Mismatch in High-Rate Entropy-Constrained Vector Quantization

Robert M. Gray, *Fellow, IEEE,* and Tamás Linder, *Senior Member, IEEE*

***Abstract*—Bucklew's high-rate vector quantizer mismatch result is extended from fixed-rate coding to variable-rate coding using a Lagrangian formulation. It is shown that if an asymptotically (high-rate) optimal sequence of variable rate codes is designed for a k-dimensional probability density function (pdf) g and then applied to another pdf f for which f/g is bounded, then the resulting mismatch or loss of performance from the optimal possible is given by the relative entropy or Kullback–Leibler divergence $I(f\|g)$. It is also shown that under the same assumptions, an asymptotically optimal code sequence for g can be converted to an asymptotically optimal code sequence for a mismatched source f by modifying only the lossless component of the code. Applications to quantizer design using uniform and Gaussian densities are described, including a high-rate analog to the Shannon rate-distortion result of Sakrison and Lapidoth showing that the Gaussian is the "worst case" for lossy compression of a source with known covariance. By coupling the mismatch result with composite quantizers, the worst case properties of uniform and Gaussian densities are extended to conditionally uniform and Gaussian densities, which provides a Lloyd clustering algorithm for fitting mixtures to general densities.**

***Index Terms*—Entropy constrained, high rate, Kullback–Leibler divergence, Lagrangian, mismatch, quantization, relative entropy, variable rate.**

I. Introduction

THE optimal performance of high-rate vector quantization using fixed-rate codes was established in Zador's classic Bell Labs Technical Memo [35] and generalized and simplified by Bucklew and Wise [2] and Graf and Luschgy [14]. These results characterized the optimal rate–distortion tradeoff of fixed-dimension vector quantization as the rate or codebook size grows asymptotically large, in contrast to the Shannon rate–distortion theory results characterizing the tradeoff for fixed rate when the dimension becomes asymptotically large. The history and generality of the results may be found, e.g., in [18]. Bucklew [3] developed further asymptotic properties of high-rate quantization, most notably providing a mismatch result that quantified the performance resulting when a sequence of quantizers that is asymptotically optimal for one source is applied instead to another "mismatched" source. Such mismatch results are important for theory and potentially important for practice as code designs are often based on source models which are estimated based on data and hence which are often inaccurate. Mismatch results provide a means of quantifying such performance variations. Another potential application of mismatch performance results is in the design of "robust" source codes. Sakrison [30] and Lapidoth [26] showed that for large dimensions, Gaussian sources provide a "worst case" or "robust" approach to code design in that the Gaussian source has the largest (worst) Shannon rate–distortion function and, more importantly, that a code designed for a Gaussian model will yield approximately the same performance when applied to any source with the same mean and covariance. (See [10] for a fixed-distortion analog of this result.) The Gaussian code will, of course, be suboptimal for the non-Gaussian source, but it will provide "robust" or reliable performance in the sense that the resulting rate and distortion will be the same whichever source the code is applied to. The results of Sakrison and Lapidoth, and related works on mismatch by Yang and Kieffer [33] and Zamir [37], are asymptotic in the typical Shannon fashion; large dimensions are required in order to apply Shannon source-coding arguments. In contrast, Bucklew's approach considers fixed dimension and asymptotically large rate.

Zador also developed the rate–distortion tradeoffs for entropy-constrained vector quantizers [35], but these results have only recently been generalized [20] to conditions of comparable generality to the fixed-rate results of Bucklew and Wise [2] and Graf and Luschgy [14]. No rigorous variable-rate mismatch results comparable to the fixed-rate results of Bucklew [3] are known to the authors prior to those reported here. The primary goal of this paper is to establish a general variable-rate mismatch result following the Lagrangian approach of [20]. Applications to universal quantization, robust quantization, and clustering mixture densities are described. In the special case of scalar quantization, the results are related to the known asymptotic optimality of uniform quantization followed by optimal entropy coding.

The basic mismatch results derived here have been previously reported, but only a heuristic derivation was given which was based on Gersho's hypothesis and the standard approximations which follow from that hypothesis [16], [22], [17].

II. Preliminaries

We begin with a review of the needed results from [20]. $(\Omega, \mathcal{B})$ is the measurable space consisting of the k-dimensional Euclidean space $\Omega = \Re^k$ and its Borel subsets. Assume that

Manuscript received April 26, 2002; revised December 23, 2002. This work was supported in part by the National Science Foundation under Grant 0073050, by the Hewlett Packard Corporation, and by Natural Sciences and Engineering Research Council (NSERC) of Canada. The material in this paper was presented in part at the IEEE Data Compression Conference, Snowbird, UT, March 2003.

R. M. Gray is with the Information Systems Laboratory, Department of Electrical Engineering, Stanford University, Stanford, CA 94305 USA (e-mail: rmgray@stanford.edu).

T. Linder is with the Department of Mathematics and Statistics, Queen's University, Kingston, ON K7L 3N6, Canada (e-mail: linder@mast.queensu.ca).

Communicated by M. Weinberger, Associate Editor for Source Coding.

Digital Object Identifier 10.1109/TIT.2003.810637

X is random vector with a distribution P_f, which is absolutely continuous with respect to the Lebesgue measure V and hence possesses a probability density function (pdf) $f = dP_f/dV$ so that

$$P_f(F) = \int_F f(x)\,dV(x) = \int_F f(x)\,dx$$

for any $F \in \mathcal{B}$. The volume of a set $F \in \mathcal{B}$ is given by its Lebesgue measure $V(F) = \int_F dx$. We assume that the differential entropy

$$h(f) \triangleq -\int dx f(x) \ln f(x)$$

exists and is finite. The unit of entropy is nats or bits according to whether the base of the logarithm is e or 2. Usually, nats will be assumed, but bits will be used when entropies appear in an exponent of 2.

A vector quantizer Q can be described by the following mappings and sets.

- An encoder α: $\Omega \to \mathcal{I}$, where $\mathcal{I}$ is a countable index set, and an associated measurable partition $\mathcal{S} = \{S_i;\ i \in \mathcal{I}\}$ such that $\alpha(x) = i$ if $x \in S_i$. If $\mathcal{I}$ is finite with $N \geq 1$ elements, we set $\mathcal{I} = \{0, 1, \ldots, N-1\}$; otherwise, we let $\mathcal{I} = \{0, 1, 2, \ldots\}$.
- A decoder β: $\mathcal{I} \to \Omega$ and an associated reproduction codebook $\mathcal{C} = \{\beta(i);\ i \in \mathcal{I}\}$. Without loss of generality, we assume that the codevectors $\beta(i);\ i \in \mathcal{I}$ are all distinct.
- An index coder ψ: $\mathcal{I} \to \{0, \ldots, D-1\}^*$, the space of all finite-length D-ary strings, and the associated length L: $\mathcal{I} \to \{1, 2, \ldots\}$ defined by $L(i) = \text{length}(\psi(i))$. ψ is assumed to be uniquely decodable (a lossless or noiseless code) and hence the resulting set of lengths must satisfy the Kraft inequality (e.g., [6]) for some channel code alphabet size D

$$\sum_i D^{-L(i)} \leq 1. \tag{1}$$

It is convenient to measure channel codeword lengths in a normalized fashion and hence we define the length function of the code in nats as $\ell(i) = L(i) \ln D$ so that Kraft's inequality becomes

$$\sum_i e^{-\ell(i)} \leq 1. \tag{2}$$

A set of code lengths $\ell(i)$ is said to be *admissible* if (2) holds. Following Cover and Thomas [6] and [20] it is convenient to remove the restriction of integer D-ary code lengths and hence we define any collection of nonnegative real numbers $\ell(i);\ i \in \mathcal{I}$ to be an *admissible length function* if it satisfies (2). The primary reason for dropping the constraint is to provide a useful tool for proving results, but the general definition can be interpreted as an approximation since if $\ell(i)$ is an admissible length function, then for a code alphabet of size D the actual integer code lengths $L(i) = \lceil \frac{\ell(i)}{\ln D} \rceil$ will satisfy the Kraft inequality. (Throughout this paper, $\lceil t \rceil$ denotes the smallest integer not less than t, and $\lfloor t \rfloor$ denotes the largest integer not greater than t.) Let $\mathcal{A}$ denote the collection of all admissible length functions ℓ.

To summarize, a vector quantizer Q is completely described by a triple (α, β, ℓ) consisting of encoder, decoder, and admissible length function. We will abbreviate the overall action of producing a reproduction from an input, the cascade of decoder and encoder, using a lower case q

$$q(x) = \beta(\alpha(x)). \tag{3}$$

A quantizer of particular interest is the uniform quantizer with side length Δ. For $\Delta > 0$, let Q_Δ denote a quantizer of Ω into contiguous cubes of side Δ. In other words, Q_Δ can be viewed as a uniform scalar quantizer with bin size Δ applied k successive times. We assume the axes of the cubes align with the coordinate axes (and that point 0 is touched by corners of cubes). In particular, Q_1 is a cubic lattice quantizer with unit volume cells.

The instantaneous rate of a quantizer is defined by $r(\alpha(x)) = \ell(\alpha(x))$. The average rate is

$$\begin{aligned} R_f(Q) &= R_f(\alpha, \ell) \\ &= E_f r(\alpha(X)) \\ &= \int dx f(x) \ell(\alpha(x)) \\ &= \sum_i p_i \ell(i) \end{aligned}$$

where $p_i = \Pr(\alpha(X) = i) = P_f(S_i)$ for all $i \in \mathcal{I}$.

Given a quantizer Q, the entropy of the quantizer is defined in the usual fashion by

$$H_f(Q) = -\sum_i p_i \ln p_i$$

and we assume that $p_i > 0$ for all i. Note that we could also write $H_f(q)$ or $H_f(\alpha)$ since the entropy depends only on the encoder or, since the reproduction words are assumed distinct, on the cascade of encoder and decoder.

For any admissible length function ℓ, the divergence inequality [6] implies that

$$R_f(Q) \geq H_f(Q)$$

with equality if and only if

$$\ell(i) = -\ln p_i. \tag{4}$$

Thus, in particular

$$H_f(Q) = \inf_{\ell \in \mathcal{A}} R_f(\alpha, \ell). \tag{5}$$

We assume a distortion measure $d(x, \hat{x}) \geq 0$ and measure performance by average distortion

$$\begin{aligned} D_f(Q) &= D_f(\alpha, \beta) \\ &= Ed(X, q(X)) \\ &= Ed(X, \beta(\alpha(X))). \end{aligned}$$

In particular, we initially assume squared-error distortion with average

$$d(x, \hat{x}) = \|x - \hat{x}\|^2 = \sum_{i=1}^{k} |x_i - \hat{x}_i|^2$$

for $x = (x_1, \ldots, x_k)$ and $\hat{x} = (\hat{x}_1, \ldots, \hat{x}_k)$. In Section VIII, another type of distortion measure will be considered for the quantization of signals into models.

