Codecell Convexity in Optimal Entropy-Constrained Vector Quantization

András György, *Member, IEEE,* and
Tamás Linder, *Senior Member, IEEE*

Abstract—**Properties of optimal entropy-constrained vector quantizers (ECVQs) are studied for the squared-error distortion measure. It is known that restricting an ECVQ to have convex codecells may preclude its optimality for some sources with discrete distribution. We show that for sources with *continuous* distribution, any finite-level ECVQ can be replaced by another finite-level ECVQ with *convex* codecells that has equal or better performance. We generalize this result to infinite-level quantizers, and also consider the problem of existence of optimal ECVQs for continuous source distributions. In particular, we show that given any entropy constraint, there exists an ECVQ with (possibly infinitely many) convex codecells that has minimum distortion among all ECVQs satisfying the constraint. These results extend analogous statements in entropy-constrained scalar quantization. They also generalize results in entropy-constrained vector quantization that were obtained via the Lagrangian formulation and, therefore, are valid only for certain values of the entropy constraint.**

Index Terms—Convex polytopes, entropy coding, infinite-level quantizers, optimal vector quantization, squared-error distortion.

I. INTRODUCTION

Characterizing properties of optimal quantizers is an important problem in quantization theory. In fixed-rate quantization, Lloyd's two necessary conditions for optimality [1], [2] give a useful characterization of quantizers having minimum distortion for a given number of codevectors. One of these optimality conditions, the nearest neighbor condition, implies that for the squared-error distortion measure, the codecells of an optimal scalar quantizer are intervals, and that in the vector case the cells are convex polytopes. This observation has proved very useful in analyzing optimal scalar and vector quantizer performance [3]. Even more importantly, the Lloyd conditions yield a popular iterative method for fixed-rate quantizer design [1], [4].

The Lloyd conditions have been generalized to optimal entropy-constrained vector quantizers (ECVQs) [5] via a Lagrangian formulation in which the performance of the quantizer is measured by a weighted sum of its distortion and rate (output entropy). In particular, for the squared-error distortion measure, the generalized form of the nearest neighbor condition implies that an ECVQ that is optimal in this Lagrangian sense has convex polytope codecells. However, the Lagrangian formulation only yields optimal quantizers that achieve a distortion-rate point on the lower convex hull of the entropy-constrained operational distortion-rate function. If the operational distortion-rate function is not convex (which seems to be the typical case [6]), it lies strictly above its lower convex hull for a whole range of rate values. An ECVQ that is optimal for any such rate no longer has to satisfy the generalized nearest neighbor condition.

Although little is known about the structure of optimal ECVQs in general, the popular assumption in the literature is that the codecells of an optimal ECVQ are convex polytopes. In [8], it was demonstrated that this assumption is wrong in general: There exist *discrete* source distributions on the real line for which no scalar quantizer with interval cells can be optimal for some values of the entropy constraint. Similar counterexamples were exhibited in [9] for network scalar quantization problems such as multiresolution, multiple-description, and Wyner–Ziv scalar quantization. However, it was also shown in [8] that for *continuous* scalar source distributions and a wide class of distortion measures (including the squared-error distortion), there always exits an optimal entropy-constrained scalar quantizer with convex (interval) codecells for *any* value of the entropy constraint. Thus, for continuous source distributions, the requirement of codecell convexity is consistent with quantizer optimality in entropy-constrained scalar quantization. Corresponding results for vector quantization do not seem to be available in the literature. The question whether or not optimal vector quantizers have convex codecells is an interesting theoretical problem that also has practical significance. For example, the encoder of a vector quantizer with N nonconvex (e.g., disconnected) codecells can, in principle, be arbitrarily complex, while if the codecells are convex polytopes, the encoding complexity (both space and time) is at most on the order of N^2.

The goal of this correspondence is to generalize the results of [8] to vector quantization. We consider (absolutely) continuous source distributions and assume the squared-error distortion measure. After introducing the relevant notation and definitions in Section II, we first focus on finite-level quantizers in Section III. Here, the main result shows that for any finite-level ECVQ there exists another finite-level ECVQ with *convex* codecells that has the same (positive) codecell probabilities (thus, the same entropy) and equal or less distortion. Among other things, this result implies that should a finite-level optimal ECVQ for a given entropy constraint exist, it can be assumed to have convex codecells. Optimal ECVQ performance may not be achieved by a finite-level quantizer. This motivates us to consider infinite-level quantizers in Section IV. We extend the main result of Section III to vector quantizers with a countably infinite number of codecells, and also consider the problem of existence of optimal ECVQs. In particular, we show that given any entropy constraint, there exists an ECVQ with (possibly infinitely many) convex codecells that has minimum distortion among all ECVQs satisfying this entropy constraint. This generalizes a recent result in [10] that was obtained via the Lagrangian formulation and is thus valid only for certain values of the entropy constraint.

II. PRELIMINARIES

A vector quantizer Q is described by a measurable partition $\mathcal{S} = \{S_1, S_2, \ldots, S_N\}$ of $\mathbb{R}^k$ and a corresponding *codebook* $\mathcal{C} = \{c_1, c_2, \ldots, c_N\} \subset \mathbb{R}^k$. If $\mathcal{S}$ and $\mathcal{C}$ are finite, we call Q an N-level quantizer. We also allow countably infinite partitions and codebooks, in which case we formally write $N = \infty$ and call Q and infinite-level quantizer. The overall quantizer $Q \colon \mathbb{R}^k \to \mathcal{C}$ is

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

The c_i and S_i are called the *codevectors* and *codecells* (or *cells*) of Q, respectively. Without loss of generality, we assume that the codevectors are distinct. To define a quantizer Q with partition $\mathcal{S}$ and codebook $\mathcal{C}$, we write $Q \equiv (\mathcal{S}, \mathcal{C})$.

We suppose a nonnegative measurable *distortion measure* d: $\mathbb{R}^k \times \mathbb{R}^k \to [0, +\infty)$. For an $\mathbb{R}^k$-valued random vector X with distribution μ, the distortion of Q is measured by the expectation

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

In particular, we restrict the treatment to the popular squared error distortion

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

for $x = (x_1, \ldots, x_k)$ and $y = (y_1, \ldots, y_k)$. However, the results can easily be extended to somewhat more general distortion measures (see the remark after Lemma 1).

The *partial distortion* of the ith codecell S_i of Q is defined by

$$D_i(Q, \mu) \triangleq \int_{S_i} d(x, c_i) \, d\mu(x) \qquad (1)$$

so that

$$D(Q, \mu) = \sum_{i=1}^{N} D_i(Q, \mu). \qquad (2)$$

When the distribution μ is clear from the context, the short notations $D(Q)$ and $D_i(Q)$ will be used. We use (1) and (2) to extend the definition of $D_i(Q, \mu)$ and $D(Q, \mu)$ to any finite (not necessarily probability) measure on $\mathbb{R}^k$. In particular, we occasionally work with subprobability measures obtained by restricting μ to some Borel set $A \subset \mathbb{R}^k$: ν is said to be the *restriction* of μ to A if $\nu(B) = \mu(A \cap B)$ for each Borel set B.

