

# Entropy Density and Mismatch in High-Rate Scalar Quantization with Rényi Entropy Constraint

Wolfgang Kreitmeier and Tamás Linder, *Senior Member, IEEE*

**Abstract**—Properties of scalar quantization with  $r$ th power distortion and constrained Rényi entropy of order  $\alpha \in (0, 1)$  are investigated. For an asymptotically (high-rate) optimal sequence of quantizers, the contribution to the Rényi entropy due to source values in a fixed interval is identified in terms of the “entropy density” of the quantizer sequence. This extends results related to the well-known point density concept in optimal fixed-rate quantization. A dual of the entropy density result quantifies the distortion contribution of a given interval to the overall distortion. The distortion loss resulting from a mismatch of source densities in the design of an asymptotically optimal sequence of quantizers is also determined. This extends Bucklew’s fixed-rate ( $\alpha = 0$ ) and Gray *et al.*’s variable-rate ( $\alpha = 1$ ) mismatch results to general values of the entropy order parameter  $\alpha$ .

**Index Terms**—Asymptotic quantization theory, distortion density, entropy density, quantizer mismatch, Rényi-entropy.

## I. INTRODUCTION

ASYMPTOTIC quantization theory studies the performance of quantizers of a fixed dimension in the limit of high rates (low distortion). This approach complements Shannon’s rate-distortion theory where optimal codes of a fixed rate (distortion) are investigated as the dimension becomes asymptotically large. Panter and Dite [26] were the first to derive a formula for the mean square distortion of optimum scalar quantizers as the number of quantization levels becomes asymptotically large. Zador’s classic work [29] for vector quantizers determined the asymptotic behavior of the minimum quantizer distortion under a constraint on either the log-cardinality of the quantizer codebook (fixed-rate quantization) or the Shannon entropy of the quantizer output (entropy-constrained quantization). Zador’s results were later clarified and generalized by Bucklew and Wise [6] and Graf and Luschgy [12] for the fixed-rate case, and by Gray *et al.* [14] for the entropy-constrained case. Gray and Neuhoff [16] provide a historical overview of related results.

One way to unify and extend the fixed and variable-rate results is to define the quantizer’s rate by the Rényi entropy of order  $\alpha$  of its output. This generalized rate concept includes

the traditional rate definitions as special cases since  $\alpha = 0$  corresponds to fixed-rate quantization, while setting  $\alpha = 1$  yields variable-rate quantization. This approach was first suggested in [13] as an alternative to the Lagrangian rate definition considered there which simultaneously controls codebook size and output (Shannon) entropy. Further motivation for using Rényi entropy as quantization rate can be obtained from axiomatic considerations [27], [1], as well as from the operational role of the Rényi entropy in variable-length lossless coding [8], [17], [2].

The theory of quantization with Rényi  $\alpha$ -entropy constraint has recently been explored in [18], [19], [20], [21]. In particular, [19] derived the sharp asymptotic behavior of the  $r$ th power distortion of optimal  $d$ -dimensional vector quantizers for  $\alpha \geq 1 + r/d$ . In [21] the technically more challenging  $\alpha < 1 + r/d$  case was considered and the asymptotically optimal  $r$ th power distortion was determined for scalar quantization ( $d = 1$ ) and a fairly large class of source densities. Thus, at least for scalar quantization, only the case  $\alpha \in (1, 1 + r)$  remains open, and it is conjectured in [21] that the main result there remains valid in this range of the parameter  $\alpha$ .

In addition to the asymptotic behavior of the optimal quantizer performance, asymptotic quantization theory has also been concerned with more subtle properties of (asymptotically) optimal quantizers. One such property is the existence, for a sequence of quantizers, of the so-called quantizer point density function, loosely defined as a probability density which, when integrated over a region, gives the fraction of the quantization levels contained in that region. More formally, a point density, if exists, is the probability density function of the limit distribution of the output levels of a sequence of quantizers. Point densities and the closely related companding quantizers have been instrumental in the early pioneering investigations into optimal scalar and vector quantization [3], [26], [22], [10] (see also [24] for a rigorous reformulation of Bennett’s result for the vector case and [16] for the history of these results). Bucklew [7] was the first to rigorously establish the existence of the point density function for an asymptotically optimal sequence of fixed-rate quantizers. To our knowledge, no such rigorous result is known for variable-rate quantization. The concept of quantizer point density has been very useful in analyzing the performance of quantizers in a distributed setting (e.g. [28], [25]).

Asymptotic quantization theory has also been successful in providing mismatch results that quantify the loss in performance when a sequence of quantizers that is asymptotically optimal for one source is applied to a different source. Mismatch results are theoretically important and in practice

W. Kreitmeier is with the Department of Informatics and Mathematics, University of Passau, Innstraße 33, 94032 Passau, Germany (email: wolfgang.kreitmeier@uni-passau.de). T. Linder is with the Department of Mathematics and Statistics, Queen’s University, Kingston, Ontario, Canada K7L 3N6 (email: linder@mast.queensu.ca).

Copyright (c) 2011 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending a request to [pubs-permissions@ieee.org](mailto:pubs-permissions@ieee.org).

This research was supported in part by the German Research Foundation (DFG) and the Natural Sciences and Engineering Research Council (NSERC) of Canada.

they may provide a means for quantifying the performance of code designs that are based on source models estimated from data. For fixed-rate vector quantization Bucklew [7] was the first to prove such a rigorous mismatch result. The variable-rate analog of this result was proved in [15] where connections with mismatch results in rate-distortion theory and robust lossy coding were also pointed out. More recently, Na [23] determined sharp asymptotic formulas for variance-mismatched scalar quantization of Laplacian sources.

In this paper we extend some of the more refined results of fixed and variable-rate asymptotic quantization theory in the framework of quantization with Rényi entropy constraint of order  $\alpha \in (0, 1)$ . The concept of a quantizer point density is a problematic one for (Rényi) entropy-constrained quantization since (near) optimal quantizers can have an arbitrarily large number of levels in any bounded region. Instead, we investigate the Rényi entropy contribution of a given interval to the overall rate. One of our main results, Theorem 2, shows that for a large class of source densities and an asymptotically optimal sequence of quantizers, this contribution can be quantified by the so called entropy density of the sequence. A dual of this result, Corollary 1, quantifies the distortion contribution of a given region to the overall distortion in terms of the so-called distortion density. Interestingly, it turns out that the entropy and distortion densities are equal in the cases we investigate (Remark 5). Our other main contribution, Theorem 3, is a mismatch formula for a sequence of asymptotically optimal Rényi entropy constrained scalar quantizers. From our density and mismatch results we can recover the known results for the traditional rate definitions by formally setting  $\alpha = 0$  or  $\alpha = 1$ .

The rest of the paper is organized as follows. In the next section we formulate the quantization problem and give a somewhat informal overview of our results in the context of prior work. In Section III the entropy and distortion density results are presented and proved. The mismatch problem is considered in Section IV. Concluding remarks are given in Section V.

## II. PRELIMINARIES AND OVERVIEW OF RESULTS

### A. Rényi entropy and quantization

We begin with the definition of Rényi entropy of order  $\alpha$ . Let  $\mathbb{N} := \{1, 2, \dots\}$  and let  $p = (p_1, p_2, \dots) \in [0, 1]^{\mathbb{N}}$  be a probability vector, i.e.  $\sum_{i=1}^{\infty} p_i = 1$ . For any  $\alpha \in (0, \infty) \setminus \{1\}$ , the Rényi entropy of order  $\alpha$ ,  $\hat{H}^\alpha(p) \in [0, \infty]$ , is defined as (see [27] or [1])

$$\hat{H}^\alpha(p) = \frac{1}{1-\alpha} \log \left( \sum_{i=1}^{\infty} p_i^\alpha \right).$$

*Remark 1:* All logarithms in this paper have base  $e$ . Setting  $0^0 := 0$ , we can extend the definition to  $\alpha = 0$ , obtaining

$$\hat{H}^0(p) = \log(\text{card}\{i \in \mathbb{N} : p_i > 0\}) \quad (1)$$

where  $\text{card}$  denotes cardinality. Also, using the convention  $0 \log 0 := 0$ , it is easy to see that letting  $\alpha \rightarrow 1$  yields the

regular (Shannon) entropy of  $p$ :

$$\hat{H}^1(p) := \lim_{\alpha \rightarrow 1} \hat{H}^\alpha(p) = - \sum_{i=1}^{\infty} p_i \log p_i$$

assuming  $\hat{H}^\alpha(p)$  is finite for some  $\alpha < 1$ .

Let  $X$  be a real-valued random variable with distribution  $\mu$ . Let  $\mathbb{I} \subset \mathbb{N}$  be an index set (thus  $\mathbb{I}$  is either finite or countably infinite) and  $\mathcal{S} = \{S_i : i \in \mathbb{I}\}$  a Borel measurable partition of the real line  $\mathbb{R}$ . Moreover let  $\mathcal{C} = \{c_i : i \in \mathbb{I}\}$  be set of distinct points in  $\mathbb{R}$ . Then  $(S_i, c_i)_{i \in \mathbb{I}}$  defines a (scalar) *quantizer*  $q : \mathbb{R} \rightarrow \mathcal{C}$  such that

$$q(x) = c_i \quad \text{if and only if} \quad x \in S_i.$$

We call  $\mathcal{C}$  the *codebook* and the  $c_i$  the codepoints (or quantization levels). Each  $S_i \in \mathcal{S}$  is called *codecell*. Clearly  $\mathcal{C} = q(\mathbb{R})$  is the range of  $q$  and

$$\mathcal{S} = \{q^{-1}(z) : z \in q(\mathbb{R})\}$$

where  $q^{-1}(z) = \{x \in \mathbb{R} : q(x) = z\}$ . Let  $\mathcal{Q}$  denote the set of scalar quantizers, i.e., the set of all Borel-measurable mappings  $q : \mathbb{R} \rightarrow \mathbb{R}$  with a countable range. The discrete random variable  $q(X)$  is a quantized version of the random variable  $X$ . With any enumeration  $\{i_1, i_2, \dots\}$  of  $\mathbb{I}$  we define

$$H_\mu^\alpha(q) = \hat{H}^\alpha(\mu(S_{i_1}), \mu(S_{i_2}), \dots)$$

as the Rényi entropy of order  $\alpha$  of  $q$  with respect to  $\mu$ . Thus  $H_\mu^0(q)$  is the log-cardinality of the codebook of  $q$  (we assume without loss of generality that each codecell of  $q$  has positive probability) and  $H_\mu^1(q)$  is the Shannon entropy of the quantizer output.

For  $r \geq 1$  and  $q \in \mathcal{Q}$  we measure the approximation error between  $X$  and  $q(X)$  by the the  $r$ th power distortion defined by

$$D_\mu(q) = E|X - q(X)|^r = \int |x - q(x)|^r d\mu(x).$$

For any  $R \geq 0$  we define

$$D_\mu^\alpha(R) = \inf\{D_\mu(q) : q \in \mathcal{Q}, H_\mu^\alpha(q) \leq R\}$$

the optimal quantization distortion of  $\mu$  under Rényi entropy constraint  $R$ . We call a quantizer  $q$  optimal for  $\mu$  under the entropy constraint  $R$  if  $D_\mu(q) = D_\mu^\alpha(R)$  and  $H_\mu^\alpha(q) \leq R$ . In particular,  $D_\mu^0(R)$  is the minimum distortion of any quantizer with codebook size not exceeding  $e^R$ , while  $D_\mu^1(R)$  is the minimum distortion under Shannon entropy constraint  $R$ .