The traditional distortion–rate approach defines the optimal performance as the minimum distortion achievable for a given rate

$$\delta_f(R) = \inf_{q:\, R_f(Q) \le R} D_f(Q) \tag{6}$$

$$= \inf_{\alpha, \beta, \ell:\, R_f(\alpha, \ell) \le R} D_f(\alpha, \beta). \tag{7}$$

The traditional form of Zador's theorem states that under suitable assumptions on f

$$\lim_{R \to \infty} 2^{\frac{2}{k} R} \delta_f(R) = b(2, k) 2^{\frac{2}{k} h(f)} \tag{8}$$

where $b(2, k)$ is Zador's constant, which depends only on k and not f. Zador did not evaluate the constant $b(2, k)$ but he did provide upper and lower bounds that become tight for large k.

Zador's basic results contained technical errors and restrictive conditions on the allowed densities. These problems were discussed and fixed and the results generalized recently [20] using a Lagrangian approach, which we turn to next.

The Lagrangian formulation of variable-rate vector quantization [5] defines for each value of a Lagrangian multiplier $\lambda > 0$ a Lagrangian distortion

$$\rho_\lambda(x, i) = d(x, \beta(i)) + \lambda \ell(i)$$

and corresponding performance

$$\begin{aligned} \rho(f, \lambda, Q) &= E d(X, q(X)) + \lambda E \ell(\alpha(X)) \\ &= D_f(Q) + \lambda R_f(Q) \end{aligned}$$

and an optimal performance

$$\rho(f, \lambda) = \inf_Q \rho(f, \lambda, Q)$$

where the infimum is over all quantizers $Q = (\alpha, \beta, \ell)$ where ℓ is assumed admissible. Unlike the traditional formulation, the Lagrangian formulation yields Lloyd optimality conditions for vector quantizers, that is, a necessary condition for optimality is that each of the three components of the quantizer be optimal for the other two.

- For a given decoder β and length function ℓ, the optimal encoder is

  $$\alpha(x) = \operatorname*{argmin}_i \left(d(x, \beta(i)) + \lambda \ell(i) \right)$$

  (ties are broken arbitrarily).

- The optimal decoder for a given encoder and length function is the usual Lloyd centroid

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

- From (4), the optimal length function for the given encoder and decoder is

  $$\ell(i) = -\ln p_i.$$

The following asymptotic result is the primary result of [20].

Theorem 1: Assume that the distribution P_f of X is absolutely continuous with respect to Lebesgue measure V with pdf $f = dP_f/dV$, that the differential entropy $h(f)$ exists and is finite, and that $H_f(Q_1) < \infty$. Then

$$\lim_{\lambda \to 0} \left(\frac{\rho(f, \lambda)}{\lambda} + \frac{k}{2} \ln \lambda \right) = h(f) + \theta_k \tag{9}$$

where the finite constant θ_k is defined by

$$\theta_k \triangleq \inf_{\lambda > 0} \left(\frac{\rho(u_1, \lambda)}{\lambda} + \frac{k}{2} \ln \lambda \right) \tag{10}$$

and u_1 is the uniform pdf on the k-dimensional unit cube $[0, 1)^k$.

In particular, the limiting constant θ_k depends only on the dimension and not on the pdf. It is also shown in [20] that under the stated assumptions, Zador's original result (8) and the Lagrangian formulation are equivalent and

$$\theta_k = \frac{k}{2} \ln \frac{2e}{k} b(2, k). \tag{11}$$

In 1968, Gish and Pierce [13] claimed that for high rate, the optimal entropy-constrained *scalar* quantizer performance was $\delta_f(R) \approx (1/12) 2^{2(h(f) - R)}$, where $\approx$ denotes asymptotic equality in the sense that the ratio of the two sides converges to 1 as $R \to \infty$, and that uniform quantization followed by optimal entropy coding achieved the optimal rate–distortion tradeoff in the limit of asymptotically large rate. Gish and Pierce gave a heuristic proof based on companding along with a rigorous one for continuous densities that satisfy a tail condition. They also claimed to have a proof for all uniformly continuous densities, but omitted it in the paper and it was not subsequently published. Since a uniform density is included in their conditions, however, their results coupled with Zador's theorem imply that $b(2, 1) = 1/12$. In the scalar case, it can be argued that a sequence of increasing rate uniform quantizers followed by optimum entropy coders will be asymptotically optimal if the assumptions of Theorem 1, which are more general than those of Gish and Pierce, are met. This follows from a result of [28] which shows that under assumptions equivalent to those of Theorem 1 a sequence of k-dimensional lattice vector quantizers Q_R will have asymptotic (large entropy R) distortion of $D(Q_R) \approx G(\Lambda) 2^{(2/k)(h(f) - R)}$, where $G(\Lambda)$ is the normalized moment of inertia of the basic Voronoi cell of the base lattice Λ. For $k = 1$, $G(\Lambda) = 1/12$, and, hence, uniform scalar quantizers are indeed asymptotically optimal.

Analogous to the scalar case, if one can find the asymptotically optimal performance for a sequence of vector quantizers with increasing rate for *any* k-dimensional density for which the theorem is true (e.g., the uniform pdf on a hypercube), then this would yield the value for $b(2, k)$ as it does in the scalar case. To

this day, however, $b(2, k)$ is known only for $k = 1$ and the limiting case of $k \to \infty$. (The corresponding constant for high-rate fixed-rate coding is known for $k = 2$ as well.)

In 1979, Gersho [12] provided a heuristic development of Zador's results and demonstrated that based on a still unproved conjecture regarding the asymptotically optimal quantizer cell shapes, an optimal lattice or tessellating quantizer followed by an optimal lossless code provides an asymptotically optimal entropy-constrained quantizer in the limit of high rate. This is a natural generalization of the one-dimensional optimality of scalar quantization and optimal entropy coding, but the result has never been rigorously demonstrated and existing derivations depend strongly on Gersho's conjecture. It is widely conjectured that lattice or tessellating quantizers followed by optimal entropy codes are asymptotically optimal, but this result has not yet been proved.

In order to state two final preliminary results, we introduce the following notation:

$$\begin{aligned}\theta(f, \lambda, \alpha, \beta, \ell) &= \theta(f, \lambda, Q)\\ &\triangleq \frac{D_f(Q)}{\lambda} + R_f(Q) + \frac{k}{2}\ln\lambda - h(f)\\ &= \frac{E_f d(X, q(X))}{\lambda} + E_f \ell(\alpha(X))\\ &\quad + \frac{k}{2}\ln\lambda - h(f) \end{aligned} \tag{12}$$

so that the theorem states that under suitable conditions

$$\lim_{\lambda\to 0}\inf_{Q}\theta(f, \lambda, Q) = \theta_k. \tag{13}$$

If one or more of the components is optimized, then it is dropped from the argument of θ, e.g.,

$$\begin{aligned}\theta(f, \lambda, \alpha, \beta) &= \inf_{\ell}\theta(f, \lambda, \alpha, \beta, \ell)\\ &= \frac{D_f(\alpha, \beta)}{\lambda} + H_f(\alpha) + \frac{k}{2}\ln\lambda - h(f) \end{aligned} \tag{14}$$

$$\theta(f, \lambda) = \inf_{\alpha, \beta, \ell}\theta(f, \lambda, \alpha, \beta, \ell). \tag{15}$$

With this notation, the theorem statement can be simplified to

$$\lim_{\lambda\to 0}\theta(f, \lambda) = \theta_k. \tag{16}$$

The theorem guarantees that if a pdf f satisfies the conditions of the theorem, then there is an *asymptotically optimal* sequence of quantizers q_n for f in the sense that for any decreasing sequence λ_n converging to 0 there exists a sequence of quantizers q_n such that

$$\lim_{n\to\infty}\theta(f, \lambda_n, q_n) = \theta_k. \tag{17}$$

Disjoint Mixtures

A mixture source is a random pair $\{X, Z\}$, where Z is a discrete random variable with probability mass function (pmf) $w_m = P(Z = m)$, $m = 1, 2, \ldots$, and conditional pdfs $f_{X|Z}(x|m) = f_m(x)$ such that $P_{f_m}(\Omega_m) = 1$ for some $\Omega_m \in \mathcal{B}$, $m = 1, 2, \ldots$. The pdf for X is given by

$$f(x) = f_X(x) = \sum_m w_m f_m(x).$$

In the special case where the Ω_m are disjoint, the mixture is said to be *orthogonal* or *disjoint*. Thus, for example, given any pdf f and a partition $\mathcal{S} = \{S_m\}$, there is induced a disjoint mixture $\{w_m, f_m\}$ with $w_m = \int_{S_m} f(x)\,dx$, $f_m(x) = f(x)/w_m$ for $x \in S_m$, and 0 otherwise.

Suppose that f is a disjoint mixture and that for each f_m we have a quantizer q_m defined on Ω_m, i.e., an encoder α_m: $\Omega_m \to \mathcal{I}$, a partition of Ω_m, $\{S_{m,i};\ i = 1, 2, \ldots\}$, and a decoder $\beta_m : \mathcal{I} \to \mathcal{C}_m$. The component quantizers $\{q_m\}$ together imply an overall composite quantizer q with an encoder α that maps x into a pair (m, i) if $x \in \Omega_m$ and $\alpha_m(x) = i$, a partition of Ω, $\{S_{m,i};\ i = 1, 2, \ldots, m = 1, 2, \ldots\}$, and a decoder β that maps (m, i) into $\beta_m(i)$

$$q(x) = \sum_m q_m(x)1_{\Omega_m}(x) \tag{18}$$

where $1_A(x)$ denotes the indicator function of $A \subset \Omega$. If each component quantizer Q_m has a length function ℓ_m and if there is then an admissible length function L for the component indexes m; $m = 1, 2, \ldots$, then a valid length function for the composite quantizer ℓ is given by

$$\ell(m, i) = L(m) + \ell_m(i). \tag{19}$$

The length function for coding the component in effect is optimized by choosing $L(m) = -\ln P_f(\Omega_m)$.

