On the Structure of Optimal Entropy-Constrained Scalar Quantizers

András György Tamás Linder

IEEE Trans. on Information Theory
Submitted: Dec. 23, 2000, Revised: Aug. 16, 2001
Final Version: Sep. 6, 2001

Abstract

The nearest neighbor condition implies that when searching for a mean-square optimal fixed-rate quantizer it is enough to consider the class of regular quantizers, i.e., quantizers having convex cells and code points which lie inside the associated cells. In contrast, quantizer regularity can preclude optimality in entropy-constrained quantization. This can be seen by exhibiting a simple discrete scalar source for which the mean-square optimal entropy-constrained scalar quantizer (ECSQ) has disconnected (and hence nonconvex) cells at certain rates. In this work, new results concerning the structure and existence of optimal ECSQs are presented. One main result shows that for *continuous* sources and distortion measures of the form $d(x, y) = \rho(|x-y|)$, where ρ is a nondecreasing convex function, any finite-level ECSQ can be "regularized" so that the resulting regular quantizer has the same entropy and equal or less distortion. Regarding the existence of optimal ECSQs, we prove that under rather general conditions there exists an "almost regular" optimal ECSQ for any entropy constraint. For the squared error distortion measure and sources with piecewise monotone and continuous densities, the existence of a regular optimal ECSQ is shown.

Index Terms: convex distortion measures, entropy coding, optimal quantization, regular quantizers.

A. György is with the Department of Computer Science and Information Theory, Budapest University of Technology and Economics,, H-1521 Budapest, Hungary (email: `gyorgy@szit.bme.hu`). T. Linder is with the Department of Mathematics and Statistics, Queen's University, Kingston, Ontario, Canada K7L 3N6 (email: `linder@mast.queensu.ca`). This research was supported in part by the Natural Sciences and Engineering Research Council (NSERC) of Canada. András György also acknowledges the support of the Soros Foundation.

1 Introduction

The main objective of quantizer design is to find a collection of code points (the codebook) and associated quantization cells providing minimum distortion subject to a rate constraint. In fixed-rate quantization, where the quantizer's rate is measured by the log-cardinality of the codebook, the rate constraint means that the number of code points is fixed. In this case, efforts to design optimal quantizers have lead to the well-known necessary conditions of quantizer optimality (namely, the nearest neighbor and centroid conditions), first for scalar quantizers and the squared error distortion measure [1, 2], and subsequently for vector quantizers and more general distortion measures [3, 4]. An important consequence of the nearest neighbor condition is that an optimal fixed-rate quantizer is essentially determined by its code points since its cells are the Voronoi regions (with respect to the source distribution) associated with the code points. For the squared error distortion measure this implies that an optimal quantizer is *regular*, i.e, each of its cells is a convex set and the associated code point lies inside the cell. The cells of a regular scalar quantizer are intervals, and the cells of a regular vector quantizer with a finite number of code points are convex polytopes. In this sense, the structure of optimal fixed-rate quantizers for the squared error distortion measure (and to a certain extent for more general norm-based distortion measures [5]) is relatively well understood. Moreover, for reasonable distortion measures and source distributions, the distortion of a quantizer satisfying the nearest neighbor condition is a continuous function of its code points, and so the existence of optimal fixed-rate quantizers can be deduced using standard continuity-compactness arguments [6, 7, 8].

The average rate of a quantizer can further be reduced if a variable-rate lossless code (entropy code) is applied to its output. In this case the rate is usually defined as the entropy of the output of the quantizer [9] in order not to tie the performance of such a scheme to a particular entropy code, and the resulting scheme is called an entropy-constrained quantizer. The objective of the design is then to minimize the quantizer's distortion for a given entropy constraint, and a quantizer achieving this minimum distortion is called an optimal entropy-constrained quantizer. Since the nearest neighbor condition is no longer necessary for the optimality of an entropy-constrained quantizer, existence and structural problems concerning optimal quantization appear to be harder in this case. In contrast to fixed-rate quantization where particular attention has been payed to structural and existence issues, works on entropy-constrained quantization have focused more on design issues. For scalar sources, Berger [10] and Farvardin and Modestino [11] found necessary conditions for the optimality of a regular entropy-constrained scalar quantizer (ECSQ) with a fixed number of output points. These conditions give rise to practical algorithms for designing locally optimal ECSQs with a fixed number of code points [10, 12, 11, 13]. Chou *et al.* [14] gave an effective iterative descent algorithm using a Lagrangian formulation for the design of locally optimal entropy-constrained vector quantizers. Sufficient conditions for the existence of an optimal quantizer among all regular ECSQs with a fixed number of code points were given for sources with log-concave densities by Kieffer *et al.* [13]. It appears, however, that no general result concerning the existence of optimal entropy-constrained quantizers is known.

The assumption of quantizer regularity seems to be ubiquitous in the literature on entropy-constrained quantization. A regular quantizer with a finite number of code points can be described using a finite number of parameters, while a more general quantizer structure may not be described this way. Thus a fundamental question is whether it is sufficient to consider only regular quantizers when searching for a (mean-square) optimal entropy-constrained quantizer. In fact, this question can be answered in the negative by a simple example; there exists a discrete scalar source distribution and an interval of entropy constraints for which no quantizer with interval cells is optimal (see Example 1 in Section 3). One main contribution of this paper is to show that such a pathological example cannot exist when the source distribution is continuous; for such a source any "good" quantizer is essentially regular.

In a recent work Chou and Betts [15] showed that if an entropy-constrained quantizer is optimal and achieves the lower convex hull of $D_h(R)$, the lowest possible distortion of any quantizer with entropy not greater than R, then it satisfies a modified version of the nearest neighbor condition. For the squared error distortion measure, all quantizers satisfying this modified nearest neighbor condition can be shown to be regular. This result is very general in that it is valid for an arbitrary source distribution and quantizer dimension. On the other hand, it does not cover optimal quantizers that lie above the lower convex hull of $D_h(R)$, which can happen if $D_h(R)$ is not convex. (For example, $D_h(R)$ is not convex for a uniform scalar source and the squared error distortion measure [16].) Moreover, the result already presumes the optimality of the quantizer, but the achievability of $D_h(R)$ is an open issue.

Our purpose in this paper is to give new results on the structural properties and the existence of optimal entropy-constrained scalar quantizers. The paper is organized as follows. In Section 2 notation and definitions are introduced. In Section 3 regularity type properties of ECSQs are investigated. Throughout our basic assumption is that the source has a nonatomic distribution (i.e., its distribution function is continuous) and the distortion measure is of the form $d(x, y) = \rho(|x-y|)$, where ρ is a nondecreasing convex function. Theorem 1 shows that for any finite-point quantizer there is a regular quantizer with the same entropy and equal or less distortion. As a consequence, Corollary 2 shows that regular finite-point quantizers can perform arbitrarily close to the operational distortion-rate curve $D_h(R)$. Theorem 2 extends

Theorem 1 to infinite-point quantizers; it shows that any quantizer can be replaced with an "almost regular" quantizer, that is, with a quantizer which has interval cells but may be undefined on a set of probability zero.

In Section 4 existence results concerning optimal ECSQs are given. Theorem 3 proves that an optimal ECSQ achieving $D_h(R)$ always exists, and that such an optimal quantizer can be assumed to be almost regular. Theorem 4 shows that a regular optimal ECSQ exists among all quantizers having a fixed number of code points. In Section 5, for the squared error distortion measure and sources with densities, the almost regularity of an optimal ECSQ is strengthened to regularity. Theorem 5 shows the existence of regular optimal ECSQs for a wide class of source densities which contains all univariate densities commonly used as parametric source models. Concluding remarks are given in Section 6.

2 Preliminaries

An *N-point* (or N-level) *scalar quantizer* Q is a (Borel) measurable mapping of the real line $\mathbb{R}$ into a finite or countably infinite set of distinct reals $\mathcal{C} = \{c_i;\, i = 1, \ldots, N\}$ called the *codebook* of Q. In case the codebook is not finite, we formally define $N = \infty$ and call Q an infinite-point quantizer. The c_i are called the *code points* and the sets $S_i = \{x : Q(x) = c_i\}$ are called the *cells* (or decision regions) of Q. The codebook $\mathcal{C}$ and the collection of cells $\mathcal{S} = \{S_i;\, i = 1, \ldots N\}$ completely characterize Q since $\mathcal{S}$ is a partition of $\mathbb{R}$ and

$$Q(x) = c_i \quad \text{for} \quad x \in S_i.$$

The distortion of Q in quantizing a real random variable X with distribution μ is measured by the expectation

$$D(Q, \mu) = E\{d(X, Q(X))\} = \int_{\mathbb{R}} d(x, Q(x))\, d\mu(x)$$

where the *distortion measure* $d(\cdot, \cdot)$ is a nonnegative measurable function of two real variables. When the distribution μ is clear from the context, the short notation $D(Q)$ will be used. The *partial distortion* of the ith cell of Q is defined by

$$D_i(Q, \mu) = \int_{S_i} d(x, Q(x))\, d\mu(x) \tag{1}$$

so that $D(Q, \mu) = \sum_i D_i(Q, \mu)$. In case μ is an arbitrary finite measure, $D(Q, \mu)$ and the $D_i(Q, \mu)$ are still well defined by the corresponding integrals.

The entropy-constrained *rate* of Q is the entropy of the discrete random variable $Q(X)$:

$$H(Q) = -\sum_{i=1}^{N} P\{X \in S_i\} \log P\{X \in S_i\}$$

where log denotes base 2 logarithm. A scalar quantizer whose rate is measured by $H(Q)$ is called an *entropy-constrained scalar quantizer* (ECSQ).

Unless otherwise stated, we always assume that the cell probabilities $P\{X \in S_i\}$, $i = 1, \ldots, N$, are all positive. One can always redefine Q on a set of probability zero (by possibly reducing the number of cells) to satisfy this requirement.