In the rest of this paper all distributions to be quantized will be absolutely continuous with respect to the Lebesgue measure  $\lambda$  on the real line. If such a distribution  $\mu$  has probability density function  $g$ , then we will use the notation  $\mu = g\lambda$ . We denote by  $\text{supp}(\mu)$  the support of  $\mu$  (the smallest closed set whose complement has zero  $\mu$  measure). If  $\mu = g\lambda$ , then we define  $\text{supp}(g) = \text{supp}(\mu)$ . We will also assume throughout the paper that the  $r$ th moment of  $\mu$  is finite, i.e.  $\int |x|^r d\mu(x) < \infty$ . This condition is sufficient (but not necessary) for  $D_\mu^\alpha(R)$  to be finite for all  $R \geq 0$ .

It has been shown in [20] that under the above conditions, the set of all quantizers  $\mathcal{Q}$  in the definition of  $D_\mu^\alpha$  can

be replaced by the set of quantizers having finitely many codecells, each of which is an interval. In view of this, we will assume throughout the whole paper that the codecells of every quantizer  $q \in \mathcal{Q}$  are intervals (but we do not restrict the number of codecells to be finite) and each codepoint is contained in the interior of the associated codecell.

### B. Asymptotic optimality and conditional distributions

The main result of [21] implies that under suitable assumptions on the source density  $g$ , for all  $\alpha \leq 1$ ,

$$\lim_{R \rightarrow \infty} e^{rR} D_\mu^\alpha(R) = \frac{1}{(1+r)2^r} e^{rh^{\beta_1}(g)} \quad (2)$$

where  $\beta_1 = \frac{1-\alpha+\alpha r}{1-\alpha+r}$  and

$$h^{\beta_1}(g) = \frac{1}{1-\beta_1} \log \left( \int g^{\beta_1} d\lambda \right)$$

is the Rényi differential entropy of order  $\beta_1$  of  $g$ .

We can formally recover Zador's classical results [29] in the scalar setting from (2). Letting  $\alpha = 0$ , we have  $\beta_1 = \frac{1}{1+r}$  and  $e^{rh^{\beta_1}(g)} = (\int g^{\frac{1}{1+r}} d\lambda)^{1+r} = \|g\|_{\frac{1}{1+r}}^{1+r}$ , yielding Zador's formula for fixed-rate scalar quantization. For  $\alpha = 1$ , we have  $\beta_1 = 1$  and  $e^{rh^{\beta_1}(g)} = e^{rh(g)}$ , where  $h(g) = -\int g \log g d\lambda$  is the Shannon differential entropy of  $g$ , and (2) becomes Zador's formula for variable-rate scalar quantization. In view of (2) we call a sequence of quantizers  $(q_n)_{n \in \mathbb{N}}$  *asymptotically optimal* if  $H_\mu^\alpha(q_n) \rightarrow \infty$  and

$$\lim_{n \rightarrow \infty} e^{rH_\mu^\alpha(q_n)} D_\mu(q_n) = \frac{1}{(1+r)2^r} e^{rh^{\beta_1}(g)}.$$

Suppose  $I$  is a bounded interval with positive  $\mu$  probability. We denote by  $\mu(\cdot|I)$  the conditional distribution for  $\mu$  given  $I$  and by  $g_I$  the corresponding conditional density (so that  $\mu(\cdot|I) = g_I \lambda$ ). We show in Theorem 2 that for  $\alpha \in (0, 1)$  any quantizer sequence  $(q_n)$  that is asymptotically optimal for  $\mu$  is also asymptotically optimal for  $\mu(\cdot|I)$ , i.e.,

$$\lim_{n \rightarrow \infty} e^{rH_{\mu(\cdot|I)}^\alpha(q_n)} D_{\mu(\cdot|I)}(q_n) = \frac{1}{(1+r)2^r} e^{rh^{\beta_1}(g_I)}. \quad (3)$$

Although this result is not very surprising, it will be very useful in establishing further, more subtle properties of asymptotically optimal quantizers.

### C. Entropy and distortion densities

Let  $N_n(I)$  denote the number of codepoints of  $q_n$  contained in an interval  $I$ . Let  $\alpha = 0$  and let  $(q_n)$  be a sequence of asymptotically optimal  $n$ -level quantizers (so that  $H_\mu^0(q_n) = \log n$ ). Specialized to the scalar case, one important result of Bucklew [7] shows that

$$\lim_{n \rightarrow \infty} \frac{N_n(I)}{n} = \frac{\int_I g^{\frac{1}{1+r}} d\lambda}{\int_{\mathbb{R}} g^{\frac{1}{1+r}} d\lambda}. \quad (4)$$

Thus the probability density  $g^{\frac{1}{1+r}} / \int g^{\frac{1}{1+r}} d\lambda$  can be interpreted as the point density function for the codepoints of asymptotically optimal quantizers (see also [12, Thm. 7.5]). Point densities are useful in gaining insight into the structure

of (asymptotically) optimal quantizers and can be used to construct such quantizers via a companding construction.

Unfortunately, no rigorous point density results are known for  $\alpha = 1$ . In fact, even the definition of a point density function is problematic for entropy-constrained quantization since for sources with a density, at any rate  $R > 0$  there exist near-optimal quantizers that have an arbitrarily large number of codepoints contained in a given bounded interval. Thus an analog of (4) cannot hold for an arbitrary sequence of asymptotically optimal quantizers, although heuristic arguments indicate that under some structural restrictions asymptotically optimal variable-rate quantizers have a uniform point density (see, e.g., [11], [10]).

To define a tractable analog of the point density function, recall that  $\mu(\cdot|I)$  denotes the conditional distribution for  $\mu$  given  $I$ . In view of (1), we have  $N_n(I) = e^{H_{\mu(\cdot|I)}^0(q_n)}$  and  $n = e^{H_\mu^0(q_n)}$ . Thus the fraction of codepoints contained in  $I$  on the left hand side of (4) can be rewritten as

$$\frac{N_n(I)}{n} = \frac{e^{H_{\mu(\cdot|I)}^0(q_n)}}{e^{H_\mu^0(q_n)}}. \quad (5)$$

This ratio represents the relative contribution of the interval  $I$  to the total Rényi entropy of order  $\alpha = 0$ .

The interpretation in (5) motivates us to define the Rényi entropy contribution of an interval  $I$  in a similar way for general  $\alpha$ . In Theorem 2, we identify the limit of this entropy contribution: Under appropriate conditions on the source density, for any  $\alpha \in (0, 1)$  and asymptotically optimal sequence  $(q_n)$ , we have

$$\lim_{n \rightarrow \infty} \frac{e^{(1-\alpha)H_{\mu(\cdot|I)}^\alpha(q_n)}}{e^{(1-\alpha)H_\mu^\alpha(q_n)}} = \frac{\int_I g^{\beta_1} d\lambda}{\int_{\mathbb{R}} g^{\beta_1} d\lambda} \mu(I)^{-\alpha}. \quad (6)$$

It is easy to see that (6) reduces to the traditional point density result (4) for  $\alpha = 0$ .

In Corollary 1 we present an almost immediate consequence of (6) and (3) which concerns the distortion contribution of an arbitrary finite interval  $I$ :

$$\lim_{n \rightarrow \infty} \frac{\int_I |x - q_n(x)|^r \mu(dx)}{D_\mu(q_n)} = \frac{\int_I g^{\beta_1} d\lambda}{\int_{\mathbb{R}} g^{\beta_1} d\lambda}.$$

Thus the probability density  $g^{\beta_1} / \int g^{\beta_1} d\lambda$  can be interpreted as either the (Rényi) *entropy density* or the *distortion density* of any asymptotically optimal quantizer sequence  $(q_n)$ .

### D. Mismatch

For scalar quantization Bucklew's fixed-rate mismatch result [7, Thm. 2] can be stated as follows: If a sequence of  $n$ -level quantizers  $(q_n)$  that is asymptotically optimal for a source with distribution  $\mu = g\lambda$  is applied to a source with distribution  $\nu = f\lambda$ , then (under some assumptions on  $g$  and  $f$ )

$$\lim_{n \rightarrow \infty} n^r D_\nu(q_n) = \frac{1}{(1+r)2^r} \int \frac{f}{g^r} d\lambda$$

where  $g_* = g^{\frac{1}{1+r}} / \int g^{\frac{1}{1+r}} d\lambda$  is the optimal point density function for  $\mu = g\lambda$  from (4). This is a generalization of a classical result of Bennett [3] who considered companding quantization and mean square distortion. The integral on the

right hand side is often called Bennett's integral. In view of (4), and after some calculations, we obtain that the asymptotic performance loss due to mismatch is

$$\lim_{n \rightarrow \infty} \frac{D_\nu(q_n)}{D_\nu^0(\log n)} = e^{r \mathcal{D}_{1+r}(f_* \| g_*)} \quad (7)$$

where  $f_* = f^{\frac{1}{1+r}} / \int f^{\frac{1}{1+r}} d\lambda$  is the optimal point density for  $\nu = f\lambda$  and

$$\mathcal{D}_\alpha(u \| v) = \frac{1}{\alpha - 1} \log \left( \int u^\alpha v^{1-\alpha} d\lambda \right) \quad (8)$$

denotes the Rényi divergence of order  $\alpha \neq 1$  between densities  $u$  and  $v$ . (Thus the loss is always greater than one unless  $\mu = \nu$ ).

For the entropy-constrained case the main result of [15] implies that if  $(q_n)$  is asymptotically optimal for  $\mu = g\lambda$ , but it is used for  $\nu = f\lambda$ , then

$$\lim_{n \rightarrow \infty} e^{r H_\nu^1(q_n)} D_\nu(q_n) = \frac{1}{(1+r)2^r} e^{r h^1(f)} e^{r \mathcal{D}_1(f \| g)}.$$

Here  $\mathcal{D}_1(f \| g) = \mathcal{D}(f \| g) = \int f \log \frac{f}{g} d\lambda$  is the Kullback-Leibler divergence (relative entropy) between  $f$  and  $g$ . From (4) the loss due to mismatch is

$$\lim_{n \rightarrow \infty} \frac{D_\nu(q_n)}{D_\nu^1(H_\nu^1(q_n))} = e^{r \mathcal{D}_1(f \| g)}. \quad (9)$$

In Theorem 3 we present a result on mismatch for quantization with constrained Rényi entropy of order  $\alpha \in (0, 1)$ . The result states that if  $(q_n)$  is asymptotically optimal for  $\mu = g\lambda$ , but is applied to  $\nu = f\lambda$ , then

$$\begin{aligned} & \lim_{n \rightarrow \infty} e^{r H_\nu^\alpha(q_n)} D_\nu(q_n) \\ &= \frac{1}{(1+r)2^r} e^{-r \mathcal{D}_\alpha(f \| g_{\alpha,r})} \int \frac{f}{(g_{\alpha,r})^r} d\lambda \end{aligned}$$

where

$$g_{\alpha,r} = \frac{g^{\frac{1}{\beta_2}}}{\int g^{\frac{1}{\beta_2}} d\lambda} \quad (10)$$

with  $\beta_2 = \frac{1-\alpha+r}{1-\alpha}$  (note that  $g_{0,r} = g_*$ ). The loss due to mismatch can be expressed as

$$\lim_{n \rightarrow \infty} \frac{D_\nu(q_n)}{D_\nu^\alpha(H_\nu^\alpha(q_n))} = \frac{e^{r(\mathcal{D}_{1+r}(f_{0,r} \| g_{\alpha,r}) - \mathcal{D}_\alpha(f \| g_{\alpha,r}))}}{e^{r(\mathcal{D}_{1+r}(f_{0,r} \| f_{\alpha,r}) - \mathcal{D}_\alpha(f \| f_{\alpha,r}))}}. \quad (11)$$

The loss can be seen to be always greater than one unless  $\mu = \nu$  (see Remark 6 following Theorem 3). Setting formally  $\alpha = 0$  or  $\alpha = 1$  (or, more precisely, letting  $\alpha \downarrow 0$  or  $\alpha \uparrow 1$ ) in the above formula yields the known cases (7) and (9).