Conversely to constructing an overall quantizer from a collection of component quantizers, an overall quantizer Q with encoder α: $\Omega \to \mathcal{I}$ can be applied to every component in the mixture. A little manipulation shows that (see, e.g., [20])

$$h(f) = \sum_m w_m h(f_m) + H(Z) \tag{20}$$

where the equality holds whenever at least two of the three quantities $h(f)$, $H(Z)$, and $\sum_m w_m h(f_m)$ are finite. Under these conditions, [20, Lemma 2] shows that

$$\theta(f, \lambda, \alpha, \beta) = \sum_m w_m \theta(f_m, \lambda, \alpha, \beta) - H(Z|q(X)) \tag{21}$$

where $H(Z|q(X))$ denotes the conditional entropy of Z given the quantizer output $q(X)$.

Relative Entropy

Given two probability measures P and G on $(\Omega, \mathcal{B})$ for which $P \ll G$ (i.e., P is absolutely continuous with respect to G) and a finite measurable partition $\mathcal{S} = \{S_i\}$, define the *relative entropy* of P with respect to G of the partition $\mathcal{S}$ as

$$H_{P\|G}(\mathcal{S}) = \sum_i P(S_i)\ln\frac{P(S_i)}{G(S_i)}$$

and the *relative entropy* of P with respect to G as

$$I(P\|G) = \sup_{\mathcal{S}} H_{P\|G}(\mathcal{S})$$

where the supremum is over all finite measurable partitions. The relative entropy is also known as the Kullback–Leibler number or Kullback–Leibler I-divergence or directed divergence or discrimination. The reader is referred to [25], [29], [7]–[9], and

[15] for thorough treatments of relative entropy and its properties. If the two measures are induced by pdfs f and g it can be shown that

$$I(P_f||P_g) = I(f||g) = \int dx f(x) \ln \frac{f(x)}{g(x)}$$

where we have abbreviated the notation to emphasize the dependence on the densities. We follow Csiszár's notation and use I for relative entropy. (Another common symbol is D for divergence, but that might cause confusion with distortion.)

III. Quantizer Mismatch

The principal result of this paper is the following high-rate variable-rate quantizer mismatch theorem.

Theorem 2 (The Mismatch Theorem): Suppose that a probability measure P_g on $\Re^k$ satisfies the conditions of Theorem 1 and has pdf g. Suppose that $Q_n = (q_n, \ell_n)$ is an asymptotically optimal sequence of variable-rate quantizers for P_g, where ℓ_n is the optimal length function for P_g and q_n. Suppose also that $P_f \ll P_g$ and that $dP_f/dP_g = f/g$ is bounded. Then

$$\lim_{n\to\infty} \frac{E_f d(X, q_n(X))}{\lambda_n} + E_f \ell_n(\alpha_n(X)) + \frac{k}{2} \ln \lambda_n = \theta_k - \int dx f(x) \ln g(x) \quad (22)$$

or, equivalently

$$\lim_{n\to\infty} \theta(f, \lambda_n, Q_n) = \theta_k + I(f||g). \quad (23)$$

The second form of the theorem provides a characterization of the *mismatch* resulting from applying an asymptotically optimal quantizer sequence for one pdf to another: the mismatch is exactly the relative entropy of the mismatched pdf to the design pdf, a continuous analog to the mismatch formula arising in noiseless coding. The result provides a new interpretation of relative entropy as a measure of mismatch for high-rate fixed-dimension lossy data compression. Relative entropy also arises in a somewhat similar way in the Shannon-type regime of asymptotically high dimensions and random codebook selection [10]. In that case, the relative entropy of the mismatched codebook distribution to the optimal codebook distribution is an upper bound on the loss due to mismatch.

The mismatch theorem can be derived based on Gersho's conjecture [17], that is, the result is consistent with that predicted by Gersho's conjecture. Unfortunately, however, as with other implications of Gersho's conjecture, this has not yet led to a proof of the theorem and the theorem is not an immediate consequence of previously known results.

The constraint that f/g be bounded is admittedly stronger than one would like as it eliminates many interesting cases. For example, for a scalar case, a Gaussian g and a Laplacian f will not meet the conditions, nor will two Gaussians with mismatched means. On the other hand, the corresponding results for fixed-rate coding essentially require the same assumption. The only known example of Bucklew's uniform integrability conditions [3] for the fixed-rate analog result is the same—that f/g be bounded. The derivation [17] based on Gersho's conjecture suggests that the theorem should hold more generally, specifically, that $I(f||g) < \infty$ should suffice, but thus far we have not succeeded in proving this the case and it remains an interesting open problem. In the particular case of a Gaussian model g and an unknown pdf f describing raster image intensities or speech samples produced by physical sensors, then f/g will be bounded due to the finite dynamic range of real sensors.

The special role of uniform scalar quantizers provides an illustration. If both f and g meet the requirements of Theorem 1, then the results of [28] imply that a sequence of uniform quantizers with an optimal entropy code for g will yield the same average squared error when applied to f with a rate increase of $H_{f||g}(\mathcal{S}_Q)$, where $\mathcal{S}_Q$ is the partition corresponding to the uniform quantizer. If the relative entropy $I(f||g)$ is finite, then $H_{f||g}(S_Q)$ will converge to $I(f||g)$ as the rate $R \to \infty$. Note that in this case, the conclusion of the mismatch theorem holds for a specific asymptotically optimal quantizer sequence without the requirement of bounded f/g.

The mismatch theorem provides the mismatched performance of *any* asymptotically optimal quantizer sequence. Even if lattice or tessellating quantizers were asymptotically optimal as suggested by results based on Gersho's conjecture, that would only be one particular scheme. For example, for an f with unbounded support but sub-Gaussian tails, the results of [4] imply that an optimal entropy-constrained vector quantization (ECVQ) for any value of λ has a *finite* number of codewords. Thus, there are asymptotically optimal sequences that are quite different from lattice quantizers.

The development of the mismatch result for ECVQ parallels that of Bucklew's fixed-rate result in that the beginning and core of the proof of the lemma of the next section provides what can be interpreted as a local form of Theorem 1, an entropy-constrained variation on Bucklew's Lemma 2. Although the statement of the result is analogous, our proof bears little resemblance to Bucklew's and is, in fact, considerably simpler. Our use of the result differs significantly from that of Bucklew. The subsequent section is devoted to the proof of the theorem. The remainder of the paper provides corollaries, examples, and an application.

The theorem is proved in a series of steps in the next three sections.

IV. Asymptotically Optimal Quantization

The first step is an extension of Theorem 1 involving only a single density. Following the notation of that theorem we assume that f is the source pdf and that an asymptotically optimal quantizer sequence is designed for f and we investigate properties of the sequence. More formally, recall that for any decreasing sequence λ_n converging to 0, Theorem 1 implies the existence of an asymptotically optimal sequence of encoders and decoders α_n, β_n, and the corresponding optimized length function

$$\ell_n^*(i) = -\ln P_f(\alpha_n(X) = i)$$

for which

$$\begin{aligned}\lim_{n\to\infty} &\theta(f, \lambda_n, \alpha_n, \beta_n)\\ &= \lim_{n\to\infty} \int_{\Re^k} dx f(x) \left(\frac{d(x, \beta_n(\alpha_n(x)))}{\lambda_n} + \ell_n^*(\alpha_n(x)) \right.\\ &\qquad \left. + \frac{k}{2} \ln \lambda_n + \ln f(x) \right)\\ &= \theta_k. \end{aligned} \tag{24}$$

Lemma 1: Suppose that $Q_n = (\alpha_n, \beta_n, \ell_n^*)$ is an asymptotically optimal sequence of variable-rate quantizers for P_f in the sense of (24), where ℓ_n^* is the optimal length function for P_f. Then, for every measurable set F

$$\lim_{n\to\infty} \int_F dx f(x) \left(\frac{d(x, \beta_n(\alpha_n(x)))}{\lambda_n} + \ell_n^*(\alpha_n(x)) \right. \left. + \frac{k}{2} \ln \lambda_n + \ln f(x) \right) = P_f(F)\theta_k. \tag{25}$$

Proof: If $P_f(F) = 0$ or 1, the claim is immediate, so assume that $1 > P_f(F) > 0$. The lemma can be stated simply by adopting Bucklew's notation. Define

$$M_f^n(F) \triangleq \int_F dx f(x) \left(\frac{d(x, \beta_n(\alpha_n(x)))}{\lambda_n} + \ell_n^*(\alpha_n(x)) \right. \left. + \frac{k}{2} \ln \lambda_n + \ln f(x) \right) \tag{26}$$

where ℓ_n^* is the optimal length function for α_n and β_n

$$\ell_n^*(i) = -\ln P_f(\alpha_n(X) = i) \tag{27}$$

and

$$M_f(F) = P_f(F)\theta_k$$

so that the claim of the lemma becomes

$$\lim_{n\to\infty} M_f^n(F) = M_f(F). \tag{28}$$

Note: Unlike Bucklew's case, we cannot say $M_f^n(F)$ is nonnegative for all events F so that we cannot argue it is a measure. We shall, however, find it useful in the next section to view it as a *signed measure*. The setwise limit $M_f(F)$ is, however, a measure.