The entropy-constrained *rate* of Q is the entropy of its output $Q(X)$

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

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

Given the source X, 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, sometimes called the *operational distortion-rate function*, is formally defined by

$$D_h(R) \triangleq \inf_Q \{D(Q) : H(Q) \leq R\}$$

where the infimum is taken over all finite- or infinite-level vector 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 $D_h(R) = +\infty$. 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* ECVQ.

It is worth noting that one of the two Lloyd conditions, the *centroid condition*, still holds for entropy-constrained quantizers. Thus, if Q is an optimal ECVQ, then the codevectors of Q must be the *centroids* of their corresponding cells, i.e.,

$$c_i = \frac{1}{\mu(S_i)} \int_{S_i} x \, d\mu(x)$$

for all S_i with positive probability $\mu(S_i) > 0$.

Central to our work are quantizers with convex codecells defined in terms of half-spaces or hyperplanes in $\mathbb{R}^k$. In the rest of the section, we summarize the relevant notation and conventions. We represent a *hyperplane* h in $\mathbb{R}^k$ by a vector of $k + 1$ real components $(a_1, \ldots, a_k, b)$ with $\|a\|^2 = \sum_{i=1}^{k} a_i^2 = 1$ such that

$$h = \{x \in \mathbb{R}^k : \langle a, x \rangle = b\}$$

where $\langle a, x \rangle = \sum_{i=1}^{k} a_i x_i$ denotes the usual inner product of $a = (a_1, \ldots, a_k)$ and $x = (x_1, \ldots, x_k)$ in $\mathbb{R}^k$. For the hyperplane h, the ordered pair $(H, \hat{H})$ will denote the corresponding complementary closed and open half-spaces

$$H = \{x \in \mathbb{R}^k : \langle a, x \rangle \leq b\}$$
$$\hat{H} = \mathbb{R}^k \setminus H = \{x \in \mathbb{R}^k : \langle a, x \rangle > b\}.$$

We say that the hyperplane h separates the sets $A, B \subset \mathbb{R}^k$, if $A \subset H$ and $B \subset \hat{H}$. $P \subset \mathbb{R}^k$ is a *convex polytope* if P is a finite intersection of (open or closed) half-spaces. Note that $\mu(h) = 0$ for any hyperplane h if the distribution μ is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^k$ (i.e., μ has a density). In this case, the boundary of a convex polytope (the union of its faces) has μ measure zero.

In the sequel, we alternatively use the notation h for the hyperplane itself and for the parameter vector $(a_1, \ldots, a_k, b)$; the actual meaning will be clear from the context. In particular, if the parameter vectors $(a_1^{(n)}, \ldots, a_k^{(n)}, b^{(n)})$ corresponding to a sequence of hyperplanes $\{h_n\}$ converge componentwise to a parameter vector $(a_1, \ldots, a_k, b)$ corresponding to a hyperplane h, we say that $\{h_n\}$ converges to h. It will be convenient to allow the parameter b of the limit hyperplane to be an extended real number $b \in \overline{\mathbb{R}} = \mathbb{R} \cup \{+\infty\} \cup \{-\infty\}$. Thus, the convergence of the first k components is meant in the usual sense in $\mathbb{R}$, while the convergence of $b^{(n)}$ is meant in $\overline{\mathbb{R}}$. Consistent with this, we set $H = \mathbb{R}^k$, $\hat{H} = \emptyset$ if $b = +\infty$, and $H = \emptyset$, $\hat{H} = \mathbb{R}^k$ if $b = -\infty$. We define h (as a subset of $\mathbb{R}^k$) to be empty in either case.

Using these conventions, it is easy to see that if $\{h_n\}$ converges to h, then for the sequence of corresponding half-spaces $\{(H_n, \hat{H}_n)\}$ and the half-spaces $(H, \hat{H})$ corresponding to h, we have

$$\lim_{n \to \infty} 1_{H_n}(x) = 1_H(x) \quad \text{and} \quad \lim_{n \to \infty} 1_{\hat{H}_n}(x) = 1_{\hat{H}}(x) \qquad (3)$$

for all $x \notin h$, where $1_A(x)$ denotes the indicator function of $A \subset \mathbb{R}^k$; i.e., $1_A(x) = 1$ if $x \in A$ and $1_A(x) = 0$ if $x \notin A$.

III. FINITE-LEVEL QUANTIZERS AND CODECELL CONVEXITY

In this section, and throughout the correspondence, we assume that the source distribution is absolutely continuous with respect to the Lebesgue measure (absolutely continuous, for short). We show that for any finite-level vector quantizer there exists another finite-level vector quantizer with convex codecells that has the same entropy and equal or less distortion. First we prove this in the next lemma for the special case of two-level quantizers. The lemma extends a similar result in [8] from scalar to vector quantization and it plays a key role in the proof of the main result.

Lemma 1: Let μ be an absolutely continuous finite measure and assume that $Q \equiv (\{S_1, S_2\}, \{c_1, c_2\})$ is an arbitrary two-level vector quantizer with finite distortion $D(Q)$. Then there exists a hyperplane h perpendicular to $c_1 - c_2$ and corresponding complementary half-spaces (H_1, H_2) with $\mu(H_i) = \mu(S_i)$, $i = 1, 2$, such that the two-level quantizer $\hat{Q} \equiv (\{H_1, H_2\}, \{c_1, c_2\})$ satisfies $D(\hat{Q}) \leq D(Q)$.

In particular, for every $\epsilon > 0$ there is a $\delta > 0$, which depends on μ but not on Q or $\hat{Q}$, such that

$$D(Q) - D(\hat{Q}) \geq 4\delta \|c_2 - c_1\| (\mu(S_1 \cap H_2) - \epsilon). \qquad (4)$$

Note that (4) implies that if the cells of a two-level quantizer cannot be separated by a hyperplane (up to a set of μ measure zero), then there

is another two-level quantizer with half-space cells that has the same entropy and *strictly less* distortion.