### III. ENTROPY DENSITY AND RELATED RESULTS

Throughout this section we assume that  $\mu = g\lambda$ . For  $r \geq 1$  and  $\alpha \in [0, 1+r) \setminus \{1\}$  let

$$\beta_1 = \frac{1-\alpha+\alpha r}{1-\alpha+r}, \quad \beta_2 = \frac{1-\alpha+r}{1-\alpha}. \quad (12)$$

*Definition 1:* Let  $C(r) = \frac{1}{2^r(1+r)}$  and define, for  $\alpha \in$

$[0, 1+r) \setminus \{1\}$ ,

$$Q_{\alpha,r}(\mu) = C(r) \left( \int g^{\beta_1} d\lambda \right)^{\beta_2}$$

whenever the integral is finite. Note that  $Q_{\alpha,r}(\mu) \in (0, \infty)$ . We call  $Q_{\alpha,r}(\mu)$  the *quantization coefficient* of  $\mu$ .

*Definition 2:* A one-dimensional probability density function  $g$  is called *weakly unimodal* if it is continuous on its support and there exists an  $l_0 > 0$  such that  $\{x : g(x) \geq l\}$  is a compact interval for every  $l \in (0, l_0)$ .

Note that every weakly unimodal density is bounded and its support is a (possibly unbounded) interval. Clearly, all continuous unimodal densities are weakly unimodal. The class of weakly unimodal densities includes many parametric source density classes commonly used in modeling information sources such as exponential, Laplacian, Gaussian, generalized Gaussian, and all bounded gamma and beta densities.

The following is one of the main results in [21].

*Theorem 1 ([21, Thm 3.4]):* For  $r > 1$  and  $\alpha \in (0, 1)$ , if  $\mu$  has a weakly unimodal density  $g$  and  $\int |x|^{r+\delta} d\mu(x) < \infty$  for some  $\delta > 0$ , then  $Q_{\alpha,r}(\mu)$  is well defined and

$$\lim_{R \rightarrow \infty} e^{rR} D_\mu^\alpha(R) = Q_{\alpha,r}(\mu).$$

*Remark 2:* (a) The theorem and (2) express the same asymptotic result since

$$\left( \int g^{\beta_1} d\lambda \right)^{\beta_2} = e^{r h^{\beta_1}(g)}.$$

The quantization coefficient  $Q_{\alpha,r}$  can also be expressed in terms of Rényi divergences (8) and the density  $g_{\alpha,r}$  introduced in (10). One can easily verify that

$$\left( \int g^{\beta_1} d\lambda \right)^{\beta_2} = e^{-r \mathcal{D}_\alpha(g \| g_{\alpha,r})} \int \frac{g}{(g_{\alpha,r})^r} d\lambda. \quad (13)$$

Furthermore, for any density  $h$  with  $\int \frac{g}{h^r} d\lambda < \infty$ ,

$$\begin{aligned} & \int \frac{g}{h^r} d\lambda \\ &= \left( \int g^{\frac{1}{1+r}} d\lambda \right)^{1+r} \int \left( \frac{g^{\frac{1}{1+r}}}{\int g^{\frac{1}{1+r}} d\lambda} \right)^{1+r} h^{(1-(1+r))} d\lambda \\ &= \|g\|_{\frac{1}{1+r}} \int (g_{0,r})^{1+r} h^{(1-(1+r))} d\lambda \\ &= \|g\|_{\frac{1}{1+r}} e^{r \mathcal{D}_{1+r}(g_{0,r} \| h)}. \end{aligned}$$

Substituting  $h = g_{\alpha,r}$  and combining with (13) we obtain

$$\left( \int g^{\beta_1} d\lambda \right)^{\beta_2} = \|g\|_{\frac{1}{1+r}} e^{r(\mathcal{D}_{1+r}(g_{0,r} \| g_{\alpha,r}) - \mathcal{D}_\alpha(g \| g_{\alpha,r}))}.$$

(b) Theorem 3.4 in [21] also covers the more exotic  $\alpha \in [-\infty, 0)$  case, but for technical reasons we require that  $\alpha \in (0, 1)$ . The weak unimodality condition is a technical one and most likely can be significantly relaxed.

*Definition 3:* A sequence of quantizer  $(q_n)_{n \in \mathbb{N}}$  with  $H_\mu^\alpha(q_n) \rightarrow \infty$  as  $n \rightarrow \infty$  is called  $\alpha$ -asymptotically optimal

for  $\mu$  if

$$\lim_{n \rightarrow \infty} \frac{D_\mu(q_n)}{D_\mu^\alpha(H_\mu^\alpha(q_n))} = 1.$$

*Remark 3:* In what follows we will simply write ‘‘asymptotically optimal’’ instead of ‘‘ $\alpha$ -asymptotically optimal.’’ Under the conditions of Theorem 1, a quantizer sequence  $(q_n)$  with  $H_\mu^\alpha(q_n) \rightarrow \infty$  is asymptotically optimal for  $\mu$  if and only if

$$\lim_{n \rightarrow \infty} e^{rH_\mu^\alpha(q_n)} D_\mu(q_n) = Q_{\alpha,r}(\mu).$$

For any measurable  $A \subset \mathbb{R}$  with  $\mu(A) > 0$  we denote by  $\mu(\cdot|A)$  the conditional probability for  $\mu$  given  $A$ . Let  $c, d \in \mathbb{R}$  be such that  $c < d$  and  $\mu((c, d]) \in (0, 1)$ , but otherwise arbitrary. In the following theorem, we let  $A_1 = (c, d]$ ,  $A_2 = \mathbb{R} \setminus A_1$ , and  $\mu_i = \mu(\cdot|A_i)$  for  $i \in \{1, 2\}$ .

*Theorem 2:* Let  $r > 1$  and  $\alpha \in (0, 1)$ . Let  $\mu = g\lambda$ , where the density function  $g$  is weakly unimodal and satisfies  $\int |x|^{r+\delta} d\mu(x) < \infty$  for some  $\delta > 0$ . Let  $(q_n)_{n \in \mathbb{N}}$  be an asymptotically optimal sequence for  $\mu$ . Then, for  $i \in \{1, 2\}$ ,

$$\lim_{n \rightarrow \infty} \frac{e^{(1-\alpha)H_{\mu_i}^\alpha(q_n)}}{e^{(1-\alpha)H_\mu^\alpha(q_n)}} = \frac{\int_{A_i} g^{\beta_1} d\lambda}{\int_{\mathbb{R}} g^{\beta_1} d\lambda} \mu(A_i)^{-\alpha}. \quad (14)$$

and  $(q_n)$  is asymptotically optimal for  $\mu_i$ , i.e.,  $\lim_{n \rightarrow \infty} H_{\mu_i}^\alpha(q_n) = \infty$  and

$$\lim_{n \rightarrow \infty} e^{rH_{\mu_i}^\alpha(q_n)} D_{\mu_i}(q_n) = Q_{\alpha,r}(\mu_i). \quad (15)$$

*Remark 4:* (a) As discussed in Section II-C, the ratio on the left hand side of (14) can be interpreted as the relative contribution to Rényi entropy of interval  $I$ . The theorem determines the limit of this relative entropy contribution for a sequence of asymptotically optimal quantizers. The method used in the proof is a generalization of the approach developed by Bucklew [7] for the case  $\alpha = 0$ .

(b) Using  $\alpha \in (0, 1)$  and the condition  $\int |x|^{r+\delta} d\mu(x) < \infty$ , the integral in the definition of  $Q_{\alpha,r}(\mu)$  can be shown to be finite by an application of Hölder’s inequality as in [12, Remark 6.3 (a)]. For the same reason,  $Q_{\alpha,r}(\mu_i)$  is finite for  $i \in \{1, 2\}$ .

In the proof of the theorem we will need the following lemma which is proved in the Appendix.

*Lemma 1:* Under the conditions of Theorem 2 the following hold: For  $i \in \{1, 2\}$ ,

$$\lim_{n \rightarrow \infty} H_{\mu_i}^\alpha(q_n) = \infty \quad (16)$$

and for all  $p \in \mathbb{R}$ ,

$$\lim_{n \rightarrow \infty} \frac{\mu(q_n^{-1}(q_n(p)))^\alpha}{\sum_{a \in q_n(\mathbb{R})} \mu(q_n^{-1}(a))^\alpha} = 0 \quad (17)$$

and

$$\lim_{n \rightarrow \infty} \frac{\mu_i(q_n^{-1}(q_n(p)))^\alpha}{\sum_{a \in q_n(\mathbb{R})} \mu_i(q_n^{-1}(a))^\alpha} = 0. \quad (18)$$

*Proof of Theorem 2.* We begin the proof by showing that (14) holds if we additionally assume that for  $i \in \{1, 2\}$ ,

$$\limsup_{n \rightarrow \infty} e^{r(H_\mu^\alpha(q_n) - H_{\mu_i}^\alpha(q_n))} < \infty. \quad (19)$$

In this case, any subsequence of  $(q_n)$  has a sub-subsequence, which we also denote by  $(q_n)$ , such that

$$\lim_{n \rightarrow \infty} e^{r(H_\mu^\alpha(q_n) - H_{\mu_1}^\alpha(q_n))} = d^{\frac{r}{1-\alpha}} \quad (20)$$

for some  $d \in [0, \infty)$ . The obvious bound

$$e^{r(H_\mu^\alpha(q_n) - H_{\mu_1}^\alpha(q_n))} \geq \mu(A_1)^{\frac{\alpha r}{1-\alpha}} \quad (21)$$

implies that  $d > 0$ . In what follows we show that  $d$  is independent of the choice of the sub-subsequence (and thus the limit in (20) holds for the original sequence) and explicitly identify  $d$ .