For any $R \geq 0$ let $D_h(R)$ denote the lowest possible distortion of any quantizer with output entropy not greater than R. This function is formally defined by

$$D_h(R) = \inf\{D(Q) : H(Q) \leq R\}$$

where the infimum is taken over all finite or infinite-point scalar quantizers whose entropy is less than or equal to R. If there is no Q with finite distortion and entropy $H(Q) \leq R$, then we formally define $D_h(R) = \infty$. The existence of a "reference letter" $c^* \in \mathbb{R}$ such that $E\{d(X, c^*)\} < \infty$ is a sufficient (but not necessary) condition for $D_h(R)$ to be finite for all $R \geq 0$. Any Q that achieves $D_h(R)$ in the sense that $H(Q) \leq R$ and $D(Q) = D_h(R)$ is called an *optimal* ECSQ.

A scalar quantizer Q is called *regular* if (i) its cells are subintervals of the real line; (ii) each of its code points lies inside the associated cell; (iii) the collection of cells $\mathcal{S}$ is locally finite in the sense that the number of cells in $\mathcal{S}$ intersecting any bounded subset of the real line is finite. This definition reduces to the usual definition of regularity [9] if Q has a finite number of code points since in this case (iii) is automatically satisfied.

Let Q be a regular finite-point scalar quantizer with code points indexed so that $c_1 < c_2 < \cdots < c_N$. Then the corresponding interval cells $S_1, S_2, \ldots, S_N$ satisfy

$$S_1 < S_2 < \cdots < S_N$$

where, for $A, B \subset \mathbb{R}$, $A < B$ means that $a < b$ for all $a \in A$ and $b \in B$. Defining $q_0 = -\infty$ and $q_N = \infty$, the interval boundaries $q_i = \sup S_i$, $i = 1, \ldots, N-1$, satisfy

$$q_0 < c_1 \leq q_1 \leq \cdots \leq q_{N-1} \leq c_n < q_N$$

and

$$Q(x) = c_i \quad \text{if} \quad q_{i-1} < x < q_i.$$

The points q_i, $1 \leq i \leq N-1$, called the *subdivision points* or *thresholds* of Q, may belong to either S_i or S_{i+1}.

Similarly, if an infinite-point quantizer Q is regular, then its code points $\{c_i\}$ can be linearly ordered so that $c_i < c_j$ if $i < j$, where the index set is either the positive integers (if there is a smallest code point), or the negative integers (if there is a largest code point), or the set of all integers (if there are no smallest and largest code points), and the cells of Q are intervals with endpoints q_{i-1} and q_i such that $q_{i-1} \le c_i \le q_i$. Throughout the paper we assume that the source random variable X has a *nonatomic* distribution, i.e., $P\{X = x\} = 0$ for all $x \in \mathbb{R}$.[1] Thus each subdivision point q_i can be mapped arbitrarily to either c_i or c_{i+1} without changing the distortion and entropy of Q. Hence we adopt the convention that the bounded cells of a regular quantizer Q are of the form $S_i = [q_{i-1}, q_i)$. For the sake of unifying the notation, if Q has a leftmost cell, we sometimes formally extend the domain of Q to $\mathbb{R} \cup \{-\infty\}$ and write $[-\infty, q_1)$ instead of $(-\infty, q_1)$.

In what follows we often assume that d is a difference distortion measure of the form

$$d(x, y) = \rho(|x - y|) \tag{2}$$

where $\rho : [0, \infty) \to [0, \infty)$ is a nondecreasing function. Then the limit

$$\lim_{t\to\infty} \rho(t) = K$$

exists, and K is either finite or $K = \infty$. It will be convenient to formally extend the domain of ρ to $[0, \infty) \cup \{\infty\}$ and accordingly define

$$d(x, \infty) = d(x, -\infty) = \rho(\infty) = K \tag{3}$$

for all $x \in \mathbb{R}$.

It is easy to show that if ρ is also lower semicontinuous, then for any Borel set $S \subset \mathbb{R}$ such that

$$\inf_{c\in\mathbb{R}} \int_S d(x, c)\, d\mu(x) < \infty$$

there is a $\hat{c} \in \mathbb{R}$ (the "generalized centroid" of S) minimizing the distortion over S, i.e.,

$$\int_S d(x, \hat{c})\, d\mu(x) = \inf_{c\in\mathbb{R}} \int_S d(x, c)\, d\mu(x). \tag{4}$$

Since ρ is nondecreasing, if S is an interval, then there is a minimizing $\hat{c}$ that lies inside S. If ρ is strictly convex then $\hat{c}$ is unique for any S such that $\mu(S) > 0$ [4]. For example, $\hat{c} = E\{X|X \in S\}$ for the squared error distortion measure.

[1] Note that if the distribution of X is "continuous" in the sense that it has a probability density function, then it is necessarily nonatomic. On the other hand, there exist nonatomic distributions (such as the Cantor measure; see, e.g., [17]) that do not have densities.

3 Regularizing ECSQs

Consider a scalar quantizer Q with a finite codebook $\mathcal{C} = \{c_i;\ i = 1, \ldots, N\}$. For any source random variable X,

$$D(Q) = E\{d(X, Q(X))\} \geq E\{\min_{c_i \in \mathcal{C}} d(X, c_i)\}.$$

Thus if $\widehat{Q}$ is a quantizer that maps any input $x \in \mathbb{R}$ to a code point in $\mathcal{C}$ minimizing $d(x, c_i)$, $i = 1, \ldots, N$, then it has minimum distortion among all quantizers with codebook $\mathcal{C}$. This is the well-known nearest neighbor condition for fixed-rate quantizers. Although the cells $\{\widehat{S}_i;\ i = 1, \ldots, N\}$ of $\widehat{Q}$ are not uniquely defined, each $\widehat{S}_i$ satisfies $\widehat{S}_i \subset \{x : d(x, c_i) \leq d(x, c_j),\ j = 1, \ldots, N\}$. For the squared error distortion measure $d(x, y) = (x - y)^2$, or more generally for $d(x, y) = \rho(|x - y|)$, where ρ is nondecreasing, each $\widehat{S}_i$ is an interval and $c_i \in \widehat{S}_i$; hence $\widehat{Q}$ is regular. This means that any finite point quantizer can be "regularized" to obtain a regular quantizer with the same number of code points and equal or less distortion, and that it is enough to consider the much smaller, parametric class of N-point regular quantizers when searching for an optimal quantizer in the nonparametric family of all N-point quantizers.

Although the nearest neighbor condition is no longer valid for entropy-constrained quantizers, it was recently shown in [15] that a modified version of it still holds in the entropy-constrained setting.[2] If a quantizer Q with codebook $\{c_i\}$ and cells $\{S_i\}$ achieves the lower convex hull of $D_h(R)$ for a source X with distribution μ, then there is a $\lambda \geq 0$ (the negative slope of a line of support to the lower convex hull at R) such that for all i and μ-almost all $x \in S_i$,

$$d(x, c_i) - \lambda \log \mu(S_i) = \inf_j \{d(x, c_j) - \lambda \log \mu(S_j)\}. \tag{5}$$

For the squared error distortion (and more generally for rth power distortions), $\lambda > 0$ if $D_h(R) > 0$. In this case, the infimum in (5) is achieved by some index j, and (5) defines a quantizer with convex cells such that only finitely many cells intersect any bounded set [15]. Thus for the squared error distortion measure, any Q achieving the lower convex hull of $D_h(R)$ can be assumed to be regular. Moreover, it was shown in [15] that Q has a finite number of code points if the source distribution has sufficiently light tail. Condition (5) works for arbitrary source distribution and quantizer dimension, but it does not cover optimal quantizers that lie above the lower convex hull of $D_h(R)$, which can happen if $D_h(R)$ is not convex. For example, $D_h(R)$ for a uniform scalar source and the squared error

[2]A similar necessary condition for the optimality of variable-rate quantizers also appeared in [14].

distortion coincides with its lower convex hull only at rates $R = \log N$ for $N = 1, 2, \ldots$ [16]. In general, little is known about the properties of $D_h(R)$, and so the achievability of the lower convex hull is very difficult to check. Thus other methods are needed to find for each rate R a tractable, sufficiently small family of quantizers that still contains an optimal entropy-constrained quantizer. In the scalar case, such a family will turn out to be the class of regular quantizers if the source distribution is nonatomic. For discrete sources, however, it may happen that no regular ECSQ achieves the minimum mean squared distortion among all ECSQs satisfying a given entropy constraint.

Example 1 Let $d(x, y) = (x - y)^2$, and let X be a discrete random variable taking values in $\{-1, 0, 2\}$ with the following distribution:

$$P\{X = -1\} = P\{X = 2\} = 2/5 \quad \text{and} \quad P\{X = 0\} = 1/5.$$

A quantizer Q for X is in effect defined by a partition of the set $\{-1, 0, 2\}$ (the definition of Q for other values is immaterial) and by the corresponding code points (the optimal code points for each partition can easily be found). Checking the five possible partitions, it turns out that for all R such that

$$h_b(1/5) \le R < h_b(2/5)$$

where $h_b(x) = -x \log x - (1 - x) \log(1 - x)$, an optimal ECSQ has two code points $c_1 = 1/2$ and $c_2 = 0$, and the corresponding cells must satisfy

$$\{-1, 2\} \subset S_1 \quad \text{and} \quad \{0\} \subset S_2.$$

Thus if S_2 is an interval, then S_1 is a union of two disjoint intervals, each with positive probability. It follows that no regular quantizer can be optimal in this case.

The first result of this section shows that no such pathological example can exist if X has a nonatomic distribution. More specifically, we show that if X has a nonatomic distribution, then, for nondecreasing convex difference distortion measures, every finite-point ECSQ can be "regularized" so that the resulting regular quantizer has the same cell probabilities (and hence the same entropy) and equal or less distortion. The key to this result is the following lemma.

Lemma 1 *Assume that the random variable X has a nonatomic distribution μ, and let $d(x, y) = \rho(|x - y|)$, where $\rho : [0, \infty) \to [0, \infty)$ is convex and nondecreasing. Let Q be an arbitrary two-point quantizer with cells $\{S_1, S_2\}$ and corresponding code points $\{c_1, c_2\}$*

such that $c_1 < c_2$. Then there exists a two-point quantizer $\widehat{Q}$ with interval cells $\{\widehat{S}_1, \widehat{S}_2\}$ and corresponding code points $\{c_1, c_2\}$ such that $\widehat{S}_1 < \widehat{S}_2$, $\mu(S_i) = \mu(\widehat{S}_i)$, $i = 1, 2$, and $D(\widehat{Q}) \le D(Q)$.

