is large, then the minimum distortion quantizer has cells S_i that are (approximately) scaled, rotated, and translated copies of S^*, the convex polytope that tesselates $\mathcal{R}^k$ with minimum normalized moments of inertia

$$M(S) = \frac{1}{kV(S)^{2/k}} \int_S \frac{||x - y(S)||^2}{V(S)}\, dx$$

where $y(S)$ denotes the centroid of S. Specifically, define

$$c_k = \min_{\text{tesselating convex polytopes } S} M(S).$$

Under these assumptions, it can be argued using Riemann approximations of integrals and sums that for large N

$$D_f(q) \approx c_k E_f \left(\Big(\frac{1}{N(q)\Lambda(X)}\Big)^{2/k} \right) \quad (17)$$

$$H_f(q(X)) \approx h(X) - E\left(\log\Big(\frac{1}{N(q)\Lambda(X)}\Big) \right) = \ln N(q) - H(f||\Lambda), \quad (18)$$

where the relative entropy $H(f||\lambda) = \int f(x) \ln(f(x)/\Lambda(x))\, dx$. Application of Holder's inequality to (17) yields the classic fixed-rate result and the combination of (17) and (18) with Jensen's inequality yields the classic variable-rate result. Suppose that a quantizer q has a quantizer point density Λ and a total of N quantization levels for N large, then

$$\theta(f, \lambda, \eta, \Lambda) \approx \frac{c_k E_f \left((N\Lambda(X))^{-2/k} \right)}{\lambda} + (1-\eta)[\ln N - H(f||\Lambda)] + \eta \ln N + \frac{k}{2} \ln \lambda.$$

If the quantizer point density Λ is fixed, then the optimum choice of N is the that which minimizes

$$\frac{c_k N^{-2/k} E_f\left((\Lambda(X))^{-2/k}\right)}{\lambda} + \ln N \geq \frac{k}{2}\ln\left(\frac{2ec_k}{k}\frac{E_f\left((\Lambda(X))^{-2/k}\right)}{\lambda}\right)$$

with equality if and only if

$$N = \left[\frac{2}{k}\frac{c_k E_f\left((\Lambda(X))^{-2/k}\right)}{\lambda}\right]^{k/2}. \qquad (19)$$

Thus

$$\theta(f,\lambda,\eta,\Lambda) = \frac{k}{2}\ln\left(\frac{kec_k}{2}\right) + (1-\eta)h(f) + \frac{k}{2}\ln\left(E_f\left((\Lambda(X))^{-2/k}\right)\right) + (1-\eta)E_f\left(\ln\Lambda(X)\right), \qquad (20)$$

and the goal becomes the minimization of $\theta(f,\lambda,\eta,\Lambda)$ over all point density functions Λ. Gersho's conjecture and this heuristic approach imply that $(1/2)\ln(kec_k/2) = \theta_k = \psi_k$, the Zador fixed-rate and variable-rate constants are equal, but no proof of this result exists except for $k = 1$ and asymptotically as $k \to \infty$. For our purposes we can identify $(k/2)\ln(kec_k/2)$ as $\theta_k(\eta)$ when comparing (20) with the rigorous results, which results in the conjecture

$$\theta(f,\eta) = \theta_k(\eta) + h(f,\eta) \qquad (21)$$

$$h(f,\eta) = (1-\eta)h(f) + \inf_{\Lambda}\left[\frac{k}{2}\ln\left(E_f\left((\Lambda(X))^{-2/k}\right)\right) + (1-\eta)E_f\left(\ln\Lambda(X)\right)\right]. \qquad (22)$$

This leads to the traditional results when $\eta = 0, 1$.

V. Some known results

Conjecture (6) has been proved for uniform densities on cubes [8].

Theorem 1:

$$\theta_k(\eta) \triangleq \inf_{\lambda>0}\theta(u,\lambda,\eta) = \lim_{\lambda\to 0}\theta(u,\lambda,\eta) \qquad (23)$$

Thus η-a.o. quantizers exist for such densities and the previous results Lemma 2 concerning the behavior $\theta(f,\eta)$ apply to $\theta(u,\eta) = \theta_k(\eta)$. For example, from the remarks following conjecture 3 of [7], the lemma implies that for almost all η, the asymptotic quantizer point density function exists and is uniform for the uniform distribution, extending the known result for fixed-rate coding to the combined case.