For any two sequences  $(u_n)$  and  $(v_n)$  of positive reals we write  $u_n \sim v_n$  if

$$\lim_{n \rightarrow \infty} \frac{u_n}{v_n} = 1. \quad (22)$$

Note that if  $u_n \sim v_n$  and  $u'_n \sim v'_n$ , then  $(u_n + u'_n) \sim (v_n + v'_n)$  and  $u_n \cdot u'_n \sim v_n \cdot v'_n$ . We can rewrite (20) as

$$e^{(1-\alpha)H_{\mu_1}^\alpha(q_n)} \sim \frac{1}{d} e^{(1-\alpha)H_\mu^\alpha(q_n)}. \quad (23)$$

We note that

$$\begin{aligned} & e^{rH_\mu^\alpha(q_n)} D_\mu(q_n) \\ &= e^{rH_\mu^\alpha(q_n)} \sum_{i=1}^2 \mu(A_i) D_{\mu_i}(q_n) \\ &= \sum_{i=1}^2 e^{r(H_\mu^\alpha(q_n) - H_{\mu_i}^\alpha(q_n))} \mu(A_i) e^{rH_{\mu_i}^\alpha(q_n)} D_{\mu_i}(q_n). \end{aligned} \quad (24)$$

Since the cells of  $q_n$  are intervals, at most two of them may intersect both  $A_1 = (c, d]$  and  $A_2 = \mathbb{R} \setminus (c, d]$  (namely, those containing  $c$  and  $d$ ). Then (18) implies

$$\begin{aligned} & e^{r(H_\mu^\alpha(q_n) - H_{\mu_2}^\alpha(q_n))} \\ &= e^{rH_\mu^\alpha(q_n)} \left( e^{(1-\alpha)H_{\mu_2}^\alpha(q_n)} \right)^{-\frac{r}{1-\alpha}} \\ &\sim e^{rH_\mu^\alpha(q_n)} \mu(A_2)^{\frac{\alpha r}{1-\alpha}} \left( \sum_{a \in q_n(\mathbb{R}): q_n^{-1}(a) \subset A_2} \mu(q_n^{-1}(a))^\alpha \right)^{-\frac{r}{1-\alpha}} \\ &\sim e^{rH_\mu^\alpha(q_n)} \mu(A_2)^{\frac{\alpha r}{1-\alpha}} \times \\ &\quad \times \left( e^{(1-\alpha)H_\mu^\alpha(q_n)} - \sum_{a \in q_n(\mathbb{R}): q_n^{-1}(a) \subset A_1} \mu(q_n^{-1}(a))^\alpha \right)^{-\frac{r}{1-\alpha}} \\ &\sim e^{rH_\mu^\alpha(q_n)} \mu(A_2)^{\frac{\alpha r}{1-\alpha}} \times \\ &\quad \times \left( e^{(1-\alpha)H_\mu^\alpha(q_n)} - e^{(1-\alpha)H_{\mu_1}^\alpha(q_n)} \mu(A_1)^\alpha \right)^{-\frac{r}{1-\alpha}}. \end{aligned}$$

In view of (23) we conclude

$$\begin{aligned} & e^{r(H_\mu^\alpha(q_n) - H_{\mu_2}^\alpha(q_n))} \\ &\sim e^{rH_\mu^\alpha(q_n)} \mu(A_2)^{\frac{\alpha r}{1-\alpha}} \times \\ &\quad \times \left( e^{(1-\alpha)H_\mu^\alpha(q_n)} - \frac{1}{d} e^{(1-\alpha)H_\mu^\alpha(q_n)} \mu(A_1)^\alpha \right)^{-\frac{r}{1-\alpha}} \\ &= \mu(A_2)^{\frac{\alpha r}{1-\alpha}} \left( 1 - \frac{1}{d} \mu(A_1)^\alpha \right)^{-\frac{r}{1-\alpha}}. \end{aligned} \quad (25)$$

Applying (25) and (20) to (24) we obtain

$$\begin{aligned}
& Q_{\alpha,r}(\mu) \\
& \sim e^{rH_{\mu}^{\alpha}(q_n)} D_{\mu}(q_n) \\
& = e^{r(H_{\mu}^{\alpha}(q_n) - H_{\mu_1}^{\alpha}(q_n))} \mu(A_1) e^{rH_{\mu_1}^{\alpha}(q_n)} D_{\mu_1}(q_n) \\
& \quad + e^{r(H_{\mu}^{\alpha}(q_n) - H_{\mu_2}^{\alpha}(q_n))} \mu(A_2) e^{rH_{\mu_2}^{\alpha}(q_n)} D_{\mu_2}(q_n) \\
& \sim d^{\frac{r}{1-\alpha}} \mu(A_1) e^{rH_{\mu_1}^{\alpha}(q_n)} D_{\mu_1}(q_n) \\
& \quad + \mu(A_2)^{\frac{\alpha r}{1-\alpha}} \left(1 - \frac{1}{d} \mu(A_1)^{\alpha}\right)^{-\frac{r}{1-\alpha}} \times \\
& \quad \times \mu(A_2) e^{rH_{\mu_2}^{\alpha}(q_n)} D_{\mu_2}(q_n) \\
& = \mu(A_1)^{\beta_1 \beta_2} (\mu(A_1)^{-\alpha} d)^{\frac{r}{1-\alpha}} e^{rH_{\mu_1}^{\alpha}(q_n)} D_{\mu_1}(q_n) \\
& \quad + \mu(A_2)^{\beta_1 \beta_2} \left(\frac{1}{1 - \frac{1}{\mu(A_1)^{-\alpha} d}}\right)^{\frac{r}{1-\alpha}} \times \\
& \quad \times e^{rH_{\mu_2}^{\alpha}(q_n)} D_{\mu_2}(q_n). \tag{26}
\end{aligned}$$

Since  $H_{\mu_i}^{\alpha}(q_n) \rightarrow \infty$  by (16), Theorem 1 implies<sup>1</sup>

$$\liminf_{n \rightarrow \infty} e^{rH_{\mu_i}^{\alpha}(q_n)} D_{\mu_i}(q_n) \geq Q_{\alpha,r}(\mu_i), \quad i \in \{1, 2\} \tag{27}$$

and thus the limit inferior of the the right hand side of (26) is lower bounded by

$$\begin{aligned}
& (\mu(A_1)^{-\alpha} d)^{\frac{r}{1-\alpha}} Q_{\alpha,r}(\mu_1) \mu(A_1)^{\beta_1 \beta_2} \\
& \quad + \left(\frac{1}{1 - \frac{1}{\mu(A_1)^{-\alpha} d}}\right)^{\frac{r}{1-\alpha}} Q_{\alpha,r}(\mu_2) \mu(A_2)^{\beta_1 \beta_2} \\
& = C(r) (\mu(A_1)^{-\alpha} d)^{\frac{r}{1-\alpha}} \left(\int_{A_1} g^{\beta_1} d\lambda\right)^{\beta_2} \\
& \quad + C(r) \left(\frac{1}{1 - \frac{1}{\mu(A_1)^{-\alpha} d}}\right)^{\frac{r}{1-\alpha}} \left(\int_{A_2} g^{\beta_1} d\lambda\right)^{\beta_2}. \tag{28}
\end{aligned}$$

In view of the definition of  $Q_{\alpha,r}(\mu)$ , combining (26) and (28) yields

$$\left(\int g^{\beta_1} d\lambda\right)^{\beta_2} \geq F(d) \tag{29}$$

where

$$\begin{aligned}
F(d) & = (\mu(A_1)^{-\alpha} d)^{\frac{r}{1-\alpha}} \left(\int_{A_1} g^{\beta_1} d\lambda\right)^{\beta_2} \\
& \quad + \left(\frac{1}{1 - \frac{1}{\mu(A_1)^{-\alpha} d}}\right)^{\frac{r}{1-\alpha}} \left(\int_{A_2} g^{\beta_1} d\lambda\right)^{\beta_2}.
\end{aligned}$$

Now let

$$d_0 = \mu(A_1)^{\alpha} \frac{\int g^{\beta_1} d\lambda}{\int_{A_1} g^{\beta_1} d\lambda} \tag{30}$$

and note that the bound (21) implies  $d^{-1} \mu(A_1)^{\alpha} \in (0, 1]$ . Moreover, from (19) we actually obtain  $d^{-1} \mu(A_1)^{\alpha} \in (0, 1)$ . Thus if  $d \neq d_0$ , then Lemma 4 in the Appendix gives  $F(d) > F(d_0)$ . Moreover, a simple calculation yields  $F(d_0) = \left(\int g^{\beta_1} d\lambda\right)^{\beta_2}$ . Hence we deduce from (29) that  $d = d_0$ .

<sup>1</sup>Strictly speaking, Theorem 1 ([21, Thm 3.4]) does not apply for  $\mu_2$  since its density  $g_2$  is not weakly unimodal. However,  $g_2$  is the mixture of two weakly unimodal densities with well-separated supports, and the proof of [21, Thm 3.4] can easily be extended to this case.

Because we chose an arbitrary convergent subsequence in (20), we obtain that (20) actually holds with  $d = d_0$  for the original quantizer sequence. This and (30) yield (14) for  $i = 1$ . Also, (25) and (30) imply (14) for  $i = 2$ .

As next step we will prove that (15) is true under the assumption (19). We proceed indirectly. Assume first that (15) is not true for  $i = 1$ . Then by (27) we can choose a subsequence of  $(q_n)$ , also denoted by  $(q_n)$ , such that

$$\begin{aligned} \lim_{n \rightarrow \infty} e^{rH_{\mu_1}^{\alpha}(q_n)} D_{\mu_1}(q_n) & > Q_{\alpha,r}(\mu_1) \\ & = C(r) \left(\int_{A_1} \left(\frac{g}{\mu(A_1)}\right)^{\beta_1} d\lambda\right)^{\beta_2}. \end{aligned}$$

We deduce from (24) and (14) that

$$\begin{aligned} \limsup_{n \rightarrow \infty} e^{rH_{\mu_2}^{\alpha}(q_n)} D_{\mu_2}(q_n) \\ < C(r) \left(\int_{A_2} \left(\frac{g}{\mu(A_2)}\right)^{\beta_1} d\lambda\right)^{\beta_2} \end{aligned} \tag{31}$$

since otherwise we would have

$$\begin{aligned} \limsup_{n \rightarrow \infty} e^{rH_{\mu}^{\alpha}(q_n)} D_{\mu}(q_n) \\ > C(r) \sum_{i=1}^2 \left(\frac{\int g^{\beta_1} d\lambda}{\int_{A_i} g^{\beta_1} d\lambda}\right)^{\beta_2-1} \left(\int_{A_i} g^{\beta_1} d\lambda\right)^{\beta_2} \\ = Q_{\alpha,r}(\mu) \end{aligned}$$

which would contradict the asymptotic optimality of  $(q_n)$ . But the right hand side of (31) is  $Q_{\alpha,r}(\mu_2)$ , which contradicts (27), so (15) must hold for  $i = 1$ . Similarly, we end in a contradiction if we assume that (15) does not hold for  $i = 2$ .

It remains to prove that (19) must hold. Assuming the contrary, we have

$$\liminf_{n \rightarrow \infty} e^{r(H_{\mu_i}^{\alpha}(q_n) - H_{\mu}^{\alpha}(q_n))} = 0.$$

Since  $D_{\mu}(q_n) \geq \mu(A_i) D_{\mu_i}(q_n)$ ,

$$\begin{aligned} 0 & = \liminf_{n \rightarrow \infty} e^{r(H_{\mu_i}^{\alpha}(q_n) - H_{\mu}^{\alpha}(q_n))} \\ & = \liminf_{n \rightarrow \infty} \frac{e^{rH_{\mu_i}^{\alpha}(q_n)} \frac{D_{\mu}(q_n)}{D_{\mu_i}(q_n)} D_{\mu_i}(q_n)}{e^{rH_{\mu}^{\alpha}(q_n)} D_{\mu}(q_n)} \\ & \geq \frac{\mu(A_i)}{Q_{\alpha,r}(\mu)} \liminf_{n \rightarrow \infty} e^{rH_{\mu_i}^{\alpha}(q_n)} D_{\mu_i}(q_n), \end{aligned}$$

which would imply

$$\liminf_{n \rightarrow \infty} e^{rH_{\mu_i}^{\alpha}(q_n)} D_{\mu_i}(q_n) = 0$$

contradicting (27). Hence (19) must hold and the proof is complete.  $\square$

Let  $(q_n)_{n \in \mathbb{N}}$  be a sequence of quantizers and for any  $n \geq 1$  and any Borel set  $E \subset \mathbb{R}$  define

$$M_g^n(E) = e^{rH_{\mu}^{\alpha}(q_n)} \int_E |x - q_n(x)|^r g(x) d\lambda(x). \tag{32}$$

Moreover, for  $\alpha \in [0, 1+r) \setminus \{1\}$  let

$$M_g(E) = C(r) \left( \int_E g^{\beta_1} d\lambda \right) \left( \int_{\mathbb{R}} g^{\beta_1} d\lambda \right)^{\frac{r}{1-\alpha}}. \quad (33)$$

Clearly,  $M_g^n$  and  $M_g$  are Borel-measures on  $\mathbb{R}$  that are absolutely continuous with respect to  $\lambda$ . We define the probability measure  $\hat{\mu}$  by setting, for any Borel set  $E \subset \mathbb{R}$ ,

$$\hat{\mu}(E) = \frac{\int_E g^{\beta_1} d\lambda}{\int_{\mathbb{R}} g^{\beta_1} d\lambda}. \quad (34)$$

*Corollary 1:* Let  $r > 1$  and  $\alpha \in (0, 1)$ . Suppose that  $\mu = g\lambda$ , where the density function  $g$  is weakly unimodal and satisfies  $\int |x|^{r+\delta} d\mu(x) < \infty$  for some  $\delta > 0$ . If  $(q_n)_{n \in \mathbb{N}}$  is an asymptotically optimal sequence of quantizers for  $\mu$ , then for any  $c, d \in \mathbb{R}$  such that  $-\infty < c < d < \infty$  we have

- (i)  $\lim_{n \rightarrow \infty} \frac{\int_{(c,d]} |x - q_n(x)|^r g(x) d\lambda(x)}{\int_{\mathbb{R}} |x - q_n(x)|^r g(x) d\lambda(x)} = \hat{\mu}((c, d]);$
- (ii)  $M_g^n$  converges weakly to  $M_g$ .