Remark $\widehat{Q}$ has interval cells but it is not necessarily regular ($c_i \in \widehat{S}_i$ is not guaranteed by the construction). However, since ρ is nondecreasing, the centroid rule (4) guarantees that each c_i can be replaced by a $\hat{c}_i \in \widehat{S}_i$ such that the distortion is not increased.

Proof of Lemma 1 Since the distribution function of X is continuous, there is an $r \in \mathbb{R}$ such that $\mu((-\infty, r]) = \mu(S_1)$. Let $\widehat{Q}$ be the quantizer with cells $\widehat{S}_1 = (-\infty, r]$, $\widehat{S}_2 = (r, \infty)$, and code points $\{c_1, c_2\}$. We show that $D(\widehat{Q}) \le D(Q)$. If $D(Q)$ is not finite, we are done; so assume $D(Q) < \infty$. The key observation is that the function

$$\phi(x) = \rho(|x - c_1|) - \rho(|x - c_2|)$$

is nondecreasing. To show this, rewrite ϕ in the following form

$$\phi(x) = \begin{cases} \rho(c_1 - x) - \rho(c_2 - x) & \text{if } x < c_1 \\ \rho(x - c_1) - \rho(c_2 - x) & \text{if } c_1 \le x \le c_2 \\ \rho(x - c_1) - \rho(x - c_2) & \text{if } x > c_2 \end{cases}$$

Since ρ is nondecreasing, ϕ is clearly nondecreasing in the interval $[c_1, c_2]$, and the convexity of ρ readily implies that ϕ is also nondecreasing in $(-\infty, c_1]$ and $[c_2, \infty)$.

Since $\mu(\widehat{S}_1 \cap S_2) = \mu(S_1 \cap \widehat{S}_2)$ and $\widehat{S}_1 < \widehat{S}_2$,

$$\int_{\widehat{S}_1 \cap S_2} \phi(x)\, d\mu(x) \le \int_{S_1 \cap \widehat{S}_2} \phi(x)\, d\mu(x).$$

(It is easy to see that $D(Q) < \infty$ and the monotonicity of ρ imply that both integrals are finite.) Adding

$$\int_{S_1 \cap \widehat{S}_1} \rho(|x - c_1|)\, d\mu(x) + \int_{S_2 \cap \widehat{S}_2} \rho(|x - c_2|)\, d\mu(x)$$

to both sides and rearranging terms yields $D(\widehat{Q}) \le D(Q)$. □

We showed that convex and nondecreasing distortion measures allow regularizing any two-point quantizer if the source distribution is nonatomic. We call a distortion measure possessing this property two-point regular. More generally, given a positive integer N we say that a distortion measure d is *N-point regular* if for any nonatomic distribution μ and any N-point quantizer Q there is a *regular* N-point quantizer $\widehat{Q}$ such that $D(\widehat{Q}) \le D(Q)$ and

Q and $\widehat{Q}$ have the same cell probabilities (i.e., there is a one-to-one mapping π between the cells of Q and $\widehat{Q}$ satisfying $\mu(S) = \mu(\pi(S))$ for any cell S of Q). Finally, d is called *finitely regular* if it is N-point regular for every positive integer N.

Lemma 1 showed that nondecreasing and convex difference distortion measures are two-point regular. The next theorem shows that if d is two-point regular, then it is also finitely regular. The proof can be found in Appendix A

Theorem 1 *Every two-point regular distortion measure is finitely regular. That is, if X has a nonatomic distribution μ, and d is two-point regular, then for any finite-point quantizer Q with cells $\{S_1, \ldots, S_N\}$ there exists a regular quantizer $\widehat{Q}$ with interval cells $\{\widehat{S}_1, \ldots, \widehat{S}_N\}$ such that $\mu(S_i) = \mu(\widehat{S}_i)$, $i = 1, \ldots, N$, and $D(\widehat{Q}) \leq D(Q)$.*

Remarks (i) Note that not only do Q and $\widehat{Q}$ have equal entropies, but if the outputs of Q and $\widehat{Q}$ are encoded using the same variable-length code (e.g., a Huffman code [18]), then the resulting average codelengths will be equal. (ii) By Lemma 1 the theorem holds for any difference distortion measure $d(x, y) = \rho(|x - y|)$ with a nondecreasing and convex ρ. Although we have not yet found other examples for two-point regular distortion measures, we conjecture that this class is actually larger than the class of nondecreasing and convex difference distortion measures.

For $d(x, y) = \rho(|x - y|)$ with a nondecreasing and convex ρ, the construction of Lemma 1 does not change the code points of the quantizer, and it orders the cells according to the corresponding code points. It is easy to see that this order preserving construction is maintained in Theorem 1 if Lemma 1 is used in each step of the induction argument given in the proof. Thus we obtain the following corollary.

Corollary 1 *Assume that the real random variable X has a nonatomic distribution μ, and let $d(x, y) = \rho(|x - y|)$ where $\rho : [0, \infty) \to [0, \infty)$ is nondecreasing and convex. Then for any finite-point quantizer Q with cells $\{S_1, \ldots, S_N\}$ and corresponding code points $\{c_1, \ldots, c_N\}$ there exists a regular quantizer $\widehat{Q}$ with interval cells $\{\widehat{S}_1, \ldots, \widehat{S}_N\}$ such that $\mu(S_i) = \mu(\widehat{S}_i)$ for all $i = 1, \ldots, N$, $D(\widehat{Q}) \leq D(Q)$, and if $c_i < c_j$, then $\widehat{S}_i < \widehat{S}_j$.*

Remark It follows from Lemma 1 and the construction in the proof of Theorem 1 that $\widehat{Q}$ can also preserve the codepoints of Q, not only their order. In this case, $\widehat{Q}$ has interval cells, but it is not necessarily regular ($c_i \in \widehat{S}_i$ is not guaranteed).

Another immediate consequence of Theorem 1 is that $D_h(R)$ can be arbitrarily well approximated by regular finite-point ECSQs. This fact can be very useful in analyzing the high rate asymptotic behavior of $D_h(R)$ [19].

Corollary 2 *Assume X has a nonatomic distribution μ, and let d be a two-point regular distortion measure. If there is a $c^* \in \mathbb{R}$ such that $E\{d(X, c^*)\} < \infty$, then for any $R \geq 0$ and $\epsilon > 0$ there is a regular finite-point quantizer Q_ϵ such that $H(Q_\epsilon) \leq R$ and*

$$D(Q_\epsilon) < D_h(R) + \epsilon.$$

Remark The conditions of the corollary are satisfied for the squared error distortion measure if $E\{X^2\} < \infty$.

Proof of Corollary 2 First we show that for any quantizer Q with finite distortion, there is a sequence of regular finite-point quantizers $\{Q_n^*\}$ such that $H(Q_n^*) \leq H(Q)$ for all n and

$$\limsup_{n\to\infty} D(Q_n^*) \leq D(Q). \tag{6}$$

If Q is a finite point quantizer, then (6) follows by regularizing it using Theorem 1. Otherwise, assume Q has cells $\{S_1, S_2, \dots\}$ and corresponding code points $\{c_1, c_2, \dots\}$, and for $n \geq 2$ define the n-point quantizer Q_n to have cells $\{S_1, \dots, S_{n-1}, \bigcup_{i\geq n} S_i\}$ and code points $\{c_1, \dots, c_{n-1}, c^*\}$. It is clear that $H(Q_n) \leq H(Q)$. Moreover, Q_n and Q are identical on $C_n = \bigcup_{i=1}^{n-1} S_i$, and so

$$\lim_{n\to\infty} |D(Q_n) - D(Q)| = \lim_{n\to\infty} \left| \int_{\mathbb{R}\setminus C_n} d(x, c^*)\, d\mu(x) - \sum_{i\geq n} D_i(Q, \mu) \right| = 0$$

since $E\{d(X, c^*)\} < \infty$, $\{C_n\}$ increases to $\mathbb{R}$, and $D(Q) < \infty$. Since each Q_n is an n-point quantizer, Theorem 1 can be used to obtain a regular n-point quantizer Q_n^* such that $H(Q_n^*) = H(Q_n)$, and $D(Q_n^*) \leq D(Q_n)$. Hence (6) is proved. Now for any $\epsilon > 0$ choose a quantizer Q'_ϵ such that $H(Q'_\epsilon) \leq R$ and $D(Q'_\epsilon) - D_h(R) < \epsilon/2$. Then by (6) there is a regular finite-point quantizer Q_ϵ such that $H(Q_\epsilon) \leq H(Q'_\epsilon)$ and $D(Q_\epsilon) - D(Q'_\epsilon) < \epsilon/2$, which in turn satisfies $D(Q_\epsilon) < D_h(R) + \epsilon$. $\square$

Unexpected problems may arise if one wants to extend Theorem 1 to infinite-point quantizers. To illustrate the problem assume that the order preserving construction of Corollary 1 carries over to the regularization of infinite-point quantizers. In this case, if c_i and c_j are two code points of the initial infinite-point quantizer Q such that $c_i < c_j$, then the corresponding cells $\widehat{S}_i$ and $\widehat{S}_j$ of the regularized quantizer $\widehat{Q}$ satisfy $\widehat{S}_i < \widehat{S}_j$. As a result, $\widehat{Q}$ can have quite an unusual structure if the initial quantizer Q is arbitrary. For example, assume that Q is an infinite-point quantizer with code points $\{c_i\}_{i=1}^\infty$ such that if $c_i < c_j$, then there exists a c_k such that $c_i < c_k < c_j$. Then the collection of interval cells $\{\widehat{S}_i\}_{i=1}^\infty$ of the regularized $\widehat{Q}$ will have the property that for any $\widehat{S}_i$ and $\widehat{S}_j$ with $\widehat{S}_i < \widehat{S}_j$, there is an $\widehat{S}_k$ such that $\widehat{S}_i < \widehat{S}_k < \widehat{S}_j$.