Proof: Since the Euclidean distance is invariant to orthogonal transformations, we can assume, without loss of generality, that $c_1 = (u_1, 0, \ldots, 0)$ and $c_2 = (u_2, 0, \ldots, 0)$ for some $u_1 < u_2$. Since μ is absolutely continuous, there is a $u \in \mathbb{R}$ such that the half-spaces

$$H_1 = \{x \colon x_1 \leq u\} \quad \text{and} \quad H_2 = \{x \colon x_1 > u\} \qquad (5)$$

satisfy $\mu(H_1) = \mu(S_1)$ and $\mu(H_2) = \mu(S_2)$. Note that the hyperplane corresponding to H_1 and H_2 is perpendicular to $c_1 - c_2$. Set $\hat{Q} \equiv (\{H_1, H_2\}, \{c_1, c_2\})$. We show that $D(\hat{Q}) \leq D(Q)$. We have

$$D(Q) = \int_{S_1} \|x - c_1\|^2 \, d\mu(x) + \int_{S_2} \|x - c_2\|^2 \, d\mu(x)$$

$$= \int_{S_1} \left((x_1 - u_1)^2 + x_2^2 + \cdots + x_k^2 \right) d\mu(x)$$

$$+ \int_{S_2} \left((x_1 - u_2)^2 + x_2^2 + \cdots + x_k^2 \right) d\mu(x)$$

$$= \int_{\mathbb{R}^k} \left(x_2^2 + \cdots + x_k^2 \right) d\mu(x) + \int_{S_1} (x_1 - u_1)^2 \, d\mu(x)$$

$$+ \int_{S_2} (x_1 - u_2)^2 \, d\mu(x).$$

Similarly,

$$D(\hat{Q}) = \int_{\mathbb{R}^k} \left(x_2^2 + \cdots + x_k^2 \right) d\mu(x)$$

$$+ \int_{H_1} (x_1 - u_1)^2 \, d\mu(x) + \int_{H_2} (x_1 - u_2)^2 \, d\mu(x).$$

Therefore,

$$D(Q) - D(\hat{Q})$$

$$= \int_{S_1 \cap H_2} \left((x_1 - u_1)^2 - (x_1 - u_2)^2 \right) d\mu(x)$$

$$+ \int_{S_2 \cap H_1} \left((x_1 - u_2)^2 - (x_1 - u_1)^2 \right) d\mu(x)$$

$$= (u_1^2 - u_2^2)(\mu(S_1 \cap H_2) - \mu(S_2 \cap H_1)) \qquad (6)$$

$$+ 2(u_2 - u_1) \left(\int_{S_1 \cap H_2} x_1 \, d\mu(x) - \int_{S_2 \cap H_1} x_1 \, d\mu(x) \right). \qquad (7)$$

Since $\mu(H_1) = \mu(S_1)$ and $\mu(H_2) = \mu(S_2)$, and $\{S_1, S_2\}$ and $\{H_1, H_2\}$ are partitions of $\mathbb{R}^k$, we have $\mu(S_1 \cap H_2) = \mu(S_2 \cap H_1)$; hence, the expression in (6) is equal to zero. Also, from (5) we have

$$\int_{S_1 \cap H_2} x_1 \, d\mu(x) \geq \int_{S_1 \cap H_2} u \, d\mu(x) = u\mu(S_1 \cap H_2)$$

and

$$\int_{S_2 \cap H_1} x_1 \, d\mu(x) \leq \int_{S_2 \cap H_1} u \, d\mu(x) = u\mu(S_2 \cap H_1)$$

and so the expression in (7) is nonnegative, implying

$$D(Q) - D(\hat{Q}) \geq 0.$$

This completes the proof of the first part.

Next we prove (4). Recall that

$$D(Q) - D(\hat{Q})$$

$$= 2(u_2 - u_1) \left(\int_{S_1 \cap H_2} x_1 \, d\mu(x) - \int_{S_2 \cap H_1} x_1 \, d\mu(x) \right). \qquad (8)$$

Since μ is absolutely continuous with respect to the Lebesgue measure, for every $\epsilon > 0$ there is a $\delta > 0$ such that for any two half-spaces $K_1 \subset K_2 \subset \mathbb{R}^k$ such that the distance between their defining hyperplanes is

at most δ, we have $\mu(K_2 \setminus K_1) \leq \epsilon$ (note that the two hyperplanes must be parallel because $K_1 \subset K_2$). Therefore,

$$\int_{S_1 \cap H_2} x_1 \, d\mu(x)$$

$$= \int_{S_1 \cap H_2 \cap \{x \colon x_1 > u + \delta\}} x_1 \, d\mu(x)$$

$$+ \int_{S_1 \cap H_2 \cap \{x \colon x_1 \leq u + \delta\}} x_1 \, d\mu(x)$$

$$\geq \int_{S_1 \cap H_2 \cap \{x \colon x_1 > u + \delta\}} (u + \delta) \, d\mu(x)$$

$$+ \int_{S_1 \cap H_2 \cap \{x \colon x_1 \leq u + \delta\}} u \, d\mu(x)$$

$$= u\mu(S_1 \cap H_2) + \int_{S_1 \cap H_2 \cap \{x \colon x_1 > u + \delta\}} \delta \, d\mu(x)$$

$$\geq u\mu(S_1 \cap H_2) + \delta(\mu(S_1 \cap H_2) - \mu(H_2 \setminus \{x \colon x_1 > u + \delta\}))$$

$$\geq u\mu(S_1 \cap H_2) + \delta(\mu(S_1 \cap H_2) - \epsilon).$$

Similarly, we have

$$\int_{S_2 \cap H_1} x_1 \, d\mu(x) \leq u\mu(S_2 \cap H_1) - \delta(\mu(S_2 \cap H_1) - \epsilon).$$

Since $\mu(S_2 \cap H_1) = \mu(S_1 \cap H_2)$, (8) implies

$$D(Q) - D(\hat{Q}) \geq 4\delta(u_2 - u_1)(\mu(S_1 \cap H_2) - \epsilon)$$

$$= 4\delta\|c_2 - c_1\|(\mu(S_1 \cap H_2) - \epsilon).$$

This completes the proof since the choice of δ depended only on ϵ and μ, but not on Q or $\hat{Q}$. $\qquad\square$

Remark (More General Distortion Measures): Lemma 1 easily generalizes to *weighted squared-error* distortion measures in the form

$$d(x, y) = \langle x - y, A(x - y) \rangle$$

where A is a positive-definite symmetric $k \times k$ matrix [2]. In this case, the separating hyperplane h and $c_2 - c_1$ will be perpendicular with respect to the inner product $\langle x, y \rangle_A \triangleq \langle x, Ay \rangle$. All subsequent results can be shown to hold for such distortion measures, but for simplicity we will restrict the treatment to the squared-error distortion, and only give below a sketch of the argument generalizing Lemma 1.