By construction and Theorem 1 as $n \to \infty$

$$M_f^n(F) + M_f^n(F^c) = \theta(f, \lambda_n, \alpha_n, \beta_n) \to \theta_k. \tag{29}$$

It is convenient to rewrite these expressions in terms of the disjoint mixture obtained by restricting f to the partition $\{F, F^c\}$, where F is a fixed measurable set. Let $\{w_m, f_m;\ m = 1, 2\}$ be the induced disjoint mixture with $w_1 = P_f(F)$, $\Omega_1 = F$, $f_1(x) = f(x)/w_1$ for $x \in F$, and 0 otherwise, and similarly for $m = 2$. Let X have pdf f and define $Z = 1$ if $X \in F$ and 2 otherwise. Then

$$\begin{aligned} M_f^n(F) &= w_1 \int dx f_1(x) \left(\frac{d(x, \beta_n(\alpha_n(x)))}{\lambda_n} + \ell_n^*(\alpha_n(x)) \right.\\ &\qquad \left. + \frac{k}{2} \ln \lambda_n + \ln f_1(x) w_1 \right) \end{aligned}$$

and

$$\begin{aligned} M_f^n(F) &+ M_f^n(F^c)\\ &= \sum_{m=1}^{2} w_m \int dx f_m(x) \left(\frac{d(x, \beta_n(\alpha_n(x)))}{\lambda_n} + \ell_n^*(\alpha_n(x)) \right.\\ &\qquad \left. + \frac{k}{2} \ln \lambda + \ln f_m(x) w_m \right). \end{aligned}$$

Since

$$\lim_{n\to\infty} \theta(f, \lambda_n, q_n) = \theta_k \tag{30}$$

[20, Lemma 6] shows that

$$\lim_{n\to\infty} H(Z|q_n(X)) = 0$$

i.e., asymptotically, the quantized X must determine which component of the mixture is in effect. This and (21) imply that

$$\begin{aligned} \lim_{n\to\infty} \theta(f, \lambda_n, \alpha_n, \beta_n) &= \lim_{n\to\infty} \sum_{m=1}^{2} w_m \theta(f_m, \lambda_n, \alpha_n, \beta_n)\\ &= \theta_k. \end{aligned}$$

We now claim that each of the two component compression functions must individually converge to θ_k, not just the overall weighted average. Note that by Theorem 1 (see (13)) for $m = 1, 2$

$$\liminf_{n\to\infty} \theta(f_m, \lambda_n, \alpha_n, \beta_n) \geq \theta_k. \tag{31}$$

Now assume that

$$\limsup_{n\to\infty} \theta(f_1, \lambda_n, \alpha_n, \beta_n) > \theta_k.$$

Then, there is a subsequence n' such that

$$\lim_{n'\to\infty} \theta(f_1, \lambda_{n'}, \alpha_{n'}, \beta_{n'}) > \theta_k.$$

By assumption

$$\lim_{n'\to\infty} \theta(f, \lambda_{n'}, \alpha_{n'}, \beta_{n'}) = \theta_k$$

and

$$\begin{aligned} \theta(f, \lambda_{n'}, \alpha_{n'}, \beta_{n'}) &= w_1 \theta(f_1, \lambda_{n'}, \alpha_{n'}, \beta_{n'})\\ &\quad + w_2 \theta(f_2, \lambda_{n'}, \alpha_{n'}, \beta_{n'}) \end{aligned}$$

which implies that

$$\lim_{n'\to\infty} \theta(f_2, \lambda_{n'}, \alpha_{n'}, \beta_{n'}) < \theta_k$$

contradicting (31). Hence, $\limsup_n \theta(f_1, \lambda_n, \alpha_n, \beta_n) \le \theta_k$, and by symmetry we also have $\limsup_n \theta(f_2, \lambda_n, \alpha_n, \beta_n) \le \theta_k$. Thus, we have that for $m = 1, 2$

$$\lim_{n\to\infty} \int dx f_m(x) \left(\frac{d(x, \beta_n(\alpha_n(x)))}{\lambda_n} - \ln P_{f_m}(S_{n,\alpha_n(x)}) + \frac{k}{2}\ln\lambda + \ln f_m(x) \right) = \theta_k \quad (32)$$

where $\mathcal{S}_n = \{S_{n,i};\ i \in \mathcal{I}\}$ denotes the encoder partition corresponding to α_n.

To combine the two similar weighted sums, those for M_F^n and those for $\theta(f_m, \lambda_n, q_n)$, suppose that $\ell^*_{m,n}$ is the optimal length function for α_n and β_n using the pdf f_m, i.e.,

$$\ell^*_{m,n}(i) = -\ln P_{f_m}(S_{n,i}). \quad (33)$$

We can write

$$\begin{aligned} M_f^n(F) &= \int_F dx f(x) \left(\frac{d(x, \beta_n(\alpha_n(x)))}{\lambda_n} + \ell_n^*(\alpha_n(x)) + \frac{k}{2}\ln\lambda_n + \ln f(x) \right) \\ &= w_1 \int dx f_1(x) \left(\frac{d(x, \beta_n(\alpha_n(x)))}{\lambda_n} + \ell_n^*(\alpha_n(x)) + \frac{k}{2}\ln\lambda_n + \ln(f_1(x)w_1) \right) \\ &= w_1(\theta(f_1, \lambda_n, \alpha_n, \beta_n) + E_{f_1}(\ell_n^*(\alpha_n(X)) - \ell^*_{1,n}(\alpha_n(X))) + \ln w_1) \end{aligned} \quad (34)$$

and similarly

$$M_f^n(F^c) = w_2(\theta(f_2, \lambda_n, \alpha_n, \beta_n) + E_{f_2}(\ell_n^*(\alpha_n(X)) - \ell^*_{2,n}(\alpha_n(X))) + \ln w_2). \quad (35)$$

Since $\ell^*_{m,n}$ is the optimal length function for q_n

$$E_{f_m}[\ell_n^*(\alpha_n(X))] \ge E_{f_m}[\ell^*_{m,n}(\alpha_n(X)))]$$

and, hence,

$$M_f^n(F) \ge w_1\theta(f_1, \lambda_n, \alpha_n, \beta_n) + w_1 \ln w_1 \quad (36)$$

$$M_f^n(F^c) \ge w_2\theta(f_2, \lambda_n, \alpha_n, \beta_n) + w_2 \ln w_2. \quad (37)$$

The leftmost term on the right-hand side of (34) has already been shown to converge to $w_1\theta_k$ as $n \to \infty$, so the lemma will be proved if the remaining term

$$w_1 E_{f_1}(\ell_n^*(\alpha_n(X)) - \ell^*_{1,n}(\alpha_n(X))) + w_1 \ln w_1$$

is shown to converge to 0. As a first step toward this demonstration, observe that

$$\begin{aligned} &w_m E_{f_m}(\ell_n^*(\alpha_n(X)) - \ell^*_{m,n}(\alpha_n(X))) + w_m \ln w_m \\ &= w_m \sum_i P_{f_m}(S_{n,i}) \ln \frac{P_{f_m}(S_{n,i})P_f(\Omega_m)}{P_f(S_{n,i})} \\ &= w_m \sum_i \frac{P_f(\Omega_m \cap S_{n,i})}{P_f(\Omega_m)} \ln \frac{P_f(\Omega_m \cap S_{n,i})}{P_f(S_{n,i})} \\ &= w_m \sum_i P_f(S_{n,i}|\Omega_m) \ln P_f(\Omega_m|S_{n,i}) \end{aligned}$$

or, equivalently

$$\begin{aligned} &w_m E_{f_m}(\ell_n^*(\alpha_n(X)) + \ln w_m - \ell^*_{1,n}(\alpha_n(X))) \\ &= w_m \sum_i \Pr(\alpha_n(X) = i | Z = m) \ln \Pr(Z = m | \alpha_n(X) = i) \end{aligned}$$

which is, obviously, negative for both $m = 1$ and 2. Hence, these terms for $m = 1$ and 2 will go to zero if and only if the sum of the two terms

$$\sum_{m=1}^{2} \left(w_m E_{f_m}(\ell_n^*(\alpha_n(X)) + \ln w_m - \ell^*_{1,n}(\alpha_n(X))) \right) = -H(Z|q_n(X))$$

tends to zero as $n \to \infty$, which has already been seen to be the case. □

In Bucklew's development, the analog to the previous lemma provides a key step in the proof of the mismatch theorem. Unfortunately, however, Bucklew's approach cannot be used directly since, in our case, the M_g^n are not nonnegative and hence the argument of the integral defining these terms cannot be assumed to be nonnegative and hence a probability density.

V. Signed Measures

We now change notation somewhat in order to reflect the fact that in the mismatch theorem there are two densities of interest, a density g for which we have designed an optimal sequence of codes and a density f to which we will apply the sequence of codes. Toward this end, replace the f in Theorem 1 and Lemma 1 by g and define the set function $M_g^n(F)$ for g as in (26). For convenience, we make the additional definition

$$\begin{aligned} \mu_n(F) &= M_g^n(F) \\ &\triangleq \int_F dx g(x) \left(\frac{d(x, \beta_n(\alpha_n(x)))}{\lambda_n} + \ell_n^*(\alpha_n(x)) + \frac{k}{2}\ln\lambda_n + \ln g(x) \right). \end{aligned} \quad (38)$$

The set function μ_n is a signed measure (see, e.g., Doob [11]). From Lemma 1 we know that $\mu_n(F)$ converges to $\mu(F) = M_g(F) = \theta_k P_g(F)$ for all measurable sets F. We now explore the consequences of this convergence.

Given a signed measure μ, for any measurable set F define the positive variation $\mu^+(F) = \sup_{G \subset F} \mu(G)$, the negative variation $\mu^-(F) = -\inf_{G \subset F} \mu(G)$, and the total variation $|\mu|(F) = \mu^+(F) + \mu^-(F)$. The space $\mathcal{M}$ of all finite-signed measures is a normed space with norm

$$\|\mu\| = |\mu|(\Re^k).$$

If $\mu = \mu^+$, then μ is called a *positive measure*.

A signed measure is *finite* if $\|\mu\| < \infty$. The Jordan decomposition states that $\mu = \mu^+ - \mu^-$, which represents the signed measure as the difference of two positive measures.

For all measurable sets F, $\mu_n(F) < \infty$, and

$$\lim_{n\to\infty} \mu_n(F) = P_g(F)\theta_k < \infty$$

and hence also

$$\lim_{n\to\infty} |\mu_n(F)| = P_g(F)\theta_k < \infty.$$

Thus, from the discussion following Doob's theorem [11, Theorem IX.9], it follows that

$$\sup_n \|\mu_n\| < \infty. \tag{39}$$

From Lemma 1, it follows that for any simple function ϕ

$$\lim_{n\to\infty} \int \phi \, d\mu_n = \int \phi \, d\mu. \tag{40}$$

We now show that this limit will also hold for any bounded nonnegative function. Suppose that ϕ is such a function and for simplicity assume that $0 \le \phi < 1$. For a fixed positive integer k define the measurable sets

$$G_i = \left\{ x \colon \phi(x) \in \left[\frac{i}{k}, \frac{i+1}{k} \right) \right\}, \qquad i = 0, 1, \ldots, k-1.$$

From the Jordan decomposition we can write $\mu_n(F) = \mu_n^+(F) - \mu_n^-(F)$ and hence,

$$\begin{aligned}
\int \phi \, d\mu_n &= \int \phi \, d\mu_n^+ - \int \phi \, d\mu_n^- \\
&= \sum_{i=0}^{k-1} \int_{G_i} d\mu_n^+ \phi - \sum_{i=0}^{k-1} \int_{G_i} d\mu_n^- \phi \\
&\le \sum_{i=0}^{k-1} \frac{i+1}{k} \mu_n^+(G_i) - \sum_{i=0}^{k-1} \frac{i}{k} \mu_n^-(G_i) \\
&= \sum_{i=0}^{k-1} \frac{i}{k} \mu_n(G_i) + \frac{1}{k} \sum_{i=0}^{k-1} \mu_n^+(G_i) \\
&= \sum_{i=0}^{k-1} \frac{i}{k} \mu_n(G_i) + \frac{1}{k} \mu_n^+(\Re^k) \\
&\le \sum_{i=0}^{k-1} \frac{i}{k} \mu_n(G_i) + \frac{1}{k} \|\mu_n\| \\
&\le \sum_{i=0}^{k-1} \frac{i}{k} \mu_n(G_i) + \frac{1}{k} \sup_n \|\mu_n\|.
\end{aligned} \tag{41}$$