*Remark 5:* Combining Theorem 2 and the corollary and using the  $\sim$  notation introduced in (22), we observe that

$$\begin{aligned} & \frac{\int_{(c,d]} |x - q_n(x)|^r d\mu(x)}{\int_{\mathbb{R}} |x - q_n(x)|^r d\mu(x)} \\ & \sim \frac{\sum_{a \in q_n(\mathbb{R})} \mu(q_n^{-1}(a) \cap (c, d])^\alpha}{\sum_{a \in q_n(\mathbb{R})} \mu(q_n^{-1}(a))^\alpha}. \end{aligned} \quad (35)$$

This means that the relative error and entropy contributions of  $(q_n)$  over any given interval asymptotically coincide.

*Proof of Corollary 1.* We start by proving (i). Let  $A = (c, d]$  and define

$$\mu_n(A) = \frac{\int_A |x - q_n(x)|^r g(x) d\lambda(x)}{\int_{\mathbb{R}} |x - q_n(x)|^r g(x) d\lambda(x)}.$$

Obviously we can assume without loss of generality that  $\mu(A) \in (0, 1)$ . Applying (14) and (15) in Theorem 2, we obtain

$$\begin{aligned} \mu_n(A) &= \frac{e^{rH_\mu^\alpha(q_n)} \int_A |x - q_n(x)|^r g(x) d\lambda(x)}{e^{rH_\mu^\alpha(q_n)} \int_{\mathbb{R}} |x - q_n(x)|^r g(x) d\lambda(x)} \\ &\sim \frac{e^{r(H_\mu^\alpha(q_n) - H_{\mu(\cdot|A)}^\alpha(q_n))} \mu(A)}{Q_{\alpha,r}(\mu)} \times \\ &\quad \times \frac{e^{rH_{\mu(\cdot|A)}^\alpha(q_n)} \int_A |x - q_n(x)|^r \frac{g(x)}{\mu(A)} d\lambda(x)}{Q_{\alpha,r}(\mu)} \\ &\sim \frac{e^{r(H_\mu^\alpha(q_n) - H_{\mu(\cdot|A)}^\alpha(q_n))} Q_{\alpha,r}(\mu(\cdot|A)) \mu(A)}{Q_{\alpha,r}(\mu)} \\ &\sim \left( \frac{\mu(A)^\alpha \int g^{\beta_1} d\lambda}{\int_A g^{\beta_1} d\lambda} \right)^{\frac{r}{1-\alpha}} \frac{Q_{\alpha,r}(\mu(\cdot|A)) \mu(A)}{Q_{\alpha,r}(\mu)}. \end{aligned} \quad (36)$$

Definition 1, (34), and a straightforward calculation yield that the right hand side of (36) is equal to  $\hat{\mu}(A)$ .

Next we prove (ii). Because  $(q_n)$  is asymptotically optimal for  $\mu$  we have  $M_g^n(\mathbb{R}) \rightarrow M_g(\mathbb{R})$  as  $n \rightarrow \infty$ . Moreover,  $M_g$  is a finite measure. Due to a refined version of the Portmanteau theorem [4, Thm. 2.4 and Example 2.3] it suffices to prove that  $M_g^n((c, d]) \rightarrow M_g((c, d])$  for any  $-\infty < c < d < \infty$ . Let  $A = (c, d]$  and assume  $\mu(A) > 0$ , since otherwise  $M_g^n(A) =$

$M_g(A) = 0$  for all  $n$ . Applying the definitions (32) and (33), we obtain

$$\begin{aligned} \frac{M_g^n(A)}{M_g(A)} &= \frac{\int_A |x - q_n(x)|^r g(x) d\lambda(x)}{D_\mu(q_n)} \times \\ &\quad \times \frac{e^{rH_\mu^\alpha(q_n)} D_\mu(q_n)}{C(r) \left( \int g^{\beta_1} d\lambda \right)^{\frac{r}{1-\alpha}} \int_A g^{\beta_1} d\lambda}. \end{aligned}$$

Since  $(q_n)$  is asymptotically optimal for  $\mu$  and by (i) we deduce

$$\lim_{n \rightarrow \infty} \frac{M_g^n(A)}{M_g(A)} = \frac{C(r) \left( \int g^{\beta_1} d\lambda \right)^{\beta_2} \frac{\int_A g^{\beta_1} d\lambda}{\int g^{\beta_1} d\lambda}}{C(r) \left( \int g^{\beta_1} d\lambda \right)^{\frac{r}{1-\alpha}} \int_A g^{\beta_1} d\lambda} = 1$$

which proves (ii).  $\square$

#### IV. ASYMPTOTIC MISMATCH

In this section we investigate the performance of a sequence of quantizers  $(q_n)$  that is asymptotically optimal for the source distribution  $\mu$  having density  $g$ , but is applied to the source distribution  $\nu$  having density  $f$ .

*Theorem 3:* Let  $r > 1$  and  $\alpha \in (0, 1)$ . Suppose  $\mu = g\lambda$ ,  $\nu = f\lambda$ , where  $g$  and  $f$  are weakly unimodal densities such that  $f/g$  is bounded. Assume  $\int |x|^{r+\delta} d\mu(x) < \infty$  for some  $\delta > 0$ . If  $(q_n)_{n \in \mathbb{N}}$  is an asymptotically optimal sequence of quantizers for  $\mu$ , then

$$\lim_{n \rightarrow \infty} e^{(1-\alpha)(H_\nu^\alpha(q_n) - H_\mu^\alpha(q_n))} = \frac{\int (f/g)^\alpha g^{\beta_1} d\lambda}{\int g^{\beta_1} d\lambda} \quad (37)$$

and

$$\begin{aligned} & \lim_{n \rightarrow \infty} e^{rH_\nu^\alpha(q_n)} D_\nu(q_n) \\ &= C(r) \left( \int f^\alpha (g^{\frac{1}{\beta_2}})^{1-\alpha} d\lambda \right)^{\frac{r}{1-\alpha}} \int f (g^{\frac{1}{\beta_2}})^{-r} d\lambda. \end{aligned} \quad (38)$$

*Remark 6:* (a) The mismatch formula (38) is best interpreted through the companding quantization approach. In [21, Remark 4.12] it was shown that for a source with density  $g$ , companding quantizers having point density  $h$  induce high-rate asymptotics performance proportional to

$$\left( \int g^\alpha h^{1-\alpha} d\lambda \right)^{\frac{r}{1-\alpha}} \int gh^{-r} d\lambda. \quad (39)$$

Asymptotically optimal companding is obtained by setting  $h = g_{\alpha,r} = g^{1/\beta_2} (\int g^{1/\beta_2} d\lambda)^{-1}$ , which is the unique minimizer of (39). If the sequence of companding quantizers with this choice of  $h$  is now applied to the mismatched distribution  $\nu = f\lambda$ , then the same asymptotic performance as in (38) is obtained. Thus the main significance of (38) is that it holds for an *arbitrary* asymptotically optimal sequence  $(q_n)$ . The analogy with companding quantization suggests that although  $(q_n)$  can have infinitely many codecells, one can interpret  $g_{\alpha,r}$  as the point density related to every asymptotically optimal sequence of quantizers for  $\mu = g\lambda$ .

(b) Using the notation introduced in Sections II-B and II-D, we can rewrite the mismatch formula (38) in the equivalent

forms

$$\begin{aligned}
& \lim_{n \rightarrow \infty} e^{rH_\nu^\alpha(q_n)} D_\nu(q_n) \\
&= C(r) e^{-r\mathcal{D}_\alpha(f\|g_{\alpha,r})} \int \frac{f}{(g_{\alpha,r})^r} d\lambda \\
&= C(r) \|f\|_{\frac{1}{1+r}} e^{r(\mathcal{D}_{1+r}(f_{0,r}\|g_{\alpha,r}) - \mathcal{D}_\alpha(f\|g_{\alpha,r}))} \\
&= Q_{0,r}(\nu) e^{r(\mathcal{D}_{1+r}(f_{0,r}\|g_{\alpha,r}) - \mathcal{D}_\alpha(f\|g_{\alpha,r}))}.
\end{aligned}$$

Formula (11) for the loss due to mismatch follows from either of the last two expressions. The loss is always greater than one unless  $\mu = \nu$  since, according to the preceding comment,  $h = f_{\alpha,r}$  is the unique minimizer of  $e^{r(\mathcal{D}_{1+r}(f_{0,r}\|h) - \mathcal{D}_\alpha(f\|h))}$  over all densities  $h$ .

(c) The condition for the boundedness of  $f/g$  is the same as in the variable-rate mismatch result of [15]. The fixed-rate result of Bucklew [7] requires essentially the same condition since the only known example when the uniform integrability condition given there is satisfied requires that  $f/g$  be bounded.

(d) The conditions of Theorem 3 are satisfied when the support of  $\mu$  and  $\nu$  is the same compact interval  $I$  and the corresponding densities  $g$  and  $f$  are continuous and bounded away from zero on  $I$ . But the theorem may also apply to distributions with unbounded support. For example, if  $g$  and  $f$  are Gaussian or Laplacian densities with mean zero and variance  $\sigma_\mu^2$  and  $\sigma_\nu^2$ , respectively, then the conditions are met if  $\sigma_\mu^2 \geq \sigma_\nu^2$ . Unfortunately, the boundedness condition is not satisfied when  $\sigma_\mu^2 < \sigma_\nu^2$  or when  $g$  is Gaussian and  $f$  is Laplacian. Na [23] obtained a mismatch result for two zero-mean Laplacian sources with arbitrarily mismatched variances by considering quantile quantizers, a special class of fixed-rate asymptotically optimal quantizers closely related to companding quantizers.

*Proof:* Let  $I = \text{supp}(\nu)$ . We will proceed in several steps.