The distortion measure can be written as

$$d(x, y) = \langle B(x - y), B(x - y) \rangle$$

where B is the unique positive-definite symmetric square root of A, i.e., $A = B^2$. (Note that B always exists; see, e.g., [7]). Then the partial distortions of Q are given by

$$D_i(Q) = \int_{S_i} \|B(x - c_i)\|^2 \, d\mu(x) = \int_{\tilde{S}_i} \|y - \tilde{c}_i\|^2 \, d\tilde{\mu}(y)$$

where $\tilde{c}_i = Bc_i$, $\tilde{S}_i = BS_i \triangleq \{Bx \colon x \in S_i\}$, and $\tilde{\mu} = \mu B^{-1}$ (i.e., $\tilde{\mu}(S) = \mu(B^{-1}S)$ for every Borel set S). Then $\tilde{\mu}$ is absolutely continuous and $\tilde{\mu}(\tilde{S}_i) = \mu(S_i)$, $i = 1, 2$, so the first part of Lemma 1 implies that there are complementary half-spaces $(\tilde{H}_1, \tilde{H}_2)$ that are perpendicular to $\tilde{c}_1 - \tilde{c}_2$ and have measure $\tilde{\mu}(\tilde{H}_i) = \mu(S_i)$, $i = 1, 2$, such that

$$\sum_{i=1}^{2} \int_{\tilde{H}_i} \|y - \tilde{c}_i\|^2 \, d\tilde{\mu}(y) \leq \sum_{i=1}^{2} \int_{\tilde{S}_i} \|y - \tilde{c}_i\|^2 \, d\tilde{\mu}(y). \qquad (9)$$

Letting $H_i = B^{-1}\tilde{H}_i$, $i = 1, 2$, it is easy to check that (H_1, H_2) are complementary half-spaces with $\mu(H_i) = \mu(S_i)$, $i = 1, 2$, that are perpendicular to $c_2 - c_1$ with respect to the inner product $\langle x, y \rangle_A$. If we

let $\hat{Q} \equiv (\{H_1, H_2\}, \{c_1, c_2\})$, then $D(\hat{Q})$ is equal to the left-hand side of (9), so that $D(\hat{Q}) \leq D(Q)$. The proof of the second statement of the lemma can be generalized in a similar manner. In particular, $\|c_2 - c_1\|$ is replaced by

$$\|c_2 - c_1\|_A = \sqrt{\langle c_2 - c_1, A(c_2 - c_1)\rangle}$$

in (4).

Next we extend Lemma 1 to quantizers with more than two codecells.

Theorem 1: Assume μ is an absolutely continuous finite measure and suppose the N-level quantizer

$$Q \equiv (\{S_1, \ldots, S_N\}, \{c_1, \ldots, c_N\})$$

has finite distortion $D(Q)$. Then there is an $(N+1)$-level quantizer

$$\hat{Q} \equiv (\{\hat{S}_1, \ldots, \hat{S}_{N+1}\}, \{c_1, \ldots, c_N, c_{N+1}\})$$

with c_{N+1} being arbitrary, such that each $\hat{S}_i$, $i = 1, \ldots, N$, is a convex polytope

$$\mu(\hat{S}_i) = \mu(S_i), \qquad i = 1, \ldots, N$$

$$\mu(\hat{S}_{N+1}) = 0$$

and

$$D(\hat{Q}) \leq D(Q).$$

Specifically, there exist hyperplanes $\{h_{i,j}; 1 \leq i < j \leq N\}$ and corresponding half-spaces $\{(H_{i,j}, H_{j,i}); 1 \leq i < j \leq N\}$ with each $h_{i,j}$ being perpendicular to $c_i - c_j$ such that the codecells of $\hat{Q}$ are given by

$$\hat{S}_i = \bigcap_{j \neq i} H_{i,j}, \qquad i = 1, \ldots, N$$

and $\hat{S}_{N+1} = \mathbb{R}^k \setminus \bigcup_{i=1}^N \hat{S}_i$.

The result generalizes [8, Theorem 1] from scalar to vector quantization (modulo the assumption of the squared-error distortion measure). The proof, however, is more involved than in the scalar case. In [8], the desired quantizer could be constructed in a finite number of steps, but that construction does not generalize to the multidimensional case. Here the construction is less explicit: $\hat{Q}$ with the stated properties is shown to exist as the "limit" of an infinite sequence of quantizers constructed in a recursive manner. The detailed proof is given in the Appendix.

DeMorgan's law implies that $\hat{S}_{N+1}$ can be written as a union of finitely many convex polytopes. Making each of these polytopes a new cell (of μ measure zero) and defining the associated codevectors arbitrarily, we obtain the following corollary.

Corollary 1: Assume that X has an absolutely continuous distribution. Then for any finite-level quantizer Q with distortion $D(Q) < \infty$ there is a finite-level quantizer $\hat{Q}$ with convex codecells (some of which may have probability zero) that has the same positive cell probabilities as Q and distortion $D(\hat{Q}) \leq D(Q)$.

Note that Q and $\hat{Q}$ in the corollary have the same entropy. Thus, any finite-level ECVQ can be replaced by another one which has convex codecells, the same entropy, and equal or less distortion.

In Theorem 1, the convex polytopes comprising $\hat{S}_{N+1}$ have μ measure zero. Each of these polytopes S is an intersection of some half-spaces from the collection $\{H_{i,j}, H_{j,i}; 1 \leq i < j \leq N\}$. Assume now that the density of μ is positive everywhere. Then, if the interior of S were not empty, we should have $\mu(S) > 0$. Since $\mu(S) = 0$, the

interior of S is empty and, in fact, $S \subset h_{i,j}$ for some $i < j$. Thus, the union of the boundaries (faces) of the polytopes $\hat{S}_i$, $i = 1, \ldots, N$ contains $\hat{S}_{N+1}$, and it follows that the union of the closures of $\hat{S}_i$, $i = 1, \ldots, N$ covers $\mathbb{R}^k$. We can, therefore, redefine the cells $\hat{S}_i$, $i = 1, \ldots, N$ to obtain a partition of $\mathbb{R}^k$ by assigning the common boundary of any two cells to the cell with the lower index. Hence, we obtain the following corollary to Theorem 1.

Corollary 2: Assume that X has an everywhere positive density. Then, for any N-level quantizer Q with finite distortion $D(Q)$, there is an N-level quantizer $\hat{Q}$ with distortion $D(\hat{Q}) \leq D(Q)$ such that $\hat{Q}$ has convex cells and the same cell probabilities as Q.

If X has a positive density concentrated on a convex subset C of $\mathbb{R}^k$, then it is enough to consider quantizers defined only on C. In this case, $\hat{Q}$ is an N-level quantizer with convex codecells defined only on C.

For a given source distribution and entropy constraint R, there might not exist a finite-level quantizer achieving optimal performance $D_h(R)$. Nevertheless, Theorem 1 readily implies that finite-level quantizers with convex codecells can arbitrarily approach the optimal performance if the source has a finite second moment. The proof is almost identical to that of [8, Corollary 2] which deals with the scalar case.