Since we know the limit for simple functions

$$\limsup_{n\to\infty} \int \phi \, d\mu_n \le \sum_{i=0}^{k-1} \frac{i}{k} \mu(G_i) + \frac{1}{k} \sup_n \|\mu_n\|.$$

In a similar manner

$$\begin{aligned}
\int \phi \, d\mu &= \int \phi \, d\mu_n^+ - \int \phi \, d\mu_n^- \\
&= \sum_{i=0}^{k-1} \int_{G_i} d\mu^+ \phi - \sum_{i=0}^{k-1} \int_{G_i} d\mu^- \phi \\
&\ge \sum_{i=0}^{k-1} \frac{i}{k} \mu^+(G_i) - \sum_{i=0}^{k-1} \frac{i+1}{k} \mu^-(G_i) \\
&= \sum_{i=0}^{k-1} \frac{i}{k} \mu(G_i) - \frac{1}{k} \sum_{i=0}^{k-1} \mu^-(G_i) \\
&= \sum_{i=0}^{k-1} \frac{i}{k} \mu(G_i) - \frac{1}{k} \mu^-(\Re^k) \\
&\ge \sum_{i=0}^{k-1} \frac{i}{k} \mu(G_i) - \frac{1}{k} \|\mu\|.
\end{aligned} \tag{42}$$

Combining the previous two inequalities we have that

$$\begin{aligned}
\limsup_{n\to\infty} \int \phi \, d\mu_n &\le \sum_{i=0}^{k-1} \frac{i}{k} \mu(G_i) + \frac{1}{k} \sup_n \|\mu_n| \\
&\le \int \phi \, d\mu + \frac{1}{k} \|\mu\| + \frac{1}{k} \sup_n \|\mu_n\|.
\end{aligned}$$

Since the two rightmost terms can be made arbitrarily small by choosing k sufficiently large

$$\limsup_{n\to\infty} \int \phi \, d\mu_n \le \int \phi \, d\mu. \tag{43}$$

Similarly, repeating the steps in (41) and (42) but exchanging the role of μ and μ_n, we obtain

$$\int \phi \, d\mu_n \ge \sum_{i=0}^{k-1} \frac{i}{k} \mu_n(G_i) - \frac{1}{k} \sup_n \|\mu_n\|$$

and

$$\int \phi \, d\mu \le \sum_{i=0}^{k-1} \frac{i}{k} \mu(G_i) + \frac{1}{k} \|\mu\|$$

so that

$$\liminf_{n\to\infty} \int \phi \, d\mu_n \ge \int \phi \, d\mu - \frac{1}{k} \sup_n \|\mu_n\| - \frac{1}{k} \|\mu\|.$$

Since k can be made arbitrarily large, indeed (40) holds as claimed for all bounded nonnegative functions.

VI. Proof of the Mismatch Theorem

Since the Radon–Nikodym derivative $\phi = f/g$ is assumed to be bounded

$$\lim_{n\to\infty} \int \frac{f}{g} \, d\mu_n = \int \frac{f}{g} \, d\mu \tag{44}$$

and the two sides of the equation evaluate as

$$\begin{aligned}
&\lim_{n\to\infty} \int \frac{f}{g} \, d\mu_n \\
&\quad = \lim_{n\to\infty} \int dx g(x) \left(\frac{d(x, \beta_n(\alpha_n(x)))}{\lambda_n} + \ell_n^*(\alpha_n(x)) \right. \\
&\qquad \left. + \frac{k}{2} \ln \lambda_n + \ln g(x) \right) \frac{f(x)}{g(x)} \\
&\quad = \lim_{n\to\infty} \int dx f(x) \left(\frac{d(x, \beta_n(\alpha_n(x)))}{\lambda_n} + \ell_n^*(\alpha_n(x)) \right. \\
&\qquad \left. + \frac{k}{2} \ln \lambda_n + \ln g(x) \right)
\end{aligned}$$

$$\begin{aligned}
\int \frac{f}{g} \, d\mu &= \theta_k \int dx g(x) \frac{f(x)}{g(x)} \\
&= \theta_k
\end{aligned}$$

which completes the proof of the theorem. □

VII. High-Rate Universal Codes

The mismatch theorem shows the asymptotic performance that is lost when an asymptotically optimal sequence of quantizers Q_n designed for a pdf g is applied to a pdf f and that this loss is just the relative entropy $I(f\|g)$. In this section, we see that this performance loss can be eliminated by modifying only the length function to match the pdf f. This implies that the asymptotically optimal sequence of reproduction codebooks for the design pdf g remains asymptotically optimal for any f

meeting the conditions of the theorem. Thus, for example, if f has bounded support, one could design an asymptotically optimal sequence of codes for a uniform pdf on the support set and it would also be optimal for f. If f has unbounded support, one could design an asymptotically optimal quantizer sequence for a Gaussian pdf and its reproduction codebooks would be asymptotically optimal for f.

Corollary 1: Suppose that $Q_n = (q_n, \ell_n)$ is a sequence of variable-rate quantizers that is asymptotically optimal for a pdf g in the sense that

$$\lim_{n\to\infty} \theta(g, \lambda_n, Q_n) = \theta_k$$

for some decreasing sequence λ_n converging to 0. Assume also that f is a pdf that meets the condition of the mismatch theorem and that $h(f) > -\infty$. Define ℓ'_n to be the optimal length function for q_n and P_f. Then $Q'_n = (q_n, \ell'_n)$ is asymptotically optimal for P_f, i.e.,

$$\lim_{n\to\infty} \theta(f, \lambda_n, Q'_n) = \theta_k. \tag{45}$$

Proof: Since Q_n and Q'_n share the same encoder α_n and decoder β_n and differ only in their length functions, we have from (12) that

$$\begin{aligned}\theta(f, \lambda_n, Q'_n) &= \theta(f, \lambda_n, Q_n) - (E_f \ell(\alpha(X)) - E_f \ell'(\alpha(X))) \\ &= \theta(f, \lambda_n, Q_n) - \left(\sum_i P_f(S_{n,i}) \ln \frac{P_f(S_{n,i})}{P_g(S_{n,i})}\right)\end{aligned}$$

where we have plugged in the definitions for ℓ and ℓ' as the optimal length function for g and f, respectively, and where $\{S_{n,i}\}$ is the partition corresponding to α_n

$$S_{n,i} = \{x: \alpha_n(x) = i\}.$$

We have immediately from Theorem 1

$$\begin{aligned}\liminf_{n\to\infty} \theta(f, \lambda_n, Q'_n) &\geq \liminf_{n\to\infty} \inf_Q \theta(f, \lambda_n, Q) \\ &= \theta_k. \end{aligned} \tag{46}$$

For the other direction we have, using the mismatch theorem, that

$$\begin{aligned}&\limsup_{n\to\infty} \theta(f, \lambda_n, Q'_n) \\ &\quad = \theta_k + I(f\|g) - \liminf_{n\to\infty} \sum_i P_f(S_{n,i}) \ln \frac{P_f(S_{n,i})}{P_g(S_{n,i})}. \end{aligned} \tag{47}$$

As a preliminary to considering the rightmost term, define the discrete distribution P_f^n by $P_f^n(y_{n,i}) = P_f(S_{n,i})$, where $y_{n,i} = \beta_n(i)$, i.e., for any measurable set F

$$P_f^n(F) = \sum_{i:\, y_{n,i} \in F} P_f^n(y_{n,i})$$

and P_f^n is the distribution for the random vector $q_n(X)$ when X is described by the pdf f. Similarly define the discrete distribution P_g^n by $P_g^n(y_{n,i}) = P_g(S_{n,i})$. With this notation, (47) becomes

$$\limsup_{n\to\infty} \theta(f, \lambda_n, Q'_n) = \theta_k + I(f\|g) - \liminf_{n\to\infty} I(P_f^n\|P_g^n). \tag{48}$$

It is easy to see that

$$\lim_{n\to\infty} \theta(g, \lambda_n, Q_n) = \theta_k$$

implies

$$\lim_{n\to\infty} D_g(q_n) = \lim_{n\to\infty} \int dx\, g(x)\|x - q_n(x))\|^2 = 0 \tag{49}$$

(see [20, eq. (24) in the proof of Lemma 6]), i.e., $q_n(X)$ converges to X in mean square (here X has pdf g). This implies that $P_g^n \to P_g$ in the sense of weak convergence (see, e.g,, [31, Theorem 4.2]).

Furthermore, since by assumption there is a finite M such that $f(x)/g(x) \leq M$

$$\begin{aligned}\lim_{n\to\infty} D_f(q_n) &= \lim_{n\to\infty} \int dx\, g(x) \frac{f(x)}{g(x)} \|x - q_n(x))\|^2 \\ &\leq M \lim_{n\to\infty} D_g(q_n) \\ &= 0\end{aligned}$$

and, hence, by the same argument $P_f^n \to P_f$ (weak convergence). From [9], relative entropy is lower semicontinuous with respect to weak convergence of distributions so that

$$\liminf_{n\to\infty} I(P_f^n\|P_g^n) \geq I(f\|g) \tag{50}$$

which with (47) yields

$$\limsup_{n\to\infty} \theta(f, \lambda_n, Q'_n) \leq \theta_k$$

which completes the proof. □

The preceding proof contains an interesting property of asymptotically optimal quantization. Since

$$I(P_f^n\|P_g^n) = H_{f\|g}(\mathcal{S}_n)$$

and $\mathcal{S}_n = \{S_{n,i}\}$ is a measurable partition

$$I(P_f^n\|P_g^n) \leq I(f\|g)$$

which, with (50), implies that

$$\lim_{n\to\infty} H_{f\|g}(\mathcal{S}_n) = I(f\|g). \tag{51}$$

This result would be immediate if the sequence of partitions $\mathcal{S}_n$ asymptotically generated the sigma field $\mathcal{B}$ (see, e.g., [29], [15]). This result shows that the partitions corresponding to an asymptotically optimal sequence of quantizers have the same property even though in general they do not generate the underlying sigma field.