1. First we prove relation (37) under the stated assumptions on  $g$  and  $f/g$ , but additionally assuming that  $I$  is a compact interval and

$$\min\{f(x) : x \in I\} > 0. \quad (40)$$

Let  $m \geq 2$  and let  $\{I_{k,m} : k = 1, \dots, m\}$  be a collection of disjoint intervals of equal length  $\lambda(I)/m$  such that  $\bigcup_{k=1}^m I_{k,m} = I$ . Let  $l_{k,m} = \inf I_{k,m}$  and  $r_{k,m} = \sup I_{k,m}$  denote, respectively, the left and right endpoints of  $I_{k,m}$ . Define

$$S_{m,n} = \bigcup_{k=1}^m \{q_n(l_{k,m}), q_n(r_{k,m})\} \subset q_n(\mathbb{R})$$

and

$$i(f, I) = \min\{f(x) : x \in I\}, \quad s(f, I) = \max\{f(x) : x \in I\}.$$

Note that  $\text{card}(S_{m,n}) \leq m+1$  and  $0 < i(f, I) \leq s(f, I) < \infty$ . Since  $f/g \leq M$  for some  $M < \infty$ , we have

$$\nu(A) \leq M\mu(A) \quad (41)$$

for any measurable  $A \subset \mathbb{R}$ . Thus by (17) in Lemma 1 we get

$$\begin{aligned}
& \limsup_{n \rightarrow \infty} \frac{\sum_{a \in S_{m,n}} \nu(q_n^{-1}(a))^\alpha}{\sum_{a \in q_n(\mathbb{R})} \mu(q_n^{-1}(a))^\alpha} \\
& \leq M^\alpha \limsup_{n \rightarrow \infty} \frac{\sum_{a \in S_{m,n}} \mu(q_n^{-1}(a))^\alpha}{\sum_{a \in q_n(\mathbb{R})} \mu(q_n^{-1}(a))^\alpha} = 0.
\end{aligned}$$

Noting that for any  $a \in q_n(\mathbb{R}) \setminus S_{m,n}$  we either have  $q_n^{-1}(a) \subset I_{k,m}$  for some  $k \in \{1, \dots, m\}$  or  $\nu(q_n^{-1}(a)) = 0$ , the above implies

$$\begin{aligned}
& \limsup_{n \rightarrow \infty} e^{(1-\alpha)(H_\nu^\alpha(q_n) - H_\mu^\alpha(q_n))} \\
&= \limsup_{n \rightarrow \infty} \frac{\sum_{a \in q_n(\mathbb{R})} \nu(q_n^{-1}(a))^\alpha}{\sum_{a \in q_n(\mathbb{R})} \mu(q_n^{-1}(a))^\alpha} \\
&= \sum_{k=1}^m \limsup_{n \rightarrow \infty} \frac{\sum_{a \in q_n(\mathbb{R})} \nu(q_n^{-1}(a) \cap I_{k,m})^\alpha}{\sum_{a \in q_n(\mathbb{R})} \mu(q_n^{-1}(a))^\alpha} \\
&= \sum_{k=1}^m \limsup_{n \rightarrow \infty} \left( \frac{\sum_{a \in q_n(\mathbb{R})} \left( \frac{\nu(q_n^{-1}(a) \cap I_{k,m})}{\mu(q_n^{-1}(a) \cap I_{k,m})} \right)^\alpha}{\sum_{a \in q_n(\mathbb{R})} \mu(q_n^{-1}(a))^\alpha} \times \right. \\
& \quad \left. \times \frac{\mu(q_n^{-1}(a) \cap I_{k,m})^\alpha}{\sum_{a \in q_n(\mathbb{R})} \mu(q_n^{-1}(a))^\alpha} \right). \quad (42)
\end{aligned}$$

Now we observe that  $f/g \leq M$  and (40) imply for all  $n \geq 1$ ,  $k \in \{1, \dots, m\}$ , and  $a \in q_n(\mathbb{R})$ ,

$$0 < \frac{i(f, I_{k,m})}{s(g, I_{k,m})} \leq \frac{\nu(q_n^{-1}(a) \cap I_{k,m})}{\mu(q_n^{-1}(a) \cap I_{k,m})} \leq \frac{s(f, I_{k,m})}{i(g, I_{k,m})} < \infty. \quad (43)$$

Combining (42) and (43) we deduce from (14) in Theorem 2 that

$$\begin{aligned}
& \limsup_{n \rightarrow \infty} e^{(1-\alpha)(H_\nu^\alpha(q_n) - H_\mu^\alpha(q_n))} \\
& \leq \sum_{k=1}^m \left( \frac{s(f, I_{k,m})}{i(g, I_{k,m})} \right)^\alpha \limsup_{n \rightarrow \infty} \frac{\sum_{a \in q_n(\mathbb{R})} \mu(q_n^{-1}(a) \cap I_{k,m})^\alpha}{\sum_{a \in q_n(\mathbb{R})} \mu(q_n^{-1}(a))^\alpha} \\
& = \sum_{k=1}^m \left( \frac{s(f, I_{k,m})}{i(g, I_{k,m})} \right)^\alpha \frac{\int_{I_{k,m}} g^{\beta_1} d\lambda}{\int g^{\beta_1} d\lambda} = \int \bar{h}_m d\hat{\mu} \quad (44)
\end{aligned}$$

where  $\hat{\mu}$  is defined in (34) and we have defined

$$\bar{h}_m = \sum_{k=1}^m 1_{I_{k,m}} \left( \frac{s(f, I_{k,m})}{i(g, I_{k,m})} \right)^\alpha.$$

Here  $1_A$  denotes the characteristic function of  $A \subset \mathbb{R}$  defined by  $1_A(x) = 1$  if  $x \in A$  and  $1_A(x) = 0$  if  $x \notin A$ . Similarly we obtain

$$\liminf_{n \rightarrow \infty} e^{(1-\alpha)(H_\nu^\alpha(q_n) - H_\mu^\alpha(q_n))} \geq \int \underline{h}_m d\hat{\mu} \quad (45)$$

with

$$\underline{h}_m = \sum_{k=1}^m 1_{I_{k,m}} \left( \frac{i(f, I_{k,m})}{s(g, I_{k,m})} \right)^\alpha.$$

Obviously,  $\underline{h}_m \leq \bar{h}_m$ , and since  $f$ ,  $g$ , and  $f/g$  are continuous on  $I$  and the common length of the intervals  $I_{k,m}$  converges



to zero as  $m \rightarrow \infty$ ,

$$\lim_{m \rightarrow \infty} \underline{h}_m(x) = \lim_{m \rightarrow \infty} \bar{h}_m(x) = (f(x)/g(x))^\alpha \quad \text{for all } x \in I.$$

Since the  $\bar{h}_m$  are uniformly bounded, from Fatou's lemma and by dominated convergence, we get

$$\begin{aligned} \int (f/g)^\alpha d\hat{\mu} &= \int \liminf_{m \rightarrow \infty} \underline{h}_m d\hat{\mu} \leq \liminf_{m \rightarrow \infty} \int \underline{h}_m d\hat{\mu} \\ &\leq \limsup_{m \rightarrow \infty} \int \bar{h}_m d\hat{\mu} = \int (f/g)^\alpha d\hat{\mu}. \end{aligned} \quad (46)$$

Combining (46) with (44) and (45) we obtain

$$\lim_{n \rightarrow \infty} e^{(1-\alpha)(H_\nu^\alpha(q_n) - H_\mu^\alpha(q_n))} = \int (f/g)^\alpha d\hat{\mu}.$$

By the definition of  $\hat{\mu}$  in (34) this yields (37).

2. We now prove relation (37) under the stated assumptions. Since  $f$  is weakly unimodal, the set  $I_\delta = \{x : f(x) \geq \delta\} \subset I$  is a compact interval for all  $\delta > 0$  small enough. Since  $\bigcup_{\delta > 0} I_\delta = I$ , we have  $\nu(I \setminus I_\delta) \rightarrow 0$  as  $\delta \rightarrow 0$ , and we also have  $\hat{\mu}(I \setminus I_\delta) \rightarrow 0$  as  $\delta \rightarrow 0$  because  $\hat{\mu}(\cdot|I)$  is absolutely continuous with respect to  $\nu$ . Consequently,

$$\begin{aligned} \limsup_{n \rightarrow \infty} \left| \int_I (f/g)^\alpha d\hat{\mu} - \nu(I_\delta)^{-\alpha} \int_{I_\delta} (f/g)^\alpha d\hat{\mu} \right| \\ \leq M^\alpha (\hat{\mu}(I \setminus I_\delta) + |1 - \nu(I_\delta)^{-\alpha}| \hat{\mu}(I_\delta)) \rightarrow 0 \end{aligned} \quad (47)$$

as  $\delta \rightarrow 0$ . Set  $[c_\delta, d_\delta] := I_\delta$ . Using (41) and (17) in Lemma 1 we obtain

$$\begin{aligned} \limsup_{n \rightarrow \infty} \left| e^{(1-\alpha)(H_\nu^\alpha(q_n) - H_\mu^\alpha(q_n))} - e^{(1-\alpha)(H_{\nu(\cdot|I_\delta)}^\alpha(q_n) - H_\mu^\alpha(q_n))} \right| \\ = \limsup_{n \rightarrow \infty} \left| \frac{\sum_{a \in q_n(\mathbb{R})} \nu(q_n^{-1}(a))^\alpha}{\sum_{a \in q_n(\mathbb{R})} \mu(q_n^{-1}(a))^\alpha} \right. \\ \left. - \frac{\nu(I_\delta)^{-\alpha} \sum_{a \in q_n(\mathbb{R})} \nu(q_n^{-1}(a) \cap I_\delta)^\alpha}{\sum_{a \in q_n(\mathbb{R})} \mu(q_n^{-1}(a))^\alpha} \right| \\ \leq M^\alpha \limsup_{n \rightarrow \infty} \frac{\sum_{a \in q_n(\mathbb{R}): q_n^{-1}(a) \subset I \setminus I_\delta} \mu(q_n^{-1}(a))^\alpha}{\sum_{a \in q_n(\mathbb{R})} \mu(q_n^{-1}(a))^\alpha} \\ + M^\alpha \limsup_{n \rightarrow \infty} \frac{\mu(q_n^{-1}(q_n(c_\delta)))^\alpha + \mu(q_n^{-1}(q_n(d_\delta)))^\alpha}{\sum_{a \in q_n(\mathbb{R})} \mu(q_n^{-1}(a))^\alpha} \\ + \limsup_{n \rightarrow \infty} \left| \frac{\sum_{a \in q_n(\mathbb{R}): q_n^{-1}(a) \subset I_\delta} \nu(q_n^{-1}(a))^\alpha}{\sum_{a \in q_n(\mathbb{R})} \mu(q_n^{-1}(a))^\alpha} \right. \\ \left. - \frac{\nu(I_\delta)^{-\alpha} \sum_{a \in q_n(\mathbb{R})} \nu(q_n^{-1}(a) \cap I_\delta)^\alpha}{\sum_{a \in q_n(\mathbb{R})} \mu(q_n^{-1}(a))^\alpha} \right| \\ = M^\alpha \limsup_{n \rightarrow \infty} \frac{\sum_{a \in q_n(\mathbb{R})} \mu(q_n^{-1}(a) \cap I \setminus I_\delta)^\alpha}{\sum_{a \in q_n(\mathbb{R})} \mu(q_n^{-1}(a))^\alpha} \\ + \limsup_{n \rightarrow \infty} \left| \frac{\sum_{a \in q_n(\mathbb{R}): q_n^{-1}(b) \subset I_\delta} \nu(q_n^{-1}(a) \cap I_\delta)^\alpha}{\sum_{a \in q_n(\mathbb{R})} \mu(q_n^{-1}(a))^\alpha} \right. \\ \left. - \frac{\nu(I_\delta)^{-\alpha} \sum_{a \in q_n(\mathbb{R})} \nu(q_n^{-1}(a) \cap I_\delta)^\alpha}{\sum_{a \in q_n(\mathbb{R})} \mu(q_n^{-1}(a))^\alpha} \right| \\ \leq M^\alpha \hat{\mu}(I \setminus I_\delta) \end{aligned}$$

$$\begin{aligned} + (1 - \nu(I_\delta)^{-\alpha}) \times \\ \times \limsup_{n \rightarrow \infty} \left| \frac{\sum_{a \in q_n(\mathbb{R}): q_n^{-1}(a) \subset I_\delta} \nu(q_n^{-1}(a) \cap I_\delta)^\alpha}{\sum_{a \in q_n(\mathbb{R})} \mu(q_n^{-1}(a))^\alpha} \right| \\ + \nu(I_\delta)^{-\alpha} M^\alpha \times \\ \times \limsup_{n \rightarrow \infty} \frac{\mu(q_n^{-1}(q_n(c_\delta)))^\alpha + \mu(q_n^{-1}(q_n(d_\delta)))^\alpha}{\sum_{a \in q_n(\mathbb{R})} \mu(q_n^{-1}(a))^\alpha} \\ \leq M^\alpha (\hat{\mu}(I \setminus I_\delta) + 1 - \nu(I_\delta)^{-\alpha}) \rightarrow 0 \end{aligned} \quad (48)$$

as  $\delta \rightarrow 0$ . Noting that the density of  $\nu(\cdot|I_\delta)$  satisfies the condition imposed on  $f$  in step 1, we obtain from this step that

$$\begin{aligned} \lim_{n \rightarrow \infty} e^{(1-\alpha)(H_{\nu(\cdot|I_\delta)}^\alpha(q_n) - H_\mu^\alpha(q_n))} \\ = \nu(I_\delta)^{-\alpha} \int_{I_\delta} (f/g)^\alpha d\hat{\mu} \end{aligned} \quad (49)$$