Corollary 3: Assume that X has an absolutely continuous distribution and finite second moment $E\|X\|^2 < \infty$. Then, for any $R \geq 0$ and $\epsilon > 0$, there is a finite-level quantizer Q_ϵ with convex codecells such that $H(Q_\epsilon) \leq R$ and

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

IV. INFINITE-LEVEL QUANTIZERS AND ECVQ OPTIMALITY

In this section, we first generalize Theorem 1 to infinite-level quantizers. In the sequel, it will be convenient to allow quantizers that are defined on subsets of $\mathbb{R}^k$. We say that a quantizer Q is defined μ-almost everywhere (μ-a.e.) if Q is defined on a set $S \subset \mathbb{R}^k$ (which may be all of $\mathbb{R}^k$) such that $\mu(\mathbb{R}^k \setminus S) = 0$. In this case, the codecells of Q form a partition of S instead of $\mathbb{R}^k$. Note that by forming the extra codecell $\mathbb{R}^k \setminus S$ of measure zero for Q, we can always extend its domain of definition to the whole of $\mathbb{R}^k$ without changing its distortion or entropy.

Theorem 2: Suppose X has an absolutely continuous distribution μ. Then, for any quantizer Q with finite distortion $D(Q)$, there is a quantizer $\hat{Q}$ defined μ-a.e., such that $\hat{Q}$ has convex cells, $\hat{Q}$ and Q have the same codevectors and codecell probabilities, and $D(\hat{Q}) \leq D(Q)$.

In case Q is an infinite-level quantizer, Theorem 2 does not imply that the cells of $\hat{Q}$ are convex polytopes. For example, a corollary of the Vitali covering theorem [11] shows that there exists a countable collection $\{B_1, B_2, \ldots\}$ of disjoint closed balls in $\mathbb{R}^k$ such that $\mathbb{R}^k \setminus (\bigcup_i B_i)$ has Lebesgue measure zero. If μ is absolutely continuous, then any quantizer with partition $\{B_1, B_2, \ldots\}$ is defined μ-a.e. and its codecells are closed balls.

The theorem is an easy consequence of the following lemma whose proof is deferred to the Appendix. The lemma will also prove useful later in showing the existence of optimal ECVQs.

Lemma 2: Suppose X has an absolutely continuous distribution μ. For each $n = 1, 2, \ldots$, let

$$Q_n \equiv (\{S_1^{(n)}, S_2^{(n)}, \ldots\}, \{c_1^{(n)}, c_2^{(n)}, \ldots\})$$

be an infinite-level quantizer, defined μ-a.e., such that for all $1 \leq i < j \leq n$ the cells $S_i^{(n)}$ and $S_j^{(n)}$ are separated by a hyperplane $h_{i,j}^{(n)}$. Furthermore, assume that for all $i \geq 1$, $\lim_n \mu(S_i^{(n)}) = p_i$ for some $p_i \geq 0$ such that $\sum_{i=1}^\infty p_i = 1$. Then there exists a quantizer $\hat{Q}$ defined

μ-a.e. such that it has convex cells $\{\hat{S}_1, \hat{S}_2, \ldots\}$ with probabilities $\mu(\hat{S}_i) = p_i$ for all $i \geq 1$, and its distortion is upper-bounded as

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

If the limit $\lim_n c_i^{(n)} = c_i \in \mathbb{R}^k$ exists for all $i \geq 1$, then $\hat{Q}$ can be defined to have codebook $\{c_1, c_2, \ldots\}$.

Proof of Theorem 2: Suppose Q is an infinite-level quantizer $Q \equiv (\{S_1, S_2, \ldots\}, \{c_1, c_2, \ldots\})$; otherwise, the statement immediately follows from Theorem 1. For each positive integer n, let μ_n be the restriction of μ to $R_n = \bigcup_{i=1}^{n} S_i$, and define the n-level quantizer

$$q_n \equiv \left(\left\{ S_1, \ldots, S_{n-1}, \bigcup_{j \geq n} S_j \right\}, \left\{ c_1, \ldots, c_{n-1}, c_n \right\} \right).$$

Applying Theorem 1 to μ_n and q_n, we obtain the $(n+1)$-level quantizer

$$\hat{q}_n \equiv \left(\left\{ \hat{S}_1^{(n)}, \ldots, \hat{S}_n^{(n)}, \hat{S}_{n+1}^{(n)} \right\}, \left\{ c_1, \ldots, c_n, \hat{c}_{n+1} \right\} \right)$$

such that $\hat{S}_i^{(n)}$ is a convex polytope with measure $\mu_n(\hat{S}_i^{(n)}) = \mu_n(S_i)$ for each $i = 1, \ldots, n$, $\mu(\hat{S}_{n+1}^{(n)}) = 0$, $\hat{c}_{n+1}$ is arbitrary, and $D(\hat{q}_n, \mu_n) \leq D(q_n, \mu_n)$. Let

$$S_i^{(n)} \triangleq \begin{cases} \hat{S}_i^{(n)} \cap R_n, & \text{if } 1 \leq i \leq n \\ S_{n+1} \cup (\hat{S}_{n+1}^{(n)} \cap R_n), & \text{if } i = n+1 \\ S_i, & \text{if } i \geq n+2 \end{cases}$$

and define the infinite-level quantizer

$$Q_n \equiv \left(\left\{ S_1^{(n)}, S_2^{(n)}, \ldots \right\}, \left\{ c_1, c_2, \ldots \right\} \right).$$

Then $\mu(S_i^{(n)}) = \mu(S_i)$ for all $i \geq 1$. Also,

$$\begin{aligned} D(Q_n, \mu) &= \sum_{i=1}^{n} D_i(Q_n, \mu) + \sum_{i>n} D_i(Q_n, \mu) \\ &= D(\hat{q}_n, \mu_n) + \sum_{i>n} D_i(Q_n, \mu) \\ &\leq D(q_n, \mu_n) + \sum_{i>n} D_i(Q_n, \mu) \\ &= \sum_{i=1}^{n} D_i(Q, \mu) + \sum_{i>n} D_i(Q, \mu) \\ &= D(Q, \mu). \end{aligned}$$

Thus, $D(Q_n, \mu) \leq D(Q, \mu)$ for all n, so

$$\liminf_{n \to \infty} D(Q_n, \mu) \leq D(Q, \mu).$$

Clearly, the sequence $\{Q_n\}$ satisfies the requirements of Lemma 2; hence, the statement of the theorem follows. □

Theorem 2 shows that in the definition of the operational distortion-rate function $D_h(R)$ it is enough to consider μ-a.e. defined quantizers with convex codecells. The next theorem uses this fact to show the existence of an optimal ECVQ for a given rate constraint.