It can be seen that such collection of intervals $\{\widehat{S}_i\}_{i=1}^{\infty}$ cannot form a partition of the real line.[3]

To deal with such cases, we introduce the notion of almost regular quantizers. Given finite measure μ, a quantizer Q is called *μ-almost regular* (or almost regular if μ is clear from the context) if there exists an $S \subset \mathbb{R}$ with $\mu(S) = 0$ such that Q is defined on $\mathbb{R} \setminus S$, and every cell of Q is an interval containing the associated code point. Thus an almost regular quantizer has interval cells but it may not be defined on a set of μ-measure zero. (We can define $Q(x)$ in an arbitrary manner for all $x \in S$ without changing the entropy and distortion of Q.) As an example, let $S \subset [0, 1]$ be the Cantor ternary set [17], and let μ be absolutely continuous with respect to the Lebesgue measure with $\mu([0, 1]) = 1$. Then $[0, 1] \setminus S$ is a union of countably many open intervals which form the cells of a μ-almost regular quantizer. As this example shows, an almost regular quantizer can have infinitely many cells in a bounded interval. Also note that the quantizer of this example cannot be (re)defined on a set of probability zero to obtain a quantizer with interval cells.

This definition can be extended similarly to higher dimensions. A k-dimensional vector quantizer Q is called *μ-almost regular* if there is a set $S \subset \mathbb{R}^k$ such that $\mu(S) = 0$, Q is defined on $\mathbb{R} \setminus S$, and Q has convex cells containing the corresponding code points. To exhibit an example for a μ-almost regular vector quantizer, we can use a corollary of the Vitali covering theorem [20]. The corollary states that there exists a countable collection $\{B_1, B_2, \dots\}$ of disjoint closed balls in the unit cube $[0, 1]^k$ of $\mathbb{R}^k$ such that the set $S = [0, 1]^k \setminus \left(\bigcup_i B_i\right)$ has Lebesgue measure zero. Now let μ be absolutely continuous with respect to the k-dimensional Lebesgue measure with $\mu([0, 1]^k) = 1$. Then $\{B_1, B_2, \dots\}$ form the cells of an almost regular k-dimensional quantizer.

Remark If Q is a μ-almost regular scalar quantizer with cells $\{S_i\}$ such that only finitely many of the S_i intersect any bounded interval, then Q can be redefined on a set of μ-measure zero to obtain a regular quantizer with the same distortion and entropy. This follows since in this case there is an indexing of the cells such that $S_i < S_{i+1}$ for all i, and now the set of points (of measure zero) lying between S_i and S_{i+1} where Q is possibly not defined can be assigned to (say) S_i to obtain interval cells whose union covers the real line.

The last result of this section shows that for two-point regular difference distortion measures, any infinite-point quantizer can be replaced by an almost regular quantizer with the same entropy and equal or less distortion. The theorem is proved in Appendix A.

[3] Let $\widehat{S}_i^{\circ}$ denote the interior of $\widehat{S}_i$. Then $\mathbb{R} \setminus \left(\bigcup_i \widehat{S}_i^{\circ}\right)$ is nonempty and perfect (since $\widehat{S}_i^{\circ}$ and $\widehat{S}_j^{\circ}$ cannot have a common endpoint if $i \neq j$), and thus uncountable. Therefore $\mathbb{R} \setminus \left(\bigcup_i \widehat{S}_i\right)$ is nonempty and uncountable.

Theorem 2 *Assume that the random variable X has a nonatomic distribution μ, and let $d(x,y) = \rho(|x-y|)$ be a two-point regular distortion measure, where $\rho : [0,\infty) \to [0,\infty)$ is nondecreasing and left-continuous. Then for any infinite-point quantizer Q there exists an almost regular quantizer $\widehat{Q}$ with the same cell probabilities such that $D(\widehat{Q}) \le D(Q)$.*

4 The existence of optimal ECSQs

The regularization result of Theorem 2 makes it possible to show the existence of an optimal ECSQ for any source with a nonatomic distribution. Theorem 2 also implies that such an optimal ECSQ can be assumed to be almost regular.

Theorem 3 *Let X have a nonatomic distribution μ and let $d(x,y) = \rho(|x-y|)$ be a two-point regular distortion measure, where $\rho : [0,\infty) \to [0,\infty)$ is nondecreasing and left-continuous. Then for any $R \ge 0$ there exists a μ-almost regular quantizer Q such that $H(Q) \le R$ and $D(Q) = D_h(R)$.*

Remarks (i) The conditions on d are satisfied when ρ is convex and nondecreasing; hence the result holds for the squared error distortion measure. (ii) This theorem would imply the existence of a regular optimal ECSQ if one could directly prove that an optimal ECSQ cannot have infinitely many cells in a bounded interval. While this is a reasonable conjecture under general conditions, we have only been able to prove it for the squared error distortion measure and sources with well-behaved densities (see Section 5).

Proof of Theorem 3 Fix $R \ge 0$ and assume $D_h(R)$ is finite; otherwise the statement is trivial. Consider a sequence of quantizers $\{Q_n\}$ such that $H(Q_n) \le R$ for all n and

$$\lim_{n\to\infty} D(Q_n) = D_h(R).$$

By Theorem 2 we can assume that each Q_n is almost regular. For positive integers n and i let $S_i^{(n)} = [a_i^{(n)}, b_i^{(n)})$ denote the cell of Q_n with the ith largest probability, let $c_i^{(n)}$ denote the corresponding code point, and let $p_i^{(n)} = \mu(S_i^{(n)})$. In case of ties, any ordering of the equiprobable cells suffices. If Q_n has k cells, then $S_i^{(n)}$ is formally defined to be empty for all $i > k$. For every n, let $\mathbf{p}_n = (p_1^{(n)}, p_2^{(n)}, \ldots)$. Then $\mathbf{p}_n \in C$, where

$$C = \left\{ (p_1, p_2, \ldots) : \, p_i \ge 0 \text{ for all } i, \, p_1 \ge p_2 \ge \cdots, \, \sum_{i=1}^{\infty} p_i = 1, \, -\sum_{i=1}^{\infty} p_i \log p_i \le R \right\}.$$

Now for any $\mathbf{p} = (p_1, p_2, \dots) \in C$ let $Y_\mathbf{p}$ be a positive integer valued random variable with distribution $P\{Y_\mathbf{p} = i\} = p_i$, $i \geq 1$. Then by a result of Wyner [21, Theorem 1] we have

$$E\{\log Y_\mathbf{p}\} \leq -\sum_{i=1}^{\infty} p_i \log p_i \leq R.$$

Thus for any $\mathbf{p} \in C$, Markov's inequality implies

$$P\{Y_\mathbf{p} > n\} = P\{\log Y_\mathbf{p} > \log n\} \leq \frac{E\{\log Y_\mathbf{p}\}}{\log n} \leq \frac{R}{\log n}.$$

Hence the family of distributions corresponding to C is tight. Therefore, by Prokhorov's theorem [17], $\{\mathbf{p}_n\}$ has a pointwise convergent subsequence, denoted also by $\{\mathbf{p}_n\}$, which converges to some probability distribution $\mathbf{p} = (p_1, p_2, \dots)$. Clearly, $p_1 \geq p_2 \geq \cdots$, and by Fatou's lemma, $-\sum_{i=1}^{\infty} p_i \log p_i \leq R$ (thus $\mathbf{p} \in C$, implying that C is compact under pointwise convergence). Now for the subsequence of quantizers $\{Q_n\}$ corresponding to $\{\mathbf{p}_n\}$, form the vectors

$$\mathbf{v}_n = (a_1^{(n)}, b_1^{(n)}, c_1^{(n)}, a_2^{(n)}, b_2^{(n)}, c_2^{(n)}, \dots).$$

If S_i is empty, define $a_i = b_i = c_i = \infty$. By Cantor's diagonalization method a subsequence $\{\mathbf{v}_{n_j}\}$ converging pointwise to a vector $\mathbf{v} = (a_1, b_1, c_1, a_2, b_2, c_2, \dots)$ can be chosen (convergence to ∞ or $-\infty$ is also allowed). For this vector we can construct a quantizer Q with code points c_i and corresponding cells $S_i = [a_i, b_i)$. If $a_i = b_i$ (in this case also $c_i = a_i$ by construction), then S_i is empty by definition. Since for every fixed n the intervals $S_i^{(n)} = [a_i^{(n)}, b_i^{(n)})$, $i = 1, 2, \dots$, are pairwise disjoint, it is easy to see that the intervals $S_i = [a_i, b_i)$, $i = 1, 2 \dots$, are also pairwise disjoint. Since μ is nonatomic, $\mu(S_i) = \lim_n \mu([a_i^{(n)}, b_i^{(n)})) = p_i$ for all i, and hence $\sum_i \mu(S_i) = 1$. Thus if Q is defined to have cells $\{S_i\}$ and code points $\{c_i\}$ (note that just as in the proof of Theorem 2, some of the c_i may not be finite), then since $\mathbf{p} \in C$, we have

$$H(Q) \leq R.$$

On the other hand, Lemma 4 in Appendix B shows that $\lim_n a_i^{(n)} = a_i$, $\lim_n b_i^{(n)} = b_i$, and $\lim_n c_i^{(n)} = c_i$ imply $D(Q) \leq \liminf_n D(Q_n) = D_h(R)$. Now we can use the centroid rule (4) to replace any nonfinite c_i by a finite one that lies inside its associated cell. The modified quantizer Q is μ-almost regular and achieves $D_h(R)$. $\square$

The following result shows that for any finite N there is an optimal quantizer among all ECSQs with no more than N code points.