Combining (47),(48), and (49) we obtain that given any  $\varepsilon > 0$  we can choose  $\delta = \delta(\varepsilon) > 0$  small enough and  $N = N(\delta, \varepsilon)$  large enough such that for all  $n > N$ ,

$$\left| e^{(1-\alpha)(H_\nu^\alpha(q_n) - H_\mu^\alpha(q_n))} - \int_I (f/g)^\alpha d\hat{\mu} \right| < \varepsilon$$

which yields (37).

3. We finish the proof by proving assertion (38). Using definition (32) we get

$$\begin{aligned} e^{rH_\nu^\alpha(q_n)} \int |x - q_n(x)|^r d\nu(x) \\ = \left( e^{(1-\alpha)(H_\nu^\alpha(q_n) - H_\mu^\alpha(q_n))} \right)^{\frac{r}{1-\alpha}} \int (f/g) dM_g^n. \end{aligned}$$

Thus Corollary 1 (ii) and (37) yield

$$\begin{aligned} \lim_{n \rightarrow \infty} e^{rH_\nu^\alpha(q_n)} \int |x - q_n(x)|^r d\nu(x) \\ = \left( \frac{\int (f/g)^\alpha g^{\beta_1} d\lambda}{\int g^{\beta_1} d\lambda} \right)^{\frac{r}{1-\alpha}} \int (f/g) dM_g. \end{aligned}$$

Using (33) we calculate

$$\begin{aligned} \lim_{n \rightarrow \infty} e^{rH_\nu^\alpha(q_n)} \int |x - q_n(x)|^r d\nu(x) \\ = \left( \frac{\int (f/g)^\alpha g^{\beta_1} d\lambda}{\int g^{\beta_1} d\lambda} \right)^{\frac{r}{1-\alpha}} C(r) \times \\ \times \left( \int_{\mathbb{R}} g^{\beta_1} d\lambda \right)^{\frac{r}{1-\alpha}} \left( \int (f/g) g^{\beta_1} d\lambda \right) \\ = C(r) \left( \int f^\alpha g^{\beta_1 - \alpha} d\lambda \right)^{\frac{r}{1-\alpha}} \left( \int f g^{\beta_1 - 1} d\lambda \right). \end{aligned} \quad (50)$$

Using definition (12) it is easy to check that  $\beta_1 - \alpha = \frac{1-\alpha}{\beta_2}$  and  $\beta_1 - 1 = -\frac{r}{\beta_2}$ , and hence (50) is equivalent to (38), completing the proof.  $\square$

## V. CONCLUDING REMARKS

We extended point density and mismatch results in fixed and variable-rate asymptotic quantization theory to scalar quantization with Rényi entropy constraint of order  $\alpha \in (0, 1)$ . We showed that the Rényi entropy contribution of a given interval

to the overall rate for a sequence of asymptotically optimal quantizers is determined by the so-called entropy density of the sequence, an analog of the traditional quantizer point density function. A dual of this result quantifies the distortion contribution of a given region to the overall distortion. We also proved a mismatch formula for a sequence of asymptotically optimal Rényi entropy constrained scalar quantizers. One can recover the known results for the traditional rate definitions by formally setting  $\alpha = 0$  or  $\alpha = 1$  in our density and mismatch results.

A natural question is whether the density and mismatch results of this paper can be generalized to higher dimensional (vector) quantization. To make progress in this direction, one first needs to generalize Theorem 1 to higher dimensions (cf. [21, Section VIII]) to obtain an analog of Zador's fixed and variable-rate vector quantization results for Rényi entropy constraint. Assuming one can prove such a result, the main difficulty in generalizing our proofs seems to be controlling the entropy contribution at the boundary of hypercubes (higher-dimensional intervals).

Another interesting question is whether the coincidence of distortion and entropy densities described by (35) in Remark 5 is particular to quantization with Rényi entropy or is a deeper phenomenon. In particular, one can ask whether replacing Rényi's entropy with some more general information measure (c.f. [9]) would preserve the existence of and the special relationship between entropy and distortion densities. Answers to these questions would provide a more complete understanding of some of the finer aspects of quantization theory.

As mentioned before, an analog of the fixed-rate point density result of Bucklew [7] (see (4)) cannot hold for arbitrary sequences of asymptotically optimal entropy-constrained quantizers. However, point densities play an important role in our intuitive understanding of the structure of optimal quantizers, and may provide (heuristic) guidance in constructing (nearly) optimal quantizers. Thus it would be interesting to find a framework within which rigorous point density result can be proved for Rényi entropy constrained quantization (and for traditional entropy-constrained quantization). For the scalar case, companding quantization provides such a framework, but for higher dimensions, the restriction to companding usually precludes asymptotic optimality [5].

## VI. APPENDIX

*Proof of Lemma 1.* We first show (16). The asymptotic optimality of  $(q_n)$  for  $\mu$  implies that  $D_\mu(q_n) \rightarrow 0$  as  $n \rightarrow \infty$ . Since  $\mu$  has a density, this yields, via Lemma 2 below, the intuitively obvious fact that

$$\lim_{n \rightarrow \infty} \max\{\mu(q_n^{-1}(a)) : a \in q_n(\mathbb{R})\} = 0.$$

This also gives for  $i \in \{1, 2\}$ ,

$$\lim_{n \rightarrow \infty} \max\{\mu_i(q_n^{-1}(a)) : a \in q_n(\mathbb{R})\} = 0. \quad (51)$$

Let  $p = (p_1, p_2, \dots)$  be a probability vector and  $p_{\max} = \max\{p_i : i \in \mathbb{N}\}$ . Since  $\alpha \in (0, 1)$ , we can lower bound

$\hat{H}^\alpha(p)$  as

$$\begin{aligned} \hat{H}^\alpha(p) &= \frac{1}{1-\alpha} \log \left( \sum_{i=1}^{\infty} p_i^\alpha \right) \\ &= \frac{1}{1-\alpha} \log \left( p_{\max}^\alpha \sum_{i=1}^{\infty} \left( \frac{p_i}{p_{\max}} \right)^\alpha \right) \\ &\geq \frac{1}{1-\alpha} \log \left( p_{\max}^\alpha \sum_{i=1}^{\infty} \frac{p_i}{p_{\max}} \right) \\ &= -\log p_{\max}. \end{aligned}$$

Combing this bound with (51) yields (16).

Next we prove (17) by contradiction. If (17) does not hold, then there is a  $T > 0$  and a subsequence of  $(q_n)$ , which we also denote by  $(q_n)$ , such that

$$\lim_{n \rightarrow \infty} \frac{\mu(q_n^{-1}(q_n(p)))^\alpha}{\sum_{a \in q_n(\mathbb{R})} \mu(q_n^{-1}(a))^\alpha} = T. \quad (52)$$

We have

$$\begin{aligned} H_\mu^\alpha(q_n) &= \frac{1}{1-\alpha} \log \left( \frac{\sum_{a \in q_n(\mathbb{R})} \mu(q_n^{-1}(a))^\alpha}{\mu(q_n^{-1}(q_n(p)))^\alpha} \right) \\ &\quad + \frac{\alpha}{1-\alpha} \log(\mu(q_n^{-1}(q_n(p)))) \\ &\leq \frac{1}{1-\alpha} \log \left( \frac{\sum_{a \in q_n(\mathbb{R})} \mu(q_n^{-1}(a))^\alpha}{\mu(q_n^{-1}(q_n(p)))^\alpha} \right) \end{aligned} \quad (53)$$

where the inequality holds since  $\alpha \in (0, 1)$ . Because  $(q_n)$  is asymptotically optimal, we know that  $H_\mu^\alpha(q_n) \rightarrow \infty$  as  $n \rightarrow \infty$ . But the right hand side of (53) converges to a finite limit by assumption (52), a contradiction.

Also, (16) and an argument identical to the proof of (17) imply that for all  $p \in \mathbb{R}$  and  $i \in \{1, 2\}$ ,

$$\lim_{n \rightarrow \infty} \frac{\mu_i(q_n^{-1}(q_n(p)))^\alpha}{\sum_{a \in q_n(\mathbb{R})} \mu_i(q_n^{-1}(a))^\alpha} = 0$$

which completes the proof.  $\square$

*Lemma 2:* Assume  $\mu$  is a probability measure on  $\mathbb{R}^d$ , let  $r > 0$ , and let  $\|\cdot\|$  be any norm on  $\mathbb{R}^d$ . Suppose  $(q_n)$  is a sequence of  $d$ -dimensional vector quantizers (mappings  $q_n : \mathbb{R}^d \rightarrow \mathbb{R}^d$  with  $q_n(\mathbb{R})$  at most countable) such that

$$\lim_{n \rightarrow \infty} \int_{\mathbb{R}^d} \|x - q_n(x)\|^r \mu(dx) = 0.$$

Then

$$\lim_{n \rightarrow \infty} \max\{\mu(q_n^{-1}(a)) : a \in q_n(\mathbb{R}^d)\} = 0 \quad (54)$$

if and only if  $\mu$  is *nonatomic*, i.e.,  $\mu(\{x\}) = 0$  for all  $x \in \mathbb{R}^d$ .