Theorem 3: Suppose that X has an absolutely continuous distribution μ. Then, for any $R \geq 0$, there exists a quantizer $\hat{Q}$ defined μ-a.e. and having convex cells such that $H(\hat{Q}) \leq R$ and $D(\hat{Q}) = D_h(R)$.

We need the following result from [8, p. 421].

Lemma 3: Let $R \geq 0$ and define the set C of probability vectors

$$C \triangleq \left\{ (p_1, p_2, \ldots) : p_i \geq 0 \text{ for all } i, \right.$$

$$\left. p_1 \geq p_2 \geq \cdots, \sum_{i=1}^{\infty} p_i = 1, -\sum_{i=1}^{\infty} p_i \log p_i \leq R \right\}.$$

Then C is compact under pointwise convergence.

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

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

By Theorem 2, each Q_n can be assumed to be defined μ-a.e. and to have convex cells. For positive integers n and i let $S_i^{(n)}$ denote the cell of Q_n with the ith largest probability, let $c_i^{(n)}$ denote the corresponding codevector, and let $p_i^{(n)} = \mu(S_i^{(n)})$. (In case of ties, any ordering of equiprobable cells suffices.) Also, if Q_n is a finite-level quantizer with l cells, we formally define $p_i^{(n)} = 0$ for all $i > l$. For every n, let

$$p^{(n)} \triangleq \left(p_1^{(n)}, p_2^{(n)}, \ldots \right).$$

By Lemma 3, the sequence $\{p^{(n)}\}$ has a pointwise convergent subsequence $\{p^{(n_j)}\}$, which converges to some probability vector $p = (p_1, p_2, \ldots)$ with entropy

$$H(p) \triangleq -\sum_{i=1}^{\infty} p_i \log p_i \leq R.$$

The corresponding sequence of quantizers $\{Q_{n_j}\}$ clearly satisfies the requirements of Lemma 2 if we redefine each Q_{n_j} on a set of probability zero by assigning the common boundary of any two of its convex codecells to the codecell with the lower index. Thus, by Lemma 2 we obtain a quantizer $\hat{Q}$ defined μ-a.e., with convex cells $\{\hat{S}_1, \hat{S}_2, \ldots\}$ such that $\mu(\hat{S}_i) = p_i$ for all $i \geq 1$. We have

$$H(\hat{Q}) = H(p) \leq R$$

and by Lemma 2,

$$D(\hat{Q}) \leq \liminf_{n \to \infty} D(Q_n) = D_h(R).$$

Hence, $\hat{Q}$ is an optimal ECVQ at rate R. □

APPENDIX

Proof of Theorem 1: We construct recursively a sequence of quantizers $\{Q_n\}$ that converges (in an appropriate sense) to a quantizer $\hat{Q}$ that has the desired properties. Each Q_n in the sequence has the same codebook $\{c_1, \ldots, c_N\}$ as Q, and partition $\{S_1^{(n)}, \ldots, S_N^{(n)}\}$ that satisfies $\mu(S_i^{(n)}) = \mu(S_i)$ for all $i = 1, \ldots, N$.

Set $Q_1 = Q$ and assume that for some $n \geq 1$ we have constructed the N-level quantizer Q_n satisfying the above conditions. We use Q_n to construct Q_{n+1}. For any $i < j$, let $\mu_{i,j}^{(n)}$ denote the restriction of μ to $S_i^{(n)} \cup S_j^{(n)}$. Since $\mu_{i,j}^{(n)}$ is also absolutely continuous, we can apply Lemma 1 to $\mu_{i,j}^{(n)}$ and the two-level quantizer

$$q_{i,j}^{(n)} \equiv \left(\{S_i^{(n)}, \mathbb{R}^k \setminus S_i^{(n)}\}, \{c_i, c_j\} \right).$$

We obtain for each pair of indexes $i < j$ a hyperplane $h_{i,j}^{(n)}$ (perpendicular to $c_i - c_j$) and corresponding complementary half-spaces $(H_{i,j}^{(n)}, H_{j,i}^{(n)})$ with measures

$$\mu_{i,j}^{(n)}(S_i^{(n)}) = \mu_{i,j}^{(n)}(H_{i,j}^{(n)}), \quad \mu_{i,j}^{(n)}(S_j^{(n)}) = \mu_{i,j}^{(n)}(H_{j,i}^{(n)})$$

such that the quantizer

$$\hat{q}_{i,j}^{(n)} \equiv \left(\{ H_{i,j}^{(n)}, H_{j,i}^{(n)} \}, \{ c_i, c_j \} \right)$$

satisfies

$$D\left(\hat{q}_{i,j}^{(n)}, \mu_{i,j}^{(n)} \right) \leq D\left(q_{i,j}^{(n)}, \mu_{i,j}^{(n)} \right).$$

Now let

$$\rho_{i,j}^{(n)} \triangleq D\left(q_{i,j}^{(n)}, \mu_{i,j}^{(n)} \right) - D\left(\hat{q}_{i,j}^{(n)}, \mu_{i,j}^{(n)} \right) \qquad (10)$$

and chose $i^* < j^*$ such that

$$\rho_{i^*, j^*}^{(n)} = \max_{1 \leq i < j \leq N} \rho_{i,j}^{(n)}. \qquad (11)$$

We define Q_{n+1} to have codevectors $\{ c_1, \ldots, c_N \}$ and cells $\{ S_1^{(n+1)}, \ldots, S_N^{(n+1)} \}$ given by

$$S_i^{(n+1)} \triangleq \begin{cases} H_{i^*, j^*}^{(n)} \cap \left(S_{i^*}^{(n)} \cup S_{j^*}^{(n)} \right), & \text{if } i = i^* \\ H_{j^*, i^*}^{(n)} \cap \left(S_{i^*}^{(n)} \cup S_{j^*}^{(n)} \right), & \text{if } i = j^* \\ S_i^{(n)}, & \text{if } i \neq i^*, j^*. \end{cases}$$

Then clearly, $\mu(S_i^{(n+1)}) = \mu(S_i^{(n)}) = \mu(S_i)$ for all $i = 1, \ldots, N$, and the distortion of Q_{n+1} is at most that of Q_n since

$$D(Q_{n+1}, \mu)$$
$$= D_{i^*}(Q_{n+1}, \mu) + D_{j^*}(Q_{n+1}, \mu) + \sum_{i \neq i^*, j^*} D_i(Q_{n+1}, \mu)$$
$$= D\left(\hat{q}_{i^*, j^*}^{(n)}, \mu_{i^*, j^*}^{(n)} \right) + \sum_{i \neq i^*, j^*} D_i(Q_{n+1}, \mu)$$
$$\leq D\left(q_{i^*, j^*}^{(n)}, \mu_{i^*, j^*}^{(n)} \right) + \sum_{i \neq i^*, j^*} D_i(Q_{n+1}, \mu)$$
$$= D_{i^*}(Q_n, \mu) + D_{j^*}(Q_n, \mu) + \sum_{i \neq i^*, j^*} D_i(Q_n, \mu)$$
$$= D(Q_n, \mu). \qquad (12)$$