Theorem 4 *Assume that the random variable X has a nonatomic distribution μ, and let $d(x, y) = \rho(|x - y|)$ be a two-point regular distortion measure, where $\rho : [0, \infty) \to [0, \infty)$ is nondecreasing and left-continuous. Then for any $R \geq 0$ and positive integer N, there exists a regular quantizer with at most N points achieving*

$$D_{h,N}(R) = \inf\{D(Q) : Q \text{ has at most } N \text{ code points}, H(Q) \leq R\}.$$

Remark Note that the quantizer achieving $D_{h,N}(R)$ may have less than N code points. For the squared error distortion case and for sources with log-concave densities, Kieffer *et al.* [13] provided conditions under which $D_{h,N}(R)$ is achieved by a quantizer having exactly N code points.

Proof of Theorem 4 The proof of Theorem 3 is used with a slight modification. By Theorem 1 there exists a sequence of regular quantizers $\{Q_n\}$, each having no more than N codepoints, such that $H(Q_n) \leq R$ for all n and $\lim_n D(Q_n) = D_{h,N}(R)$. Then the construction in the proof of Theorem 3 yields an almost regular quantizer Q with at most N codepoints. Since $H(Q) \leq R$ and $D(Q) \leq \lim_n D(Q_n)$, Q can be redefined to obtain a regular quantizer which achieves $D_{h,N}(R)$. □

We conclude this section with a short discussion on the (lack of) stability of $D_h(R)$. For two probability distributions μ and ν with $\int x^2\, d\mu(x) < \infty$ and $\int y^2\, d\nu(y) < \infty$ define

$$\Delta(\mu, \nu) = \inf\big(E\{(X - Y)^2\}\big)^{1/2}$$

where the infimum is over all joint distributions of the pairs of random variables (X, Y) such that X has distribution μ and Y has distribution ν. Then Δ is a metric on probability distributions with finite second moments which has been widely used in fixed-rate quantization (see, e.g., [22, 6, 5]). For the squared error distortion measure, optimal fixed-rate quantizer performance is a continuous function of the source distribution in this metric [6], that is, letting $D_f(N, \mu)$ denote the minimum mean squared distortion for a source with distribution μ of any quantizer with N code points, one has $D_f(N, \mu_n) \to D_f(N, \mu)$ for any sequence of source distributions $\{\mu_n\}$ with $\Delta(\mu, \mu_n) \to 0$. The convergence in Δ is easy to characterize ($\Delta(\mu, \mu_n) \to 0$ if and only if $\int x^2\, d\mu_n(x) \to \int x^2\, d\mu(x)$ and $\mu_n \to \mu$ weakly [6]), and the continuity of $D_f(N, \mu)$ in μ is an important tool in proving consistency and convergence rate results for empirical quantizer design. In particular, if a quantizer Q_n is optimal for μ_n, and μ_n and μ are close in Δ, then Q_n is nearly optimal for μ.

One can ask whether an analogous stability result holds for $D_h(R, \mu)$, the minimum ECSQ distortion (here we have made explicit the dependence of $D_h(R)$ on μ). It is easy to see that the entropy $H(Q, \mu)$ of a fixed quantizer Q need not be a continuous function of the source distribution μ. The following simple example uses this fact to demonstrate that $D_h(R, \mu)$ is not continuous in μ.

Example 2 Let μ denote the distribution of X in Example 1, and let μ_n, $n = 1, 2, \ldots$, be defined by the density f_n such that $f_n(x) = 2n/5$ if $|x+1| < \frac{1}{2n}$ or $|x-2| < \frac{1}{2n}$, $f_n(x) = n/5$ if $|x| < \frac{1}{2n}$, and f_n is zero otherwise. Then $\Delta(\mu, \mu_n) \to 0$. Let the two-point quantizer Q be defined by $Q(x) = -2/3$ if $x < 2$ and $Q(x) = 2$ if $x \geq 2$. Consider the squared error distortion measure. Then $H(Q, \mu_n) = h_b(1/5)$ for all n, implying

$$D_h(h_b(1/5), \mu_n) \leq D(Q, \mu_n).$$

Also, it is easy to see that

$$\lim_{n\to\infty} D(Q, \mu_n) = D(Q, \mu) = \frac{2}{15}.$$

Since $D_h(h_b(1/5), \mu) = 1.8$ from Example 1, we obtain

$$\limsup_{n\to\infty} D_h(h_b(1/5), \mu_n) < D_h(h_b(1/5), \mu).$$

5 Optimality and regularity for sources with densities

In the previous section we showed the existence of almost regular optimal ECSQs for nonatomic source distributions and a wide class of difference distortion measures. We can obtain the stronger result that a regular optimal ECSQ exists if we restrict the attention to the squared error distortion measure and sources with well-behaved densities. For convenience, we assume in this section that the source density is supported in a subinterval (σ, τ) of the real line $\mathbb{R}$, where we allow the possibility that $\sigma = -\infty$ and $\tau = \infty$. Accordingly, quantizers need only be defined on (σ, τ).

A function $f : (\sigma, \tau) \to \mathbb{R}$ is called *piecewise monotone* if (σ, τ) can be partitioned into countably many intervals such that any bounded set intersects only a finite number of these intervals and f is monotone in each of these intervals. *Piecewise continuity* is similarly defined with continuity in place of monotonicity. All continuous unimodal densities are of course piecewise monotone and piecewise continuous, and all densities commonly used in source modeling belong to this class, including the generalized Gaussian, Cauchy, and beta densities [11].

The following theorem resolves the problem of the regularity of optimal ECSQs in the special, but important case of the squared error distortion measure and piecewise monotone and piecewise continuous source densities.

Theorem 5 *Let X be a random variable with a density f which is piecewise monotone and piecewise continuous in (σ, τ). Assume that $d(x, y) = (x - y)^2$. Then for any entropy constraint $R \geq 0$ there is an optimal ECSQ which is regular.*

To prove this result, we need two useful technical lemmas. The proofs of these are deferred to Appendix C. The first lemma is an extension of a result of Kieffer *et al.* [13, Lemma A.5].

Lemma 2 *Let X be a random variable with a density function f which is continuous, positive, and nonincreasing (resp. nondecreasing) on (σ, τ), and $E\{X^2\} < \infty$. Assume $d(x, y) = (x - y)^2$. Then for any scalar quantizer Q there exists a quantizer $\widehat{Q}$ satisfying*

(a) *Q and $\widehat{Q}$ have the same cell probabilities and $D(\widehat{Q}) \leq D(Q)$;*

(b) *$\widehat{Q}$ has interval cells $\{\widehat{S}_i\}$ such that $\widehat{S}_1 < \widehat{S}_2 < \cdots$ (resp. $\widehat{S}_1 > \widehat{S}_2 > \cdots$);*

(c) *The cell probabilities $\mu(\widehat{S}_i)$ are nonincreasing as i increases.*

The next lemma is a simple consequence of a necessary condition for the optimality of a finite-point regular ECSQ due to Farvardin and Modestino [11], who generalized a similar result of Berger [10] from the squared error distortion to more general distortion measures.

Lemma 3 *Let X have a density f which is positive and continuous in (σ, τ) and let $d(x, y) = \rho(|x - y|)$, where $\rho : [0, \infty) \to [0, \infty)$ is nondecreasing and continuous. Assume Q is an ECSQ with finite distortion that is optimal for some entropy constraint $R > 0$. If Q has code points $\{c_i\}$ and interval cells $\{S_i = [q_{i-1}, q_i)\}$, where $q_{i-1} < q_i$ for all i, then there is a $\lambda \geq 0$ such that for all i,*

$$\rho(c_{i+1} - q_i) - \lambda \log p_{i+1} = \rho(q_i - c_i) - \lambda \log p_i \tag{7}$$

where $p_i = \mu(S_i)$.

Proof of Theorem 5 Denote the distribution of X by μ and assume $D_h(R) < \infty$; otherwise the result trivially holds. By Theorem 3 for any rate constraint $R \geq 0$ there is a μ-almost regular optimal quantizer Q. If only finitely many cells of Q intersect any bounded interval in (σ, τ), then Q can be redefined on a set of μ-measure zero to obtain a regular quantizer (see the remark preceding Theorem 2), and the theorem holds.

Assume now that Q has an infinite number of cells in some bounded interval. Since f is piecewise monotone and piecewise continuous, it is easy to see that there is an interval partition $\{I_i\}$ of (σ, τ) such that only finitely many of the I_i intersect any bounded subset of (σ, τ), and f is continuous and monotone in the interior of each I_i. Then there is a bounded interval J such that $J \subset I$ for some $I \in \{I_i\}$, J contains infinitely many cells, say $\{S_i\}$, of Q, and the intersection of J with any cell of Q not contained in J has μ-measure zero. Thus we can define a new quantizer $\tilde{Q}$ on J to have cells $\{S_i\}$ with the corresponding codepoints. If $\tilde{f}$ denotes the conditional density of X on J, and $\tilde{\mu}$ is the corresponding distribution, then $\tilde{Q}$ is $\tilde{\mu}$-almost regular, and $\tilde{f}$ is positive, continuous, and monotone on J. Assume $\tilde{f}$ is nonincreasing; the argument is similar for nondecreasing $\tilde{f}$. Then by Lemma 2 there is a $\widehat{Q}$ with $D(\widehat{Q},\tilde{\mu}) \le D(\tilde{Q},\tilde{\mu})$, such that $\tilde{Q}$ and $\widehat{Q}$ have the same cell probabilities, $\widehat{Q}$ has interval cells with subdivision points $q_0 < q_1 < q_2 < \cdots$ such that the cell probabilities $\tilde{\mu}([q_{i-1}, q_i))$ are nonincreasing as i increases, and $q_0 = a$, the left endpoint of J. Note that the condition $\int_J x^2 \tilde{f}(x)\,dx < \infty$ required in Lemma 2 is automatically satisfied since J is bounded. Since Q was optimal, $\widehat{Q}$ must be optimal for $\tilde{\mu}$ and the entropy constraint $H(\widehat{Q},\tilde{\mu})$. Denoting by c_i the code point for the cell $[q_{i-1}, q_i)$, by Lemma 3 there is a $\lambda \ge 0$ such that for all $i \ge 1$ we have

$$\rho(c_{i+1} - q_i) - \lambda \log p_{i+1} = \rho(q_i - c_i) - \lambda \log p_i \tag{8}$$