Step 2 assumes that the pdf f is nonzero on the union of a finite number M of disjoint cubes $\{C(m) : m = 1, 2, \ldots, M\}$ on which it is constant:

$$f(x) = \sum_m w_m f_m(x) = \sum_m w_m a^{-k} 1_{C(m)}(x), \qquad (24)$$

where $\sum_m w_m = 1$ and and a^k is the volume of each $C(m)$. Design for each cube $C(m)$ a nearly optimal code with partition $\mathcal{S}_m$ for the conditionally uniform pdf f_m using a Lagrange multiplier λ_m and a common value of η for all m. For the moment we leave open the choice of the λ_m except for the assumption that they are all small enough for Lemma 2 to apply. After some algebra, the overall Lagrangian distortion becomes

$$\theta(f,\lambda,\eta,\mathcal{S}) \approx \theta_k(\eta) - \frac{k}{2} + (1-\eta)H(w) + k\ln a + \frac{k}{2}\sum_m \frac{\lambda_m}{\lambda}w_m - (1-\eta)\frac{k}{2}\sum_m w_m \ln\frac{\lambda_m}{\lambda} + \eta\ln\sum_m\left(\frac{\lambda_m}{\lambda}\right)^{-k/2}.$$

So the goal is to minimize the function $\theta(f,\eta,\mu) = \theta_k(\eta) + \phi(w,\eta,\{\lambda_m\})$ over $\{\lambda_m\}$. Unfortunately, ϕ is not a convex function of the Lagrangian multipliers. However, transforming variables as $\nu_m = \ln(\lambda_m/\lambda)$ results in a convex optimization problem

$$h(f,\eta) \triangleq \min_\nu \phi(f,\eta,\nu) = \min_\nu\left((1-\eta)F_0(w,\nu) + \eta F_1(w,\nu)\right), \qquad (25)$$

where

$$F_0(w,\nu) = \phi(w,0,\nu) = \frac{k}{2}\sum_m w_m e^{\nu_m} + H(w) - \frac{k}{2}\sum_m w_m\nu_m + \frac{k}{2}\ln\frac{a^2}{e} \qquad (26)$$

$$F_1(w,\nu) = \phi(w,1,\nu) = \frac{k}{2}\sum_m w_m e^{\nu_m} + \ln\sum_m e^{-k\nu_m/2} + \frac{k}{2}\ln\frac{a^2}{e}. \qquad (27)$$

Since $\phi(w,\eta,\nu)$ is strictly convex in ν, there must be a unique minimum. Since $\phi(w,\eta,\nu)$ is differentiable with respect to ν, the minimum must be at a point with 0 gradient, which implies

$$w_m e^{\nu_m} = (1-\eta)w_m + \eta\frac{e^{-\frac{k}{2}\nu_m}}{\sum_n e^{-\frac{k}{2}\nu_n}}. \qquad (28)$$

Strict convexity of $\phi(w,\eta,\nu)$ guarantees the existence of a ν satisfying this equation and furthermore that ν minimizes $\phi(w,\eta,\nu)$. Unfortunately there seems to be no nice closed form solution for ν in terms of w.

Combining the above arguments with the careful limiting arguments proves the following.

Theorem 2: If f is a piecewise constant pdf of the form given in (24), then

$$\limsup_{\lambda\to 0} \theta(f,\eta,\lambda) \leq \theta_k(\eta) + h(f,\eta), \tag{29}$$

with $h(f,\eta)$ given by (25).

This proves the positive part of (6), but to prove the conjecture requires a converse to the effect that

$$\liminf_{\lambda\to 0} \theta(f,\eta,\lambda) \geq h(f,\eta). \tag{30}$$

Unfortunately the converse has proved more difficult than in either of the traditional cases, but we conjecture that it holds based on the fact that a development based on Gersho's conjecture and approximations is consistent with our conjecture.