Continuing in this manner we obtain an infinite sequence $\{ Q_n \}$ of N-level quantizers and an associated sequence of $N(N-1)/2$ hyperplanes $\{ (h_{1,2}^{(n)}, h_{1,3}^{(n)}, \ldots, h_{N-1,N}^{(n)}) \}$. Now pick a subsequence of $\{ Q_n \}$, also denoted by $\{ Q_n \}$ to simplify the notation, such that the corresponding sequence of hyperplanes $\{ (h_{1,2}^{(n)}, h_{1,3}^{(n)}, \ldots, h_{N-1,N}^{(n)}) \}$ converges, that is, for all $i < j$, $\{ h_{i,j}^{(n)} \}$ converges to some hyperplane $h_{i,j}$. (Recall from Section II the definition of convergence for a sequence of hyperplanes.)

For every $1 \leq i < j \leq N$, let $(H_{i,j}, H_{j,i})$ denote the half-spaces corresponding to $h_{i,j}$, and for $1 \leq i \leq N$, define the convex polytopes

$$\hat{S}_i \triangleq \bigcap_{1 \leq j \leq N, j \neq i} H_{i,j}.$$

Furthermore, let

$$\hat{S}_{N+1} \triangleq \mathbb{R}^k \setminus \bigcup_{j=1}^N \hat{S}_j$$

and let c_{N+1} be an arbitrary vector in $\mathbb{R}^k$. Finally, define the $(N+1)$-level quantizer

$$\hat{Q} \equiv (\{ \hat{S}_1, \ldots, \hat{S}_{N+1} \}, \{ c_1, \ldots, c_{N+1} \}).$$

In the remainder of the proof we show that $\hat{Q}$ has the desired properties. Since the sequence $\{ D(Q_n, \mu) \}$ is nonnegative and nonincreasing, it converges, so $D(Q_n, \mu) - D(Q_{n+1}, \mu) \to 0$ as $n \to \infty$. Thus, since $\rho_{i,j}^{(n)} \geq 0$ for all i, j and n, (10)–(12) imply for all $1 \leq i < j \leq N$

$$\lim_{n \to \infty} \rho_{i,j}^{(n)} = 0.$$

Fix $\epsilon > 0$, and let

$$d^* \triangleq \min_{1 \leq i < j \leq N} \| c_i - c_j \|.$$

Then, from the second part of Lemma 1, there is a $\delta = \delta(\epsilon)$ such that

$$\rho_{i,j}^{(n)} \geq 4 \delta d^* \left(\mu_{i,j}^{(n)} \left(S_i^{(n)} \cap H_{j,i}^{(n)} \right) - \epsilon \right)$$
$$= 4 \delta d^* \left(\mu \left(S_i^{(n)} \cap H_{j,i}^{(n)} \right) - \epsilon \right).$$

(Since each $\mu_{i,j}^{(n)}$ is dominated by μ, it is easy to see from the proof of Lemma 1 that the same δ works for all i, j, and n.) Therefore, since $\delta, d^* > 0$

$$\limsup_{n \to \infty} \mu(S_i^{(n)} \cap H_{j,i}^{(n)}) - \epsilon \leq 0$$

for every $\epsilon > 0$, implying

$$\lim_{n \to \infty} \mu \left(S_i^{(n)} \cap H_{j,i}^{(n)} \right) = 0 \qquad (13)$$

for all $1 \leq i < j \leq N$. Furthermore, since

$$\mu(S_i^{(n)} \cap H_{j,i}^{(n)}) = \mu(S_j^{(n)} \cap H_{i,j}^{(n)})$$

(13) holds for every $i \neq j$. Define the random variable $Y_n = Y_n(X)$ by

$$Y_n \triangleq \sum_{i=1}^N \sum_{1 \leq j \leq N, j \neq i} 1_{S_i^{(n)} \cap H_{j,i}^{(n)}}(X).$$

Then $Y_n \to 0$ in probability as $n \to \infty$, since for all $\epsilon > 0$

$$\Pr\{ |Y_n| > \epsilon \} \leq \Pr\{ Y_n > 0 \}$$
$$= \Pr \left\{ \bigcup_{i=1}^N \bigcup_{1 \leq j \leq N, j \neq i} \left\{ 1_{S_i^{(n)} \cap H_{j,i}^{(n)}}(X) = 1 \right\} \right\}$$
$$\leq \sum_{i=1}^N \sum_{1 \leq j \leq N, j \neq i} \mu \left(S_i^{(n)} \cap H_{j,i}^{(n)} \right) \to 0.$$

Therefore, there is a subsequence $\{ Y_{n_l} \}$ such that $\lim_l Y_{n_l}(X) = 0$ almost surely (see [12, Theorem 9.2.1]). Since $\lim_n 1_{H_{i,j}^{(n)}}(x) = 1_{H_{i,j}}(x)$ μ-a.e. by (3), we obtain for all $i \neq j$

$$0 = \lim_{l \to \infty} 1_{S_i^{(n_l)} \cap H_{j,i}^{(n_l)}}(x) = \lim_{l \to \infty} 1_{S_i^{(n_l)}}(x) 1_{H_{j,i}^{(n_l)}}(x)$$
$$= \lim_{l \to \infty} 1_{S_i^{(n_l)}}(x) 1_{H_{j,i}}(x) \ \mu\text{-a.e.}$$

It is easy to see that the preceding implies

$$\lim_{l \to \infty} 1_{S_i^{(n_l)} \setminus \hat{S}_i}(x) = 0 \ \mu\text{-a.e.}$$

and

$$\lim_{l \to \infty} 1_{\hat{S}_i \setminus S_i^{(n_l)}}(x) = 0 \ \mu\text{-a.e.}$$

Combining this with the identity $|1_A - 1_B| = 1_{A \setminus B} + 1_{B \setminus A}$ we conclude that for all $i = 1, \ldots, N$

$$\lim_{l \to \infty} 1_{S_i^{(n_l)}}(x) = 1_{\hat{S}_i}(x) \ \mu\text{-a.e.} \qquad (14)$$