where $p_i = \tilde{\mu}([q_{i-1}, q_i))$. If $\lambda > 0$, we follow an idea of [15] to show that only finitely many p_i can be nonzero, which contradicts the assumption that $\widehat{Q}$ has an infinite number of cells in J. Notice that (8) implies that for any $i \ge 2$,

$$\begin{aligned} -\lambda \log p_i &= \sum_{j=1}^{i-1} \Big(\rho(q_j - c_j) - \rho(c_{j+1} - q_j) \Big) - \lambda \log p_1 \\ &\le \sum_{j=1}^{i-1} \rho(q_j - c_j) - \lambda \log p_1 \\ &\le \rho(\ell) - \lambda \log p_1 \end{aligned}$$

where ℓ denotes the length of J and the last inequality holds because $\rho(t) = t^2$ $(t \ge 0)$ is convex and nondecreasing. Therefore, if $\lambda > 0$, then $-\log p_i$ is bounded from above; hence there is an $\epsilon > 0$ such that $p_i > \epsilon$ for all i. But then $\sum_{i=1}^{\infty} p_i = \infty$, which is impossible. Thus $\lambda = 0$. Then (8) implies that the cells $[q_{i-1}, q_i)$ satisfy the standard nearest neighbor condition. Since $d(x, y) = (x - y)^2$, the optimal code point c_i corresponding to $[q_{i-1}, q_i)$ is the centroid $E\{X|X \in [q_{i-1}, q_i)\}$, and since $\tilde{f}$ is nonincreasing on J, $c_i - q_{i-1} \le q_i - c_i$. By the nearest neighbor condition, $c_{i+1} - q_i = q_i - c_i$, and so $q_i - q_{i-1} \le q_{i+1} - q_i$ for all i. That is, the cell lengths increase with i. Therefore, $\lim_i q_i = \infty$, contradicting the fact the J is bounded and $q_i \in J$ for all $i \ge 1$. Thus only finitely many cells of Q can intersect any bounded interval. $\square$

6 Conclusion

In this paper we presented new results on the structure and existence of optimal ECSQs. First we considered the problem of regularization. For nonatomic source distributions and a wide class of difference distortion measures we showed that for any finite-point ECSQ there is a regular ECSQ with the same entropy and equal or less distortion. As a consequence, we showed (under the standard assumption that there is a reference point with finite expected distortion) that regular finite-point ECSQs can perform arbitrarily close to the operational distortion-rate curve $D_h(R)$. We introduced the notion of an almost regular quantizer (such quantizers have convex cells but may be undefined for a set of input points of probability zero), and we showed that any infinite-point quantizer can be replaced with an almost regular quantizer that has the same entropy and equal or less distortion.

Using the technique of regularization we gave results concerning the existence of optimal ECSQs. We proved that there exists an almost regular ECSQ that achieves the operational distortion-rate function $D_h(R)$. This result basically settles the problem of existence of optimal entropy-constrained quantizers in the scalar case. We also showed that a regular optimal ECSQ exists among all ECSQs having no more than a given finite number of code points. For the squared error distortion measure and a wide class of sources with well-behaved densities we showed that $D_h(R)$ can be achieved by a regular ECSQ.

It is of interest to extend these results to entropy-constrained vector quantizers (ECVQs). The techniques used in this paper rely rather heavily on the fact that the set of reals can be linearly ordered in a natural way, and the proofs do not easily carry over to the vector case. For example, the existence of optimal ECVQs is an open problem. Another interesting (and perhaps easier) problem is to show that for the squared error distortion measure, the kth order entropy-constrained operational distortion-rate function, $D_h^{(k)}(R)$, can be arbitrarily well approached by regular k-dimensional ECVQs. Such a result would help tie up some loose ends in asymptotic (high-rate) quantization theory [23, 24].

Appendix A

Proof of Theorem 1 Assume $D(Q)$ is finite; otherwise the statement is trivial. The proof uses induction on N. A one-point quantizer is always regular; hence d is always one-point regular. Also, d is two-point regular by assumption. Now assume that d is n-point regular for all $n \le N-1$, where $N \ge 3$. We show that then d is also N-point regular.

Assume without loss of generality that the indexing of the cells of Q is such that

$$\mu(S_1) = \min_{1 \le i \le N} \mu(S_i) \tag{9}$$

and for $i = 1, \dots, N$, let c_i denote the code point corresponding to S_i.

The main idea in constructing $\widehat{Q}$ is the following: First we fix S_1 and, using the induction hypothesis, redefine the cells $S_2, \dots, S_N$ such that the new cells are "intervals" in $\mathbb{R} \setminus S_1$. That is, the new cells satisfy $S_1^{(1)} = S_1$ and $S_2^{(1)} < \cdots < S_N^{(1)}$. Then we redefine $S_1^{(1)}$ and $S_2^{(1)}$ such that the quantizer thus obtained have a leftmost cell which is a proper interval. Finally, the other $N-1$ cells are replaced by intervals using the induction hypothesis.

Step 1. Let μ_1 be the restriction of μ to $\mathbb{R} \setminus S_1$, that is, $\mu_1(B) = \mu(B \cap (\mathbb{R} \setminus S_1))$ for any Borel set $B \subset \mathbb{R}$. Furthermore, let the $(N-1)$-point quantizer q_1 be defined to have cells $\{S_1 \cup S_2, S_3, \dots, S_N\}$ and code points $\{c_2, \dots, c_N\}$. Then

$$D(Q, \mu) = D_1(Q, \mu) + D(q_1, \mu_1) \tag{10}$$

(recall definition (1)). Since μ_1 is nonatomic, the induction hypothesis implies that there is a regular quantizer $\widehat{q}_1$ with interval cells $T_1^{(1)} < \cdots < T_{N-1}^{(1)}$ and code points $\{e_1^{(1)}, \dots, e_{N-1}^{(1)}\}$ such that

$$D(\widehat{q}_1, \mu_1) \le D(q_1, \mu_1) \tag{11}$$

and the cells of q_1 and $\widehat{q}_1$ have the same measures according to μ_1. Now define Q_1 to have cells $\{S_1^{(1)}, \dots, S_N^{(1)}\}$ and code points $\{c_1^{(1)}, \dots, c_N^{(1)}\}$ such that

$$S_i^{(1)} = \begin{cases} S_1 & \text{if } i = 1 \\ T_{i-1}^{(1)} \cap (\mathbb{R} \setminus S_1) & \text{if } 2 \le i \le N \end{cases} \quad \text{and} \quad c_i^{(1)} = \begin{cases} c_1 & \text{if } i = 1 \\ e_{i-1}^{(1)} & \text{if } 2 \le i \le N. \end{cases}$$

From the construction of Q_1 it follows that Q and Q_1 have the same cell probabilities according to μ,

$$S_2^{(1)} < \cdots < S_N^{(1)} \tag{12}$$

and

$$D_i(Q_1, \mu) = \begin{cases} D_1(Q, \mu) & \text{if } i = 1 \\ D_{i-1}(\widehat{q}_1, \mu_1) & \text{if } 2 \le i \le N. \end{cases}$$

Thus (10) and (11) imply that

$$D(Q_1, \mu) \le D(Q, \mu).$$

Step 2. Let μ_2 be the restriction of μ to $S_1^{(1)} \cup S_2^{(1)}$, and let the two-point quantizer q_2 have cells $\{S_1^{(1)}, \mathbb{R} \setminus S_1^{(1)}\}$ and code points $\{c_1^{(1)}, c_2^{(1)}\}$. Then, since d is assumed to be two-point regular, there is a regular quantizer $\widehat{q}_2$ with interval cells $T_1^{(2)} < T_2^{(2)}$ and code points $\{e_1^{(2)}, e_2^{(2)}\}$ such that $D(\widehat{q}_2, \mu_2) \le D(q_2, \mu_2)$ and $\widehat{q}_2$ and q_2 have the same cell measures according to μ_2. We assume (as we may without loss of generality) that $T_1^{(2)} = (-\infty, t)$ and $T_2^{(2)} = [t, \infty)$, where $t = \inf\{x : \mu_2((-\infty, x)) = \mu_2(T_1^{(2)})\}$.

Since either $\mu_2(T_1^{(2)}) = \mu(S_1^{(1)})$ or $\mu_2(T_1^{(2)}) = \mu(S_2^{(1)})$, by (9) we have $\mu_2(T_1^{(2)}) \le \mu(S_2^{(1)})$. Thus $t \le \sup S_2^{(1)}$, because $S_2^{(1)} \subset (-\infty, \sup S_2^{(1)}]$, and so

$$\mu_2((-\infty, \sup S_2^{(1)}]) \ge \mu_2(S_2^{(1)}) = \mu(S_2^{(1)}) \ge \mu_2(T_1^{(2)}) = \mu_2((-\infty, t]).$$

Therefore we can define the quantizer Q_2 with cells and code points

$$S_i^{(2)} = \begin{cases} (-\infty, t) & \text{if } i = 1 \\ [t, \infty) \cap (S_1^{(1)} \cup S_2^{(1)}) & \text{if } i = 2 \\ S_i^{(1)} & \text{if } 3 \le i \le N \end{cases} \quad \text{and} \quad c_i^{(2)} = \begin{cases} e_1^{(2)} & \text{if } i = 1 \\ e_2^{(2)} & \text{if } i = 2 \\ c_i^{(1)} & \text{if } 3 \le i \le N. \end{cases}$$

Then, as in *Step 1*, it is easy to see that

$$D(Q_2, \mu) \le D(Q_1, \mu)$$

and Q_1 and Q_2 have the same cell probabilities according to μ.