*Proof:* If  $\mu(\{x\}) > 0$  for some  $x$ , then  $\mu(q_n^{-1}(q_n(x))) \geq \mu(\{x\})$  shows that (54) cannot hold. Now assume that  $\mu$  is nonatomic. We proceed indirectly to prove (54). Since  $\int \|x - q_n(x)\|^r \mu(dx) \geq \int_{q_n^{-1}(a)} \|x - q_n(x)\|^r \mu(dx)$  for all  $n$  and  $a \in q_n(\mathbb{R}^d)$ , if (54) does not hold, then (considering subsequences if necessary) there exist an  $\varepsilon > 0$ , points  $a_n \in \mathbb{R}^d$ , and measurable sets  $A_n \subset \mathbb{R}^d$ , such that

$$\lim_{n \rightarrow \infty} \int_{A_n} \|x - a_n\|^r \mu(dx) = 0, \quad \mu(A_n) \geq \varepsilon \text{ for all } n. \quad (55)$$

Let  $B(z, \delta) = \{x \in \mathbb{R}^d : \|x - z\| < \delta\}$  denote the open ball of radius  $\delta > 0$  centered at  $z \in \mathbb{R}^d$ . We have for all  $\delta > 0$ ,

$$\int_{A_n} \|x - a_n\|^r \mu(dx) \geq \delta^r \mu(A_n \setminus B(a_n, \delta))$$

which, combined with (55), implies  $\lim_n \mu(A_n \setminus B(a_n, \delta)) = 0$ . Thus for all  $\delta > 0$ ,

$$\liminf_{n \rightarrow \infty} \mu(B(a_n, \delta)) \geq \varepsilon.$$

This immediately implies that  $\{a_n : n \in \mathbb{N}\}$  is a bounded set, since  $\limsup_n \|a_n\| = \infty$  would yield  $\liminf_n \mu(B(a_n, \delta)) = 0$  because, as a probability measure on  $\mathbb{R}^d$ ,  $\mu$  is tight. Thus we can choose a subsequence of  $(a_n)$ , which we also denote by  $(a_n)$ , such that  $a_n \rightarrow a \in \mathbb{R}^d$  as  $n \rightarrow \infty$ . For this subsequence,  $B(a_n, \delta) \subset B(a, 2\delta)$  for all  $n$  large enough, implying, for all  $\delta > 0$ ,

$$\mu(B(a, 2\delta)) \geq \liminf_{n \rightarrow \infty} \mu(B(a_n, \delta)) \geq \varepsilon.$$

Since  $\mu(\{a\}) = \lim_{\delta \rightarrow 0} \mu(B(a, 2\delta))$ , we obtain  $\mu(\{a\}) \geq \varepsilon$ , which contradicts our assumption that  $\mu$  is nonatomic.  $\square$

*Lemma 3:* Let  $A \geq 0$ ,  $B \geq 0$ ,  $\gamma > 0$ , and define  $F : (0, 1) \rightarrow \mathbb{R}$  by

$$F(z) = \frac{A}{z^\gamma} + \frac{B}{(1-z)^\gamma}.$$

Then

$$\inf\{F(z) : z \in (0, 1)\} = \left(B^{\frac{1}{1+\gamma}} + A^{\frac{1}{1+\gamma}}\right)^{1+\gamma}.$$

If  $\min(A, B) > 0$ , then  $F(z_0) < F(z)$  for every  $z \in (0, 1) \setminus \{z_0\}$ , where

$$z_0 = \frac{A^{\frac{1}{1+\gamma}}}{A^{\frac{1}{1+\gamma}} + B^{\frac{1}{1+\gamma}}}.$$

*Proof:* The assertion is obvious for the cases  $A = 0$ ,  $B = 0$  or  $A + B = 0$ . Thus we can assume that  $A > 0$  and  $B > 0$ . But in this case the assertion follows from elementary calculus.  $\square$

A special case of the following lemma has already been used in [7]. For the reader's convenience we provide a detailed proof.

*Lemma 4:* Let  $r > 1$  and  $\alpha \in (0, 1)$ . Let  $E \subset \mathbb{R}$  be measurable. Then,

$$\begin{aligned} & \left(\int g^{\beta_1} d\lambda\right)^{\beta_2} \\ &= \inf \left\{ \frac{\left(\int_E g^{\beta_1} d\lambda\right)^{\beta_2}}{R^{\beta_2-1}} + \frac{\left(\int_{\mathbb{R} \setminus E} g^{\beta_1} d\lambda\right)^{\beta_2}}{(1-R)^{\beta_2-1}} : R \in (0, 1) \right\}. \end{aligned}$$

If  $\mu(E) \in (0, 1)$ , then

$$\left(\int g^{\beta_1} d\lambda\right)^{\beta_2} < \frac{\left(\int_E g^{\beta_1} d\lambda\right)^{\beta_2}}{R^{\beta_2-1}} + \frac{\left(\int_{\mathbb{R} \setminus E} g^{\beta_1} d\lambda\right)^{\beta_2}}{(1-R)^{\beta_2-1}}$$

for every  $R \in (0, 1) \setminus \{R_0\}$ , where  $R_0 = \int_E g^{\beta_1} d\lambda / \int_{\mathbb{R}} g^{\beta_1} d\lambda$ .

*Proof:* The assertion follows from Lemma 3 with

$$\gamma = \beta_2 - 1 = r/(1 - \alpha) > 0$$

and

$$A = \left(\int_E g^{\beta_1} d\lambda\right)^{\beta_2}, \quad B = \left(\int_{\mathbb{R} \setminus E} g^{\beta_1} d\lambda\right)^{\beta_2}$$

(note that  $A > 0$  and  $B > 0$  if  $\mu(E) \in (0, 1)$ ).  $\square$

## REFERENCES

- [1] J. Aczél and Z. Daróczy, *On Measures of Information and Their Characterizations*. London: Academic Press, 1975.
- [2] M. B. Baer, *Coding for General Penalties*. Ph.D. Thesis, Stanford University, June 2003.
- [3] W. R. Bennett, "Spectra of quantized signals," *Bell. Syst. Tech. J.*, vol. 27, pp. 446–472, 1948.
- [4] P. Billingsley, *Convergence of Probability Measures*. Wiley, 2nd ed., 1999.
- [5] J. A. Bucklew, "Companding and random quantization in several dimensions," *IEEE Trans. Inform. Theory*, vol. IT-27, pp. 207–211, Mar. 1981.
- [6] J. A. Bucklew and G. L. Wise, "Multidimensional asymptotic quantization theory with  $r$ th power distortion measures," *IEEE Trans. Inform. Theory*, vol. 28, pp. 239–247, Mar. 1982.
- [7] J. A. Bucklew, "Two results on the asymptotic performance of quantizers," *IEEE Trans. Inform. Theory*, vol. 30, pp. 341–348, March 1984.
- [8] L. L. Campbell, "A coding theorem and Rényi's entropy," *Information and Control*, vol. 8, pp. 423–429, 1965.
- [9] I. Csiszár, "Axiomatic characterizations of information measures," *Entropy*, vol. 10, no. 3, pp. 261–273, 2008.
- [10] A. Gersho, "Asymptotically optimal block quantization," *IEEE Trans. Inform. Theory*, vol. 25, pp. 373–380, Jul. 1979.
- [11] H. Gish and J. N. Pierce, "Asymptotically efficient quantizing," *IEEE Trans. Inform. Theory*, vol. 14, pp. 676–683, Sept. 1968.
- [12] S. Graf and H. Luschgy, *Foundations of Quantization for Probability Distributions*. Berlin, Heidelberg: Springer Verlag, 2000.
- [13] R. M. Gray, T. Linder, and J. T. Gill, "Lagrangian vector quantization with combined entropy and codebook size constraints," *IEEE Trans. Inform. Theory*, vol. 54, pp. 2220–2242, May 2008.
- [14] 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.
- [15] R. M. Gray and T. Linder, "Mismatch in high-rate entropy-constrained vector quantization," *IEEE Trans. Inform. Theory*, vol. 49, pp. 1204–1217, May 2003.
- [16] R. M. Gray and D. L. Neuhoff, "Quantization," *IEEE Trans. Inform. Theory*, (Special Commemorative Issue), vol. 44, pp. 2325–2383, Oct. 1998.
- [17] F. Jelinek, "Buffer overflow in variable length coding of fixed rate sources," *IEEE Trans. Inform. Theory*, vol. 14, pp. 490–501, May 1968.
- [18] W. Kreitmeier, "Optimal quantization for the one-dimensional uniform distribution with Rényi- $\alpha$ -entropy constraints," *Kybernetika*, vol. 46, pp. 96–113, 2010.
- [19] W. Kreitmeier, "Error bounds for high-resolution quantization with Rényi- $\alpha$ -entropy constraints," *Acta Math. Hungar.*, vol. 127, pp. 34–51, 2010.
- [20] W. Kreitmeier, "Optimal vector quantization in terms of Wasserstein distance," *Journal of Multivariate Analysis*, doi:10.1016/j.jmva.2011.04.005, vol. 102, no. 8, pp. 1225–1239, Sep. 2011.
- [21] W. Kreitmeier and T. Linder, "High-resolution scalar quantization with Rényi entropy constraint," *IEEE Trans. Inform. Theory*, vol. 57, no. 10, pp. 6837–6859, Oct. 2011.
- [22] S. P. Lloyd, Least squares quantization in PCM. *IEEE Transactions on Information Theory*, 1957. Bell Laboratories Technical Note. Reprinted in *IEEE Trans. on Inform. Theory*, vol. 28, pp. 127–135, Mar. 1982.
- [23] S. Na, "Variance-mismatched fixed-rate scalar quantization of Laplacian sources," *IEEE Trans. on Inform. Theory*, vol. 57, no. 7, pp. 4561–4572, Jul. 2011.
- [24] S. Na and D. L. Neuhoff, "Bennett's integral for vector quantizers," *IEEE Trans. Inform. Theory*, vol. 41, no. 4, pp. 886–900, Jul. 1995.
- [25] V. Misra, V. K. Goyal, and L. R. Varshney, Distributed scalar quantization for computing: High-resolution analysis and extensions," *IEEE Trans. Inform. Theory*, to appear, August 2011.
- [26] P. F. Panter and W. Dite, "Quantizing distortion in pulse-count modulation with nonuniform spacing of levels" *Proc. IRE*, vol. 39, pp. 44–48, Jan. 1951.

- [27] A. Rényi, "On measures of entropy and information." in *Proc. 4th Berkeley Symp. Math. Stat. Probab.*, vol. 1, pp. 547–561, 1960.
- [28] C. Tian and S. S. Hemami, "Universal multiple description scalar quantization: analysis and design," *IEEE Trans. Inform. Theory*, vol. 50, no. 9, pp. 2089–2102, Sep. 2004.
- [29] P. Zador, *Development and evaluation of procedures for quantizing multivariate distributions*. PhD thesis, Stanford Univ., 1963. Univ. Microfilm no. 64-9855.

## AUTHOR BIOGRAPHIES

**Wolfgang Kreitmeier** was born in Dachau, Germany, in 1972. He received the M.S. degree in physics from the Ludwig-Maximilians University of Munich in 1998, and the Ph.D degree in mathematics from the University of Passau in 2007. Currently he is supported by a grant from the German Research Foundation and works as a postdoctoral researcher at the University of Passau. His research interests include vector quantization and fractal geometry.

**Tamás Linder** (S'92-M'93-SM'00) was born in Budapest, Hungary, in 1964. He received the M.S. degree in electrical engineering from the Technical University of Budapest in 1988, and the Ph.D degree in electrical engineering from the Hungarian Academy of Sciences in 1992.

He was a post-doctoral researcher at the University of Hawaii in 1992 and a Visiting Fulbright Scholar at the Coordinated Science Laboratory, University of Illinois at Urbana-Champaign during 1993-1994. From 1994 to 1998 he was a faculty member in the Department of Computer Science and Information Theory at the Technical University of Budapest. From 1996 to 1998 he was also a visiting research scholar in the Department of Electrical and Computer Engineering, University of California, San Diego. In 1998 he joined Queen's University where he is now a Professor of Mathematics and Engineering in the Department of Mathematics and Statistics. His research interests include communications and information theory, source coding and vector quantization, machine learning, and statistical pattern recognition.

Dr. Linder received the Premier's Research Excellence Award of the Province of Ontario in 2002 and the Chancellor's Research Award of Queen's University in 2003. He was an Associate Editor for Source Coding of the IEEE Transactions on Information Theory in 2003–2004.