An additional observation on the corollary is that although the length function (and, hence, the lossless component) of the quantizer has been matched to the true source, the encoder has not been optimized for the new length function. Thus, there remains a mismatch in the code sequence, which nonetheless is asymptotically optimal.

VIII. Examples

As examples and applications, we first consider two important special cases: a uniform pdf over a bounded subset of $\Re^k$ and a Gaussian pdf over the entire space $\Re^k$. The examples provide both interesting similarities and important differences which suggest specific applications and future exploration. Both examples yield reasonably simple formulas when used as the design pdfs for quantizers, but applied to a different pdf. Both examples represent "worst cases" for quantization, so that the resulting design provides a robust quantizer sequence for other pdfs satisfying the conditions of the mismatch theorem, where here "robust" is in the sense of Sakrison [30] and Lapidoth [26]:

a code is robust if it yields predictable, if suboptimal, performance on a source with only partially known statistics. The single pdf worst case design is then extended to mixtures of uniform or Gaussian by effectively quantizing the space of pdfs using Lloyd clustering based on relative entropy as a measure of distortion between models.

Uniform Codes

Suppose that g is a uniform pdf on a bounded measurable set G with positive Lebesgue measure so that

$$g(x) = \begin{cases} \frac{1}{V(G)}, & x \in G \\ 0, & \text{otherwise} \end{cases}$$

and hence

$$h(g) = \ln V(G).$$

Of all pdfs having G as a support set, it is well known that the uniform pdf results in the largest differential entropy (see, e.g., [6]).

Since the uniform pdf maximizes the differential entropy, it is the worst case in the sense of having the largest possible asymptotically optimal high-rate performance $\theta_k + h(g)$ for any pdf g with support G. Any bounded pdf f with support G meets the conditions of the mismatch theorem and, hence, if a sequence $\{\lambda_n, Q_n\}$ is asymptotically optimal for g

$$\begin{aligned} \lim_{n\to\infty} \theta(f, \lambda_n, Q_n) &= \theta_k + I(f\|g) \\ &= \theta_k + \ln V(G) - h(f). \end{aligned} \tag{52}$$

Thus, the code sequence designed for the uniform pdf g will have asymptotic performance when applied to f that is greater than the optimal asymptotic performance for f and this performance mismatch is $I(f\|g) = \ln V(G) - h(f)$. This implies that the code is robust as we next illustrate using the traditional Zador/Gersho non-Lagrangian argument.

The mismatch theorem and the correspondence (11) between the Lagrangian and traditional formulations implies that given a fixed large rate R and a quantizer Q_R that is optimized for g but used for f, the resulting average distortion is approximately

$$D(Q_R) \approx b(2, k)2^{-(2/k)R}2^{(2/k)h(f)}2^{(2/k)I(f\|g)}. \tag{53}$$

For a uniform pdf g

$$I(f\|g) = h(g) - h(f) \tag{54}$$

and hence,

$$D(Q_R) \approx b(2, k)2^{-(2/k)R}2^{(2/k)h(g)} \tag{55}$$

which is the best asymptotic performance of an ECVQ at rate R for g. So these are indeed robust quantizers in the Lapidoth sense.

In the special case where the support set G is the unit k-dimensional cube, the mismatch is simply $-h(f)$ (the divergence inequality implies that the differential entropy $h(f)$ is necessarily nonpostive in this case). Here, the pdf g is exactly that used in the definition of θ_k.

One important aspect of pdfs with bounded support is that the optimal codes exist and require only a finite number of reproduction levels [4], [23].

Gaussian Codes

Consider a Gaussian pdf

$$\begin{aligned} g(x) &= \mathcal{N}(x, \mu, K) \\ &= \frac{1}{(2\pi)^{\frac{k}{2}}|K|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(x-\mu)^t K^{-1}(x-\mu)\right) \end{aligned}$$

where $\mu = EX$, $K = E[(X - EX)(X - EX)^t]$ is the $k \times k$ covariance matrix, and $|K|$ the determinant of k. Assume that the covariance is nonsingular. In this case, the differential entropy is well known to be

$$h(g) = -\int dx f(x) \ln f(x) = \frac{1}{2}\ln(2\pi e)^k |K| \tag{56}$$

and it is well known that this differential entropy is the maximum possible over all pdfs corresponding to random vectors with covariance K (see, e.g., [6]). This, in turn, implies that if a sequence $\{\lambda_n, Q_n\}$ is asymptotically optimal for g, then for any pdf f with covariance K for which f/g is bounded, the asymptotic performance of this sequence is

$$\begin{aligned} \lim_{n\to\infty} \theta(f, \lambda_n, Q_n) &= \theta_k + I(f\|g) \\ &= \theta_k + \frac{1}{2}\ln(2\pi e)^k |K| - h(f). \end{aligned} \tag{57}$$

In this case, if all that is known about the pdf f is its covariance, then, designing a code for a Gaussian pdf g with the same covariance will, provided f/g is bounded, result in a code whose performance is $I(f\|g)$ worse than it would have been if the true pdf had been used to design the code. This code is robust since, as in the uniform example, (53)–(55) hold. This provides a high rate analog to the Shannon rate-distortion results of Sakrison [30] and Lapidoth [26] that an approximately optimal code designed for a large dimensional independent and identically distributed (i.i.d.) Gaussian vector will yield roughly the same performance on any other i.i.d. vector with the same mean and covariance. Here, high rate replaces the assumptions of memorylessness and large dimension.

Instead of knowing the full covariance

$$K = \{K(i, j);\ i = 0, \ldots, k-1, j = 0, \ldots, k-1\}$$

one might know only a partial covariance

$$K_{\mathcal{N}} = \{K(i, j);\ (i, j) \in \mathcal{N}\}$$

e.g., the covariance for small lags or in some band of the covariance matrix. In this case, the worst case pdf from a high-rate quantization perspective will be the worst case Gaussian pdf consistent with the known constraints, which is a Gaussian pdf with the covariance $\arg\max_{\Lambda:\ \Lambda_{\mathcal{N}} = K_{\mathcal{N}}} |\Lambda|$ if such a "maximum determinant" extension exists. This optimization problem is the well-known MAXDET problem for which much theory and efficient algorithms exist [32]. This case is of particular interest when the covariance is being estimated based on observed data and one can only trust a limited number of the covariance values, e.g., those of nearby pixels in an image. This provides a robust high-rate coding result for the case of partially known covariance, provided the partial covariance has a maximum determinant (or "maximum entropy") extension.

Composite Codes

A problem with choosing a worst case pdf to provide a robust quantizer sequence subject to some assumed constraint (e.g.,

known support or covariance structure) is that it can be too conservative. For example, fitting a single Gaussian pdf to a 100-ms chunk of sampled speech for the purposes of code design is well known to produce an overly conservative code, one that does not perform as well as a code which fits codes more customized to local behavior. One might instead use a collection of Gaussian models instead of a single model. Each model in the collection could yield a code that was robust for some conditional behavior of the source such as a conditional covariance structure. This approach is implicit in traditional linear prediction coding (LPC) speech coders, which can be interpreted as fitting Gaussian pdfs to local second-order behavior [21]. This idea provides an extension of the uniform and Gaussian mismatch examples to piecewise-uniform models and Gauss mixture models.

As before let f be the "true" pdf and suppose that Ω_f is its support (which might be all of $\Re^k$). Assume that $\mathcal{S}=\{S_m;\ m\in\mathcal{J}\}$, where $\mathcal{J}=\{1,\ \ldots,\ M\}$, is a finite partition of Ω_f and that $P_f(S_m)>0$ for all m. Assume also that we have a collection of model pdfs $\{g_m;\ m\in\mathcal{J}\}$ on $\Re^k$. The two examples of interest here will be uniform pdfs with bounded support and Gaussian pdfs. We assume further that we have an asymptotically optimal sequence of quantizers for each of the "design" pdfs g_m, that is, for a common decreasing sequence $\lambda_n \to 0$ we have for each m a quantizer sequence $Q_{m,n};\ n=1,2,\ldots$ for which $\lim_{n\to\infty}\theta(g_m,\lambda_n,Q_{m,n})=\theta_k$.

Let $Q_n=(\alpha_n,\beta_n,\ell_n)$ be the composite quantizer constructed from the $Q_{m,n}=(\alpha_{m,n},\beta_{m,n},\ell_{m,n})$, the partition $\mathcal{S}$, and a component length function L, that is,

$$\begin{aligned}\alpha_n(x)&=(m,\alpha_{m,n}(x)),\qquad \text{if } x\in S_m\\ \beta_n(m,i)&=\beta_{m,n}(i)\\ \ell_n(m,i)&=L(m)+\ell_{m,i}(i).\end{aligned}$$

Consider the performance resulting when the composite quantizer Q_n is applied to the pdf f. Letting $w_m=P_f(S_m)$ and $f_m(x)=f(x)/w_m$ if $x\in S_m$ and 0 otherwise and using (12) and (20) yields

$$\begin{aligned}&\theta(f,\lambda_n,Q_n)\\ &\quad=\frac{E_f d(X,q_n(X))}{\lambda_n}+E_f\ell_n(\alpha_n(X))+\frac{k}{2}\ln\lambda_n-h(f)\\ &\quad=\sum_m w_m\left(\frac{E_{f_m}d(X,q_n(X))}{\lambda_n}+E_{f_m}\ell_n(\alpha_n(X))\right.\\ &\qquad\qquad\left.+\frac{k}{2}\ln\lambda_n-h(f_m)\right)-H(Z)\end{aligned}$$

where Z has distribution

$$\Pr(Z=m)=w_m,\qquad \text{for } m=1,\ \ldots,\ M.$$

Also, by construction, from (19) the length function $\ell_n(m,i)=L(m)+\ell_{n,m}(i)$ and with the optimal choice of $L(m)=-\ln w_m$, the average code length of the composite quantizer is $EL(Z)+E_f\ell_{m,n}(\alpha_{m,n}(X))$. With this choice, we have for the composite quantizer Q_n that

$$\begin{aligned}&\theta(f,\lambda_n,Q_n)\\ &\quad=\sum_m w_m\left(\frac{E_{f_m}d(X,q_{m,n}(X))}{\lambda_n}+E_{f_m}\ell_{m,n}(\alpha_{m,n}(X))\right.\\ &\qquad\qquad\left.+\frac{k}{2}\ln\lambda_n-h(f_m)\right)+EL(Z)-H(Z).\end{aligned}$$