Step 3. Since $\mathbb{R} \setminus S_1^{(2)} = [t, \infty)$ is an interval, using the restriction of μ to $[t, \infty)$, by the induction hypothesis we can replace the cells $S_2^{(2)}, \cdots, S_N^{(2)}$ with interval cells $\widehat{S}_2 < \cdots < \widehat{S}_N$ and corresponding code points to obtain a regular quantizer $\widehat{Q}$ (where $\widehat{S}_1 = S_1^{(2)}$ and $\hat{c}_1 = c_1^{(2)}$) such that $\widehat{Q}$ has the same cell probabilities as Q_2 according to μ, and

$$D(\widehat{Q}, \mu) \le D(Q_2, \mu).$$

Consequently, $D(\widehat{Q}, \mu) \le D(Q, \mu)$, and $\widehat{Q}$ and Q have the same cell probabilities. □

Proof of Theorem 2 Assume $D(Q) < \infty$; otherwise the statement is trivial. Suppose that Q is an infinite-point quantizer with cells $\{S_1, S_2, \dots\}$ and code points $\{c_1, c_2, \dots\}$. The proof consists of two parts. First, for every nonatomic finite measure μ and for every positive integer n we construct an infinite-point quantizer $\widehat{Q}_n$ with $D(\widehat{Q}_n) \le D(Q)$ having cells $\{\widehat{S}_1^{(n)}, \widehat{S}_2^{(n)}, \dots\}$ such that $\mu(S_i) = \mu(\widehat{S}_i^{(n)})$ for all i, and either $\widehat{S}_i^{(n)} < \widehat{S}_j^{(n)}$ or $\widehat{S}_i^{(n)} > \widehat{S}_j^{(n)}$ if $i \ne j$, $1 \le i, j \le n$. Then we show that $\widehat{Q}_n$ converges, in a sense, to an almost regular quantizer $\widehat{Q}$ which has the same cell probabilities as Q and satisfies $D(\widehat{Q}) \le D(Q)$.

The construction is a simple application of Theorem 1. For $n \ge 1$ let μ_n denote the restriction of μ to $\bigcup_{i \le n} S_i$, and apply Theorem 1 to μ_n and the n-point quantizer with cells $\{S_1, \dots, S_{n-1}, \bigcup_{i \ge n} S_i\}$ and code points $\{c_1, \dots, c_n\}$. The resulting regular quantizer $\bar{Q}_n$

has code points, say, $\{\bar{c}_1^{(n)}, \ldots, \bar{c}_n^{(n)}\}$ and interval cells $\{\bar{S}_1^{(n)}, \ldots, \bar{S}_n^{(n)}\}$ such that $\mu\big(\bar{S}_i^{(n)} \cap \big(\bigcup_{i\le n} S_i\big)\big) = \mu(S_i)$ for $i = 1, \ldots, n$. Now define $\widehat{Q}_n$ to have cells $\widehat{S}_i^{(n)} = \bar{S}_i^{(n)} \cap \big(\bigcup_{i\le n} S_i\big)$ for $i = 1, \ldots, n$, and $\widehat{S}_i^{(n)} = S_i$ for $i > n$, with corresponding code points $c_i^{(n)} = \bar{c}_i^{(n)}$ for $i = 1, \ldots, n$, and $c_i^{(n)} = c_i$ for $i > n$. Then either $\widehat{S}_i^{(n)} < \widehat{S}_j^{(n)}$ or $\widehat{S}_i^{(n)} > \widehat{S}_j^{(n)}$ for all $1 \le i, j \le n$, and

$$\begin{aligned} D(\widehat{Q}_n) &= D(\bar{Q}_n, \mu_n) + \sum_{i>n} D_i(Q, \mu) \\ &\le D(Q, \mu_n) + \sum_{i>n} D_i(Q, \mu) \\ &= D(Q, \mu). \end{aligned}$$

In the second part of the proof we show that there is a μ-almost regular quantizer $\widehat{Q}$ with (interval) cells $\{\widehat{S}_1, \widehat{S}_2, \ldots\}$ such that

$$\mu(\widehat{S}_i) = \mu(S_i) \quad \text{for all } i \tag{13}$$

and

$$D(\widehat{Q}) \le \liminf_{n\to\infty} D(\widehat{Q}_n). \tag{14}$$

Since $D(\widehat{Q}_n) \le D(Q)$, this completes the proof.

To show the existence of $\widehat{Q}$ satisfying (13) and (14), let $a_i^{(n)} = \inf \widehat{S}_i^{(n)}$ and $b_i^{(n)} = \sup \widehat{S}_i^{(n)}$ and form the vector

$$\mathbf{v}_n = (a_1^{(n)}, b_1^{(n)}, c_1^{(n)}, a_2^{(n)}, b_2^{(n)}, c_2^{(n)}, \ldots)$$

(recall that $c_i^{(n)}$ is the code point associated with $\widehat{S}_i^{(n)}$, and that $a_i^{(n)} \le c_i^{(n)} \le b_i^{(n)}$ for $1 \le i \le n$ by construction since $\bar{Q}_n$ was a regular quantizer). By Cantor's diagonalization method there is a subsequence of $\{\mathbf{v}_n\}$, for convenience denoted also by $\{\mathbf{v}_n\}$, converging componentwise to a vector $\mathbf{v} = (a_1, b_1, \hat{c}_1, a_2, b_2, \hat{c}_2, \ldots)$ (convergence to ∞ or $-\infty$ is also allowed). Denote the corresponding subsequence of quantizers also by $\{\widehat{Q}_n\}$. Then $a_i \le \hat{c}_i \le b_i$ for all i, and the intervals (a_i, b_i), $i = 1, 2, \ldots$, are pairwise disjoint since $(a_i^{(n)}, b_i^{(n)})$, $i = 1, \ldots, n$, are pairwise disjoint by the construction of $\widehat{Q}_n$. Also by construction, we have that for all $1 \le i, j \le n$, $(a_i^{(n)}, b_i^{(n)}) \cap \widehat{S}_j^{(n)} = \emptyset$ if $j \ne i$. Hence for all $n \ge i$,

$$\mu((a_i^{(n)}, b_i^{(n)})) \le \mu\big(\widehat{S}_i^{(n)} \cup \bigcup_{j>n} \widehat{S}_j^{(n)}\big) = \mu(S_i) + \sum_{j>n} \mu(S_i).$$

Since $\mu((a_i^{(n)}, b_i^{(n)})) \ge \mu(S_i)$ for all i, this implies $\lim_n \mu((a_i^{(n)}, b_i^{(n)})) = \mu(S_i)$ for all i, and since μ is nonatomic and $\lim_n a_i^{(n)} = a_i$, $\lim_n b_i^{(n)} = b_i$, we obtain $\mu((a_i, b_i)) = \mu(S_i)$. Thus the quantizer $\widehat{Q}$ with cells $\widehat{S}_i = [a_i, b_i)$ and code points $\hat{c}_i$, $i = 1, 2, \ldots$ satisfies (13). (Note that it may happen that $\hat{c}_i = \infty$ or $\hat{c}_i = -\infty$).

To show (14), observe that if $x \in (a_i, b_i)$, then $x \in (a_i^{(n)}, b_i^{(n)})$ for all n large enough. Since the sequence $\{\bigcup_{j \le n} \widehat{S}_j^{(n)}\}_{n=1}^{\infty}$ increases to $\mathbb{R}$ by construction, and either $\widehat{S}_i^{(n)} < \widehat{S}_j^{(n)}$ or $\widehat{S}_i^{(n)} > \widehat{S}_j^{(n)}$ for $i \neq j$, $1 \le i, j \le n$, this implies that if $x \in (a_i, b_i)$, then $x \in \widehat{S}_i^{(n)}$ for all n large enough. In other words, letting 1_S denote the indicator function of a set $S \subset \mathbb{R}$,

$$\lim_{n\to\infty} 1_{(a_i,b_i)\cap \widehat{S}_j^{(n)}}(x) = \begin{cases} 1_{(a_i,b_i)}(x) & \text{if } i = j \\ 0 & \text{otherwise.} \end{cases}$$

Also, for all $x \in \mathbb{R}$ we have,

$$\liminf_{n\to\infty} \rho(|x - c_i^{(n)}|) \ge \rho(|x - \hat{c}_i|)$$

since $\lim_n c_i^{(n)} = \hat{c}_i$ and ρ is nondecreasing and left-continuous (this also holds if $\hat{c}_i$ is not finite; recall (3)). Thus, since $\mu\big(\bigcup_i (a_i, b_i)\big) = 1$, we obtain

$$\begin{aligned} \liminf_{n\to\infty} D(\widehat{Q}_n) &= \liminf_{n\to\infty} \int_{\mathbb{R}} \sum_{i,j} 1_{(a_i,b_i)\cap \widehat{S}_j^{(n)}}(x) \rho(|x - c_j^{(n)}|)\, d\mu(x) \\ &\ge \int_{\mathbb{R}} \sum_{i,j} \liminf_{n\to\infty} 1_{(a_i,b_i)\cap \widehat{S}_j^{(n)}}(x) \rho(|x - c_j^{(n)}|)\, d\mu(x) \\ &\ge \int_{\mathbb{R}} \sum_{i} 1_{(a_i,b_i)}(x) \rho(|x - \hat{c}_i|)\, d\mu(x) \\ &= D(\widehat{Q}) \end{aligned}$$

where the first inequality follows by applying Fatou's lemma [17] twice. Finally, we can use the centroid rule (4) to replace any nonfinite $\hat{c}_i$ by a finite one that lies inside its associated cell. The modified quantizer $\widehat{Q}$ is μ-almost regular and satisfies (13) and (14). $\square$

Appendix B

Lemma 4 *Assume X has a nonatomic distribution μ and $d(x, y) = \rho(|x - y|)$, where $\rho : [0, \infty) \to [0, \infty)$ is nondecreasing and left-continuous. Let $\{Q_n\}$ be a sequence of infinite-point almost regular quantizers, each with cells $\{S^{(n)} = [a_i^{(n)}, b_i^{(n)})\}$ and code points $\{c_i^{(n)}\}$ such that $\lim_n a_i^{(n)} = a_i$, $\lim_n b_i^{(n)} = b_i$, and $\lim_n c_i^{(n)} = c_i$ for all i, where a_i, b_i, and c_i are not necessarily finite. If the intervals $S_i = [a_i, b_i)$ satisfy $\mu\big(\bigcup_i S_i\big) = 1$, then for quantizer Q with cells $\{S_i\}$ and code points $\{c_i\}$, we have*

$$D(Q) \le \liminf_{n\to\infty} D(Q_n).$$

Proof Since ρ is nondecreasing and left-continuous, we have

$$\liminf_{n\to\infty} 1_{S_i^{(n)}}(x)\rho(|x-c_i^{(n)}|) \geq 1_{S_i}(x)\rho(|x-c_i|)$$

for μ-almost all x (recall definition (3) if c_i is not finite). Then by applying Fatou's lemma twice, we obtain

$$\begin{aligned}
D(Q) &= \int_{\mathbb{R}} \sum_i 1_{S_i}(x)\rho(|x-c_i|)\, d\mu(x) \\
&\leq \int_{\mathbb{R}} \sum_i \liminf_{n\to\infty} 1_{S_i^{(n)}}(x)\rho(|x-c_i^{(n)}|)\, d\mu(x) \\
&\leq \liminf_{n\to\infty} \int_{\mathbb{R}} \sum_i 1_{S_i^{(n)}}(x)\rho(|x-c_i^{(n)}|)\, d\mu(x) \\
&= \liminf_{n\to\infty} D(Q_n).
\end{aligned}$$

□

Appendix C

Proof of Lemma 2 Assume that f is nonincreasing, and denote the cells of Q by $\{S_1, S_2, \dots\}$. By Theorem 2 we can redefine Q to be almost regular. If Q has a finite number of code points, then it can be redefined to be a regular quantizer. For a finite-point regular quantizer the statement of the lemma reduces to [13, Lemma A.5].