The next step is to generalize from piecewise constant pdfs on a cube to more general pdfs on the unit cube. The arguments for the piecewise continuous case extend to this case and also to unbounded support sets with a moment condition and show that

Theorem 3: If f is a pdf satisfying the moment condition of $E_f(\|X\|^{2+\delta}) \leq \infty$ for some $\delta > 0$, then the result of Theorem 2 holds with

$$\phi(f,\eta,\nu) = (1-\eta)F_0(f,\nu) + \eta F_1(f,\nu)$$

where

$$F_0(f,\nu) = \phi(f,0,\nu) = \frac{k}{2}\int f(x)\left(e^{\nu(x)} - \nu(x) - 1\right)dx + h(f)$$

$$F_1(f,\nu) = \phi(f,1,\nu) = \frac{k}{2}\int f(x)\left(e^{\nu(x)} - 1\right)dx + \ln\left(\int e^{-k\nu(x)/2}\,dx\right)$$

where

$$\Lambda(x) = \frac{e^{-k\nu(x)/2}}{\int e^{-k\nu(y)/2}\,dy}. \tag{31}$$

$\phi(f,\eta,\nu)$ is a strictly convex function of ν. Unfortunately, in this infinite dimensional case convexity does not guarantee the existence of a minimizing ν and hence further assumptions are needed. It does, however, guarantee that if a minimizing ν exists, it is unique (at least up to a set of measure zero). In particular, if there is a local minimum of $\phi(f,\eta,\nu)$ with respect to ν, then it is the unique global minimum. By adding more assumptions (in particular, that f is such that $\phi(f,\eta,\nu)$ is twice continuously differentiable), a calculus of variations argument results in the conditions

$$f(x)e^{\nu(x)} = (1-\eta)f(x) + \eta\frac{e^{-k\nu(x)/2}}{\int e^{-k\nu(y)/2}\,dy}, \tag{32}$$

which is the continuous analog of the piecewise constant result. Note that as in the piecewise constant case, the minimizing ν must satisfy

$$\int f(x)e^{\nu(x)}\,dx = 1. \tag{33}$$

Transforming the variables to Λ using (31) yields the identical optimization to (22). This suggests that the Λ arising in the Lagrangian optimization is in fact Gersho's quantizer point density function.

Acknowledgments

The author gratefully acknowledges discussions with Stephen Boyd, Tamas Linder, John Gill,Michelle Effros and Deirdre O'Brien. This work was partially supported by the National Science Foundation under NSF Grant CCR-0073050 and by gifts from the Hewlett-Packard coorporation.

References

[1] J. A. Bucklew and G. L. Wise, "Multidimensional asymptotic quantization theory with rth power distortion measures," *IEEE Trans. Inform. Theory*, vol. IT-28, pp. 239–247, Mar. 1982.

[2] J. A. Bucklew, "Two results on the asymptotic performance of quantizers," *IEEE Trans. Inform. Theory*, vol. IT-30, pp. 341–348, Mar. 1984.

[3] P. A. Chou, T. Lookabaugh, and R. M. Gray. Entropy-constrained vector quantization. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 37:31–42, Jan. 1989.

[4] A. Gersho, "Asymptotically optimal block quantization," *IEEE Trans. Inform. Theory.*, vol. 25, pp. 373–380, July 1979.

[5] S. Graf and H. Luschgy, *Foundations of Quantization for Probability Distributions*, Springer, Lecture Notes in Mathematics, 1730, Berlin, 2000.

[6] R. M. Gray, T. Linder, and J. Li, "A Lagrangian formulation of Zador's entropy-constrained quantization theorem," *IEEE Trans. Inform. Theory*, pp. 695–707, vol. 48, Mar. 2002.

[7] R. M. Gray and T. Linder, "Results and Conjectures on High Rate Quantization," *Proceedings of the 2004 IEEE Data Compression Conference*.

[8] R. M. Gray and J. T. Gill, III, "A Lagrangian formulation of fixed rate and entropy/memory constrained quantization," presented at the *2005 IEEE Data Compression Conference*, http://ee.stanford.edu/~gray/dcc05.pdf.

[9] S. P. Lloyd. Least squares quantization in PCM. *IEEE Transactions on Information Theory*, 1957. Bell Laboratories Technical Note. Reprinted in *IEEE Transactions on Information Theory*, 28:127–135, March 1982.

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