We assume the optimal choice of $L(m)=-\ln w_m$ so that

$$\begin{aligned}&\theta(f,\lambda_n,Q_n)\\ &\quad=\sum_m w_m\left(\frac{E_{f_m}d(X,q_{m,n}(X))}{\lambda}+E_{f_m}\ell_{m,n}(\alpha_{m,n}(X))\right.\\ &\qquad\qquad\left.+\frac{k}{2}\ln\lambda-h(f_m)\right)\\ &\quad=\sum_m w_m\theta(f_m,\lambda_n,Q_{m,n}).\end{aligned}$$

If f/g_m is bounded for each $m=1,\ \ldots,\ M$, then we can apply the mismatch theorem to each component to obtain the asymptotic high-rate performance

$$\lim_{n\to\infty}\theta(f,\lambda_n,Q_n)=\theta_k+\sum_m w_m I(f_m\|g_m).\tag{58}$$

This equation can be viewed as an extension of the mismatch theorem to composite quantizers. To extend the previous examples, recall that there the idea was to design a code for the "worst case" source given some constraint on f and then show that the resulting code applied to an unknown source with the given constraint would yield a known, if suboptimal, performance. Now the strategy is to divide and conquer: suppose that instead of a single uniform (or Gaussian) worst case, we are allowed to find a collection $\mathcal{G}=\{g_m;\ m\in\mathcal{J}\}$ of pdfs from an allowed collection $\mathcal{M}$ of uniform (or Gaussian) pdfs and a partition $\mathcal{S}=\{S_m;\ m\in\mathcal{J}\}$ of $\Re^k$ for use in a composite quantizer. What is the best way to do so? Specifically, for a fixed pdf f and model class $\mathcal{M}$, find a partition $\mathcal{S}$ with M elements and model codebook $\mathcal{G}$ which minimizes the mismatch

$$\overline{I}_f\triangleq\inf_{\mathcal{S},\mathcal{G}}\overline{I}_f(\mathcal{S},\mathcal{G})$$

where

$$\overline{I}_f(\mathcal{S},\mathcal{G})=\sum_m P_f(S_m)I(f_m\|g_m).$$

This minimization can be solved by clustering and, in fact, posed as a quantization problem with an encoder a: $\Re^k\to\mathcal{J}$ described by the partition $\mathcal{S}=\{S_m\}$ by $a(x)=m$ if $x\in S_m$, $m\in\mathcal{J}$, and a decoder b: $\mathcal{J}\to\mathcal{M}$ defined by $b(m)=g_m$.

The Lloyd decoder optimization is obvious in this context, given an encoder index m corresponding to encoder cell S_m, the best possible g_m is

$$g_m=\operatorname*{argmin}_{g\in\mathcal{M}}I(f_m\|g)$$

if the minimum exists, as will shortly be seen to be the case for both uniform and Gaussian model spaces. If the optimum decoder is assumed, the minimum mismatch problem becomes

$$\overline{I}_f=\inf_{\mathcal{S}}\sum_m P_f(S_m)\min_{g\in\mathcal{M}}I(f_m\|g).$$

To describe a quantizer encoder requires a distortion measure which describes the distortion, say $d_I(x,m)$, between an input vector $x\in\Re^k$ and an encoder output The average distortion with respect to the encoder should yield the mismatch, which we are attempting to minimize. A candidate distortion which will be shown to accomplish the desired goal is

$$d_I(x,m)=\ln\frac{f(x)}{g_m(x)}+L(m)$$

where L is an admissible length function which can be optimized along with the encoder and decoder. The first term involves only the shape of the model pdf and it has been used in clustering with the name of a "maximum-likelihood" (ML) distortion since minimizing this distortion over m for a given x is equivalent to choosing the ML estimate for m assuming the vector was produced by one of the models g_m [1], [22]. d_I is not a distortion in the strict sense since it need not be nonnegative, but its average with respect to f is nonnegative from the divergence inequality.

Given such a distortion measure is specified, the optimal encoder is a minimum distortion encoder and hence for a given decoder codebook $\mathcal{G}$

$$\begin{aligned} a(x) &= \operatorname*{argmin}_m d_I(x, m) \\ &= \operatorname*{argmin}_m \left(\ln \frac{f(x)}{g_m(x)} + L(m) \right) \\ &= \operatorname*{argmin}_m \left(L(m) - \ln g_m(x) \right) \end{aligned}$$

where ties are broken in an arbitrary fashion. The corresponding encoder partition $\mathcal{S}$ will then yield average distortion

$$\begin{aligned} &\int dx f(x)\, d_I(x, a(x)) \\ &\quad = \sum_m \int_{S_m} dx f(x) \left(\ln \frac{f(x)}{g_m(x)} + L(m) \right) \\ &\quad = \sum_m w_m \left(L(m) + \int_{S_m} dx f_m(x) \ln \frac{f_m(x) w_m}{g_m(x)} \right) \end{aligned}$$

where, as before, $w_m = P_f(S_m)$ and

$$f_m(x) = 1_{S_m}(x) f(x)/w_m.$$

Thus,

$$\begin{aligned} &\int dx f(x)\, d_I(x, a(x)) \\ &\quad = \sum_m w_m I(f_m \| g_m) + \sum_m w_m \ln \frac{w_m}{e^{-L(m)}} \\ &\quad \geq \sum_m w_m I(f_m \| g_m) \end{aligned}$$

with equality if and only if we choose the optimal length function $L(m) = -\ln w_m$. Thus, if we choose an optimal decoder and length function for a partition, the average distortion according to d_I is exactly the mismatch. Thus, iterating the Lloyd optimality properties of optimizing encoder, decoder, and length function can only decrease average distortion and hence also the mismatch.

The Lloyd algorithm for minimizing mismatch produces a collection of models $g_m \in \mathcal{M}$ drawn from some set $\mathcal{M}$ together with a pmf w_m. A collection of pdfs together with a pmf can be viewed as a *mixture* and, hence, the proposed algorithm can be viewed as a means of fitting mixtures of specified families of densities to an arbitrary pdf.

Piecewise Uniform Codes

Let $\mathcal{M}$ consist of all uniform pdfs on bounded sets with positive Lebesgue measure. For any pdf f having bounded support G, its centroid in $\mathcal{M}$ is easily seen to be a uniform pdf on G. In particular, suppose that $u_F \in \mathcal{M}$ has support F. Since we require that $f \ll u$ and f/u is bounded for the mismatch theorem to hold, necessarily $G \subset F$ and hence $V(G) \leq V(F)$. Then

$$\begin{aligned} I(f \| u_F) &= \int_G dx f(x) \ln \frac{f(x)}{u_F(x)} \\ &= \int_G dx f(x) \ln f(x) V(F) \\ &\geq \int_G dx f(x) \ln f(x) V(G) \\ &= I(f \| u_G) \end{aligned}$$

with equality if $F = G$. Thus, the centroid exists and is given by

$$\operatorname*{argmin}_{g \in \mathcal{M}} I(f \| g) = u_G. \tag{59}$$

Thus, the robust uniform model for f is also the minimum relative entropy model for f from the space of all uniform models.

Consider the conditional relative entropy arising with a composite quantizer. In order to fit uniform quantizers with finite conditional relative entropy, we allow only partitions $\mathcal{S}$ having only bounded cells in the support set of f. This will automatically be true, e.g., if the support set of f is bounded. Also, we allow only partitions with a fixed finite number of cells since the infimum of $\overline{\ell}_f(\mathcal{SM})$ over all countable partitions can be seen to be 0. Since all the partition cells S_m are assumed to be bounded, the centroids then follow as before

$$g_m = \operatorname*{argmin}_{g \in \mathcal{M}} I(f_m \| g) = u_{S_m}. \tag{60}$$

Observe that the single uniform model case considered earlier can be considered as an example of the clustered case with only a single reproduction vector corresponding to the centroid of the entire space. Since this adds a constraint to the optimization, the performance must be worse and hence

$$I(f \| g) \geq \sum_m w_m I(f_m \| g_m) \tag{61}$$

if the partitions and models are chosen optimally. (In fact, it is easy to see that for uniform model densities, (61) holds for an arbitrary partition if the models are chosen optimally.) This implies that composite quantizers will indeed provide reduced mismatch from the single "worst case," confirming the motivation for considering them.

Gauss Mixture Codes

Let $\mathcal{M}$ consist of all nonsingular Gaussian pdfs. Again begin by considering the centroid $g \in \mathcal{M}$ as the Gaussian pdf g minimizing $I(f \| g)$. This is accomplished by some algebraic manipulation using relative entropies for Gaussian pdfs as found, e.g., in Kullback [25]. The centroid result is a minor variation on results derived in [1], [16], [22], but the derivation is provided for completeness and is tailored to the specific version of the distortion used here.

Suppose that a Gaussian g has mean μ and covariance K and that f has mean μ_f and covariance K_f. Then

$$\begin{aligned}
I(f||g) &= -h(f) - \int dx f(x) \ln g(x) \\
&= -h(f) + \frac{1}{2}\ln\left((2\pi)^k |K|\right) \\
&\quad + \int dx f(x) \left(\frac{1}{2}(x-\mu)^t K^{-1}(x-\mu)\right) \\
&= -h(f) + \frac{1}{2}\ln(2\pi e)^k |K| \\
&\quad + \frac{1}{2}\,\mathrm{Trace} E_f[(X-\mu)(X-\mu)^t]K^{-1} - \frac{k}{2} \\
&= -h(f) + \frac{1}{2}\ln(2\pi e)^k |K| + \frac{1}{2}\,\mathrm{Trace}(K_f K^{-1}) \\
&\quad + \frac{1}{2}(\mu-\mu_f)^t K^{-1}(\mu-\mu_f) - \frac{k}{2} \\
&= -h(f) + \frac{1}{2}\ln(2\pi e)^k |K_f| \\
&\quad + \left[\frac{1}{2}\ln\frac{|K|}{|K_f|} + \frac{1}{2}\,\mathrm{Trace}(K_f K^{-1}) - \frac{k}{2}\right] \\
&\quad + \frac{1}{2}(\mu-\mu_f)^t K^{-1}(\mu-\mu_f).
\end{aligned}$$

The bracketed term is exactly the relative entropy between a Gaussian pdf with mean μ_f and covariance K_f and a second Gaussian pdf with mean μ and covariance K (e.g., see Kullback [25, p. 189]). Thus, in particular, the quantity is nonnegative and will, in fact, be zero with the choices $\mu_f = \mu$ and $K_f = K$, i.e., if we choose the mean and covariance of the model g to match the mean and covariance of f. The rightmost term is nonnegative and will also be 0 if $\mu_f = \mu$. With these choices we are left with

$$I(f||g) = -h(f) + \frac{1}{2}\ln(2\pi e)^k |K_f|$$

and the centroid g is the Gaussian that has as mean and covariance the mean and covariance with respect to f.