Now assume that Q has infinitely many code points. Denote the distribution of X by μ and let us index the cells of Q in such a way that $\mu(S_1) \geq \mu(S_2) \geq \cdots$. The proof of Corollary 2 shows that there is a sequence $\{Q_n^*\}$ of regular n-point quantizers such that each Q_n^* has cell probabilities $\mu(S_1), \dots, \mu(S_{n-1}), \mu(\bigcup_{j\geq n} S_j)$ (and so $H(Q_n^*) \leq H(Q)$), and $\limsup_n D(Q_n^*) \leq D(Q)$.

Apply the lemma to Q_n^* to obtain a unique n-point regular quantizer $\widehat{Q}_n$ with $D(\widehat{Q}_n) \leq D(Q_n^*)$, having subdivision points $\sigma = q_0^{(n)} < q_1^{(n)} < \cdots < q_{n-1}^{(n)} < q_n^{(n)} = \tau$ and code points $c_i^{(n)} = E\{X | X \in [q_{i-1}^{(n)}, q_i^{(n)})\}$, $i = 1, \dots, n$, such that $\mu([q_{i-1}^{(n)}, q_i^{(n)})) = p_i^{(n)}$ for $i = 1, \dots, n$, where $p_1^{(n)}, \dots, p_n^{(n)}$ is obtained by ordering the probabilities $\mu(S_1), \dots, \mu(S_{n-1}), \mu(\bigcup_{j\geq n} S_j)$ in a nonincreasing manner. Let us formally make Q_n an infinite-point quantizer by defining $q_i^{(n)} = \tau$ for all $i > n$. Denote the distribution function of X by F, and let F^{-1} be the inverse of F in the interval (σ, τ). Clearly, for each fixed i and all n large enough, $\mu(S_i) \geq \mu(\bigcup_{j\geq n} S_j)$, and so $p_i^{(n)} = \mu(S_i)$. Consequently, for such i and n,

$$F(q_i^{(n)}) = \sum_{j\leq i} \mu(S_j)$$

and so letting $\hat{q}_i = F^{-1}(\sum_{j \le i} \mu(S_j))$, we have $q_i^{(n)} = \hat{q}_i$ for all n large enough. Therefore Lemma 4 can be applied to show that the infinite-point quantizer $\widehat{Q}$ with subdivision points $\hat{q}_0 < \hat{q}_1 < \hat{q}_2 < \cdots$ and code points $\hat{c}_i = E\{X | X \in [\hat{q}_{i-1}, \hat{q}_i)\}$, $i = 1, 2, \ldots$, satisfies

$$D(\widehat{Q}) \le \liminf_{n\to\infty} D(\widehat{Q}_n) \le \limsup_{n\to\infty} D(Q_n^*) \le D(Q).$$

Since $\widehat{Q}$ clearly satisfies the other requirements of the lemma, the proof is complete for nonincreasing f. A similar argument can be used when f is nondecreasing. $\square$

Proof of Lemma 3 Consider first the case where Q is a finite-point regular quantizer with ordered subdivision points $\{q_1, \ldots, q_{n-1}\}$ and code points $\{c_1, \ldots, c_n\}$. In this case, [11, eq. 13] shows that (7) holds for all $i = 1, \ldots, n-1$. (This result is an immediate consequence of the Kuhn-Tucker conditions of constrained optimization [25] applied to the distortion and the entropy of Q as functions of the vector $(q_1, \ldots, q_{n-1})$. Although explicit conditions on the source density and the distortion measure were not stated in [11], it is easy to check that the positivity and continuity of the source density and the continuity of ρ are sufficient.)

If p_i is constant for all i, then (7) does not depend on λ. Otherwise there is an i such that $p_i \ne p_{i+1}$, and then λ is uniquely given by

$$\lambda = \frac{\rho(c_{i+1} - q_i) - \rho(q_i - c_i)}{\log p_{i+1} - \log p_i}. \tag{15}$$

Notice that the numerator in the above expression is independent of the distribution of X, and the denominator depends only on the ratio of the probabilities of adjacent cells. Therefore, if Q is an infinite-point quantizer, then $p_i \ne p_{i+1}$ for some i, and the application of (15) to the conditional density of X on $[q_j, q_k)$ for all $j < k-1$ proves (7) for all i (clearly, the finite-point quantizer obtained by restricting Q to $[q_j, q_k)$ must also be optimal for the conditional density and the corresponding rate). $\square$

References

[1] S. P. Lloyd, "Least squared quantization in PCM." unpublished memorandum, Bell Lab., 1957. Also, *IEEE Trans. Inform. Theory,* vol. IT-28, no. 2, pp. 129-137., Mar. 1982.

[2] J. Max, "Quantizing for minimum distortion," *IEEE Trans. Inform. Theory*, vol. IT-6, pp. 7–12, Mar. 1960.

[3] Y. Linde, A. Buzo, and R. M. Gray, "An algorithm for vector quantizer design," *IEEE Transactions on Communications*, vol. COM-28, pp. 84–95, Jan. 1980.

[4] R. M. Gray, J. C. Kieffer, and Y. Linde, "Locally optimum block quantizer design," *Inform. Contr.*, vol. 45, pp. 178–198, 1980.

[5] S. Graf and H. Luschgy, *Foundations of Quantization for Probability Distributions.* Berlin, Heidelberg: Springer Verlag, 2000.

[6] D. Pollard, "Quantization and the method of k-means," *IEEE Trans. Inform. Theory*, vol. IT-28, pp. 199–205, Mar. 1982.

[7] E. A. Abaya and G. L. Wise, "Convergence of vector quantizers with applications to optimal quantization," *SIAM Journal on Applied Mathematics*, vol. 44, pp. 183–189, 1984.

[8] M. J. Sabin, *Global convergence and empirical consistency of the generalized Lloyd algorithm.* PhD thesis, Stanford Univ., 1984.

[9] A. Gersho and R. M. Gray, *Vector Quantization and Signal Compression.* Boston: Kluwer, 1992.

[10] T. Berger, "Optimum quantizers and permutation codes," *IEEE Trans. Inform. Theory*, vol. IT-18, pp. 759–765, Nov. 1972.

[11] N. Farvardin and J. W. Modestino, "Optimum quantizer performance for a class of non-Gaussian memoryless sources," *IEEE Trans. Inform. Theory*, vol. IT-30, pp. 485–497, May 1984.

[12] A. N. Netravali and R. Saigal, "Optimum quantizer design using a fixed-point algorithm," *Bell Syst. Tech. J*, vol. 55, pp. 1423–1435, Nov. 1976.

[13] J. C. Kieffer, T. M. Jahns, and V. A. Obuljen, "New results on optimal entropy-constrained quantization," *IEEE Trans. Inform. Theory*, vol. 34, pp. 1250–1258, Sep. 1988.

[14] P. A. Chou, T. Lookabaugh, and R. M. Gray, "Entropy-constrained vector quantization," *IEEE Trans. Acoust. Speech, Signal Processing*, vol. ASSP-37, pp. 31–42, Jan. 1989.

[15] P. A. Chou and B. J. Betts, "When optimal entropy-constrained quantizers have only a finite number of codewords." in *Proc. IEEE Int. Symp. Information Theory* (Cambridge, MA, USA, Aug. 16-21, 1998), p. 97.

[16] A. György and T. Linder, "Optimal entropy-constrained scalar quantization of a uniform source," *IEEE Trans. Inform. Theory*, vol. IT-46, pp. pp. 2704–2711, Nov. 2000.

[17] R. B. Ash, *Real analysis and probability.* New York: Academic Press, 1972.

[18] T. Cover and J. A. Thomas, *Elements of Information Theory.* New York: Wiley, 1991.

[19] T. Linder and R. Zamir, "Causal source coding of stationary sources with high resolution." submitted to *2001 IEEE Int. Symp. Inform. Theory.*

[20] F. Jones, *Lebesgue Integration on Euclidean Space.* London: Jones and Bartlett, 1993.

[21] A. D. Wyner, "An upper bound on entropy series," *Information and Control*, vol. 20, pp. 176–181, 1972.

[22] R. M. Gray and L. D. Davisson, "Quantizer mismatch," *IEEE Trans. Communications*, vol. 23, pp. 439–443, 1975.

[23] H. Gish and J. N. Pierce, "Asymptotically efficient quantizing," *IEEE Trans. Inform. Theory*, vol. IT-14, pp. 676–683, Sep. 1968.

[24] P. Zador, "Asymptotic quantization error of continuous signals and the quantization dimension," *IEEE Trans. Inform. Theory*, vol. IT-28, pp. 139–149, Mar. 1982.

[25] D. G. Luenberger, *Linear and Nonlinear Programming.* Reading, Mass.: Addison-Wesley, 2nd ed., 1984.