Again consider the conditional relative entropy arising with a composite quantizer. Given an encoder partition $\mathcal{S}$, the centroids are given as above with f replaced by f_m: define the conditional mean $\mu_{f_m} = E_{f_m} X$ and the conditional covariance $K_{f_m} = E_{f_m}[(X-\mu_{f_m})((X-\mu_{f_m})^t]$ (conditioned on $X \in S_m$). Then

$$b(m) = g_m = \operatorname*{argmin}_{g\in\mathcal{M}} I(f_m||g) = \mathcal{N}(x, \mu_{f_m}, K_{f_m}). \quad (62)$$

As with the piecewise-uniform codes, it follows that (61) holds for optimally chosen codes and hence clustered composite quantizers indeed yield smaller mismatch than would a single worst case. For a model quantizer with an optimal decoder, the mismatch can be expressed simply as

$$\begin{aligned}
&\sum_m w_m I(f_m||g_m) \\
&\qquad = -\sum_m w_m h(f_m) + \frac{1}{2}\sum_m w_m \ln(2\pi e)^k |K_{f_m}|
\end{aligned}$$

which, with (20), can be expressed as

$$\begin{aligned}
&\sum_m w_m I(f_m||g_m) \\
&\qquad = -h(f) + H(Z) + \frac{1}{2}\sum_m w_m \ln(2\pi e)^k |K_{f_m}|. \quad (63)
\end{aligned}$$

As with the Lagrangian formulation of variable-rate vector quantization, the average distortion forces a balance between the rightmost term, which tries to match Gaussian models to partition cells, and the entropy term, which puts a cost on partition cells.

When using individual Gaussian models with optimal codebooks and length functions, the the optimal encoder is

$$\begin{aligned}
a(x) = \operatorname*{argmin}_m \Bigg(&-\ln w_m + \frac{1}{2}\ln\left((2\pi)^k |K_{f_m}|\right) \\
&+\frac{1}{2}(x-\mu_{f_m})^t K_{f_m}^{-1}(x-\mu_{f_m})\Bigg). \quad (64)
\end{aligned}$$

The rightmost term is a weighted quadratic distortion measure. Similar distortion measures have been used in pattern recognition with names such as the "local Mahalanobis" distortion since it is a Mahalanobis distortion with respect to the covariance and mean of model m. The results developed here show that the distortion arises naturally in a quantization or minimum conditional relative entropy context. The model selection rule by minimizing d_I in this case corresponds to "quadratic discrimination analysis (QDA)" to find the best of a collection of Gaussian models for a given input vector [24]. Hence, for the Gaussian case, d_I can be considered a QDA distortion as well as an ML or log-likelihood distortion.

This completes the description of the Lloyd algorithm for minimizing mismatch using composite Gaussian quantizers and, hence, provides an algorithm for designing Gauss mixtures. In practice, the pdf f is not known and it must be estimated from the data. Observe, however, that the encoder is well defined given a decoder and length function without any additional knowledge of f. The decoder centroid requires only the conditional expectation and covariance with respect to f, which can be estimated by sample means and covariances. The length function requires only the $P_f(S_m) = w_m$, which can be estimated by the counts for each encoder index. Preliminary results for Lloyd clustering using this and related approaches may be found in [1], [16], [22], [17].

ACKNOWLEDGMENT

The authors gratefully acknowledge the comments and corrections of the Stanford Compression and Classification Group and Ken Zeger.

REFERENCES

[1] A. Aiyer, "Robust image compression using gauss mixture models," Ph.D. dissertation, Dept. Elec. Eng., Stanford Univ., Stanford, CA, Aug. 2001.

[2] 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.

[3] J. A. Bucklew, "Two results on the asymptotic performance of quantizers," *IEEE Trans. Inform. Theory*, vol. IT-30, pp. 341–348, Mar. 1984.
[4] P. A. Chou and B. Betts. When optimal entropy-constrained quantizers have only a finite number of codewords. presented at the IEEE Int. Symp. Information Theory, Cambridge, MA, Aug. 1998. [Online]. Available: http://research.microsoft.com/ pachou/publications.htm
[5] P. A. Chou, T. Lookabaugh, and R. M. Gray, "Entropy-constrained vector quantization," *IEEE Trans. Acoust. Speech Signal Processing*, vol. 37, pp. 31–42, Jan. 1989.
[6] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.
[7] I. Csiszár, "Information-type measures of difference of probability distributions and indirect observations," *Studia Scient. Math. Hung.*, vol. 2, pp. 299–318, 1967.
[8] ——, "Generalized entropy and quantization problems," in *Proc. 6th Prague Conf. Information Theory, Statistical Decision Functions, Random Processes*, 1973, pp. 159–174.
[9] ——, "On an extremum problem of information theory," *Studia Scient. Math. Hung.*, pp. 57–70, 1974.
[10] A. Dembo and I. Kontoyiannis, "Source coding, large deviations, and approximate string matching," *IEEE Trans. Inform. Theory*, vol. 48, pp. 1590–1615, Jun. 2002.
[11] J. L. Doob, *Measure Theory*. New York: Springer-Verlag, 1994.
[12] A. Gersho, "Asymptotically optimal block quantization," *IEEE Trans. Inform. Theory.*, vol. IT-25, pp. 373–380, July 1979.
[13] H. Gish and J. N. Pierce, "Asymptotically efficient quantizing," *IEEE Trans. Inform. Theory*, vol. IT-14, pp. 676–683, Sept. 1968.
[14] S. Graf and H. Luschgy, *Foundations of Quantization for Probability Distributions (Lecture Notes in Mathematics)*. Berlin, Germany: Springer-Verlag, 2000, vol. 1730.
[15] R. M. Gray, *Entropy and Information Theory*. New York: Springer-Verlag, 1990.
[16] ——, "Gauss mixture vector quantization," in *Proc. 2001 IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 3, Salt Lake City, UT, May 2001, pp. 1769–1772.
[17] ——, "Gauss mixture quantization: Clustering Gauss mixtures," in *Nonlinear Estimation and Classification (Lecture Notes in Mathematics)*, D. D. Dennison, M. H. Hansen, C. Holmes, B. Mallick, and B. Yu, Eds. New York: Springer-Verlag, 2003.
[18] R. M. Gray and D. L. Neuhoff, "Quantization," *IEEE Trans. Inform. Theory*, vol. 44, pp. 2325–2384, Oct. 1998.
[19] R. M. Gray and J. Li, "On Zador's entropy-constrained quantization theorem," in *Proc. Date Compression Conf. 2001*. Los Alimitos, CA: IEEE Computer Soc. Press, Mar. 2001, pp. 3–12.
[20] R. M. Gray, T. Linder, and J. Li, "A Lagrangian formulation of Zador's entropy-constrained quantization theorem," *IEEE Trans. Inform. Theory*, vol. 48, pp. 695–707, Mar. 2002.
[21] R. M. Gray, A. H. Gray, Jr., G. Rebolledo, and J. E. Shore, "Rate distortion speech coding with a minimum discrimination information distortion measure," *IEEE Trans. Information Theory*, vol. IT-27, no. 6, pp. 708–721, Nov. 1981.
[22] R. M. Gray, J. C. Young, and A. K. Aiyer, "Minimum discrimination information clustering: modeling and quantization with Gauss mixtures," in *Proc. 2001 IEEE Int. Conf. Image Processing*, vol. 3, Thessaloniki, Greece, Oct. 2001, pp. 14–17.
[23] A. György, T. Linder, P. A. Chou, and B. J. Betts. When optimal entropy-constrained quantizers have a finite or infinite number of codewords. Preprint. [Online]. Available: http://magenta.mast.queensu.ca/~linder/psfiles/GyLiChBe02.ps
[24] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction (Springer Series in Statistics)*. New York: Springer-Verlag, 2001.
[25] S. Kullback, *Information Theory and Statistics*. New York: Dover, 1968. Reprint of 1959 edition published by Wiley.
[26] A. Lapidoth, "On the role of mismatch in rate distortion theory," *IEEE Trans. Inform. Theory*, vol. 43, pp. 38–47, Jan. 1997.
[27] T. Linder, "Lagrangian empirical design of variable-rate vector quantizers: Consistency and convergence rates," *IEEE Trans. Inform. Theory*, vol. 48, pp. 2998–3003, Nov. 2002.
[28] T. Linder and K. Zeger, "Asymptotic entropy constrained performance of tessellating and universal randomized lattice quantization," *IEEE Trans. Inform. Theory*, vol. 40, pp. 575–579, Mar. 1994.
[29] M. S. Pinsker, *Information and Information Stability of Random Variables and Processes*. San Francisco, CA: Holden Day, 1964. (Translated by A. Feinstein from the Russian edition published in 1960 by Izd. Akad. Nauk. SSSR.).
[30] D. J. Sakrison, "Worst sources and robust codes for difference distortion measures," *IEEE Trans. Inform. Theory*, vol. IT-21, pp. 301–309, May 1975.
[31] A. N. Shiryaev, *Probability*, 2nd ed. New York: Springer-Verlag, 1996.
[32] L. Vandenberghe, S. Boyd, and S.-P. Wu, "Determinant maximization with linear matrix inequality constraints," *SIAM J. Matrix Anal. Applic.*, vol. 19, pp. 499–533, 1998.
[33] E. H. Yang and J. Kieffer, "On the performance of data compression algorithms based upon string matching," *IEEE Trans. Inform. Theory*, vol. 44, pp. 47–65, Jan. 1998.
[34] P. L. Zador, "Development and evaluation of procedures for quantizing multivariate distributions," Ph.D. dissertation, Stanford Univ., Stanford, CA, 1963. Also, Stanford Univ. Dept. Statistics Tech. Rep.
[35] ——, "Topics in the asymptotic quantization of continuous random variables," report, Bell Labs. Tech. Memo, 1966.
[36] ——, "Asymptotic quantization error of continuous signals and the quantization dimension," *IEEE Trans. Inform. Theory*, vol. IT–28, pp. 139–148, Mar. 1982.
[37] R. Zamir, "The index entropy of a mismatched codebook," *IEEE Trans. Inform. Theory*, vol. 48, pp. 523–528, Feb. 2002.