Since $\mu(S_i^{(n)}) = \mu(S_i)$ for all n and $i = 1, \ldots, N$, the dominated convergence theorem gives

$$\mu(\hat{S}_i) = \mu(S_i), \qquad i = 1, \ldots, N$$

which also implies $\mu(\hat{S}_{N+1}) = 0$. Since $Q_{n_l}(x) = c_i$ if $x \in S_i^{(n_l)}$, and $\hat{Q}(x) = c_i$ if $x \in \hat{S}_i$, from (14) we obtain

$$\lim_{l \to \infty} Q_{n_l}(x) = \hat{Q}(x) \ \mu\text{-a.e.}$$

Therefore, Fatou's lemma [12] implies

$$
\begin{aligned}
D(\hat{Q}) &= \int_{\mathbb{R}^k} \|x - \hat{Q}(x)\|^2 \, d\mu(x) \\
&= \int_{\mathbb{R}^k} \liminf_{l \to \infty} \|x - Q_{n_l}(x)\|^2 \, d\mu(x) \\
&\leq \liminf_{l \to \infty} \int_{\mathbb{R}^k} \|x - Q_{n_l}(x)\|^2 \, d\mu(x) \\
&= \liminf_{l \to \infty} D(Q_{n_l}) \leq D(Q) \qquad (15)
\end{aligned}
$$

since $D(Q_{n+1}) \leq D(Q_n)$ for all n and $D(Q_1) = D(Q)$. Thus, $\hat{Q}$ satisfies the requirements of the theorem. □

Proof of Lemma 2: Assume $\liminf_n D(Q_n)$ is finite; otherwise, the statement is trivial. For positive integers n and i, let $p_i^{(n)} \triangleq \mu(S_i^{(n)})$. For all $1 \leq i < j \leq n$, let $(H_{i,j}^{(n)}, H_{j,i}^{(n)})$ denote the complementary half-spaces corresponding to $h_{i,j}^{(n)}$. Since $h_{i,j}^{(n)}$ separates $S_i^{(n)}$ and $S_j^{(n)}$, $S_i^{(n)} \subset H_{i,j}^{(n)}$ and $S_j^{(n)} \subset H_{j,i}^{(n)}$. For each n form the vector

$$v_n \triangleq \Big(c_1^{(n)}, c_2^{(n)}, h_{1,2}^{(n)}, c_3^{(n)}, h_{1,3}^{(n)}, h_{2,3}^{(n)}, c_4^{(n)}, \ldots, c_l^{(n)},$$

$$h_{1,l}^{(n)}, \ldots, h_{l-1,l}^{(n)}, c_{l+1}^{(n)}, \ldots \Big)$$

where $h_{i,j}^{(n)}$ stands for the parameter vector representing the hyperplane (see Section II) and is defined arbitrary for all $i < j$ if $j > n$. Now use Cantor's diagonal method to pick a subsequence of $\{v_n\}$ converging componentwise to a vector $v = (c_1, c_2, h_{1,2}, c_3, h_{1,3}, h_{2,3}, \ldots)$. (The convergence of each sequence $\{c_i^{(n)}\}_{n=1}^{\infty}$ is considered in $\mathbb{R}^k \cup \{\infty\}$, the one-point compactification of $\mathbb{R}^k$ [12], while the convergence of the last components of $h_{i,j}^{(n)}$, $n = 1, 2, \ldots$, is considered in $\overline{\mathbb{R}}$.) For simplicity, this subsequence is also denoted by $\{v_n\}$. Using v we can construct a quantizer $\hat{Q}$ by setting

$$\hat{Q} \equiv \Big(\big\{ \hat{S}_1, \hat{S}_2, \ldots \big\}, \{c_1, c_2, \ldots\} \Big)$$

where

$$\hat{S}_i = \bigcap_{j \neq i} H_{i,j}$$

where for all $i < j$, $(H_{i,j}, H_{j,i})$ is the pair of complementary half-spaces corresponding to $h_{i,j}$. In the rest of the proof, we show that $\hat{Q}$ is well defined and have the desired properties.

To simplify the notation, define $H_{i,i}^{(n)} = H_{i,i} = \mathbb{R}^k$ for all n and i. The $\hat{S}_i$ are clearly disjoint and convex. To see that they cover $\mathbb{R}^k$ with probability 1, notice that for all $i, l < n$

$$S_i^{(n)} \subset \bigcap_{j=1}^{l} H_{i,j}^{(n)} \quad \text{and} \quad \bigcup_{1 \leq j \leq l, j \neq i} S_j^{(n)} \subset \mathbb{R}^k \backslash \bigcap_{j=1}^{l} H_{i,j}^{(n)}.$$

Therefore, since $S_j^{(n)}$, $j = 1, 2, \ldots$, are disjoint

$$
\begin{aligned}
p_i^{(n)} = \mu\Big(S_i^{(n)}\Big) &\leq \mu\Big(\bigcap_{j=1}^{l} H_{i,j}^{(n)}\Big) \leq 1 - \sum_{1 \leq j \leq l, j \neq i} \mu\Big(S_j^{(n)}\Big) \\
&= 1 - \sum_{1 \leq j \leq l, j \neq i} p_j^{(n)}. \qquad (16)
\end{aligned}
$$

Since $\lim_n 1_{H_{i,j}^{(n)}}(x) = 1_{H_{i,j}}(x)$ μ-a.e., the dominated convergence theorem implies

$$p_i \leq \mu\Big(\bigcap_{j=1}^{l} H_{i,j}\Big) \leq 1 - \sum_{1 \leq j \leq l, j \neq i} p_j$$

for all i and l. Since the sequence of sets $\{\bigcap_{j=1}^{l} H_{i,j}\}_{l=1}^{\infty}$ decreases to $\hat{S}_i$, letting $l \to \infty$ we obtain for each i

$$p_i \leq \mu(\hat{S}_i) \leq 1 - \sum_{i \neq j} p_j = p_i.$$

Thus, $\mu(\hat{S}_i) = p_i$ and $\sum_{i=1}^{\infty} \mu(\hat{S}_i) = 1$, so $\hat{Q}$ has the desired cell probabilities.

Next, we show that for every i and Borel set $B \subset \hat{S}_i$

$$\lim_{n \to \infty} \mu(B \cap S_i^{(n)}) = \mu(B). \qquad (17)$$

If $i, l < n$, then

$$
\begin{aligned}
&\mu\Big(B \cap S_i^{(n)}\Big) \\
&= \mu\Big(B \cap \bigcap_{j=1}^{l} H_{i,j}^{(n)} \backslash \Big(\bigcap_{j=1}^{l} H_{i,j}^{(n)} \backslash S_i^{(n)}\Big)\Big) \\
&\geq \mu\Big(B \cap \bigcap_{j=1}^{l} H_{i,j}^{(n)}\Big) - \mu\Big(\bigcap_{j=1}^{l} H_{i,j}^{(n)} \backslash S_i^{(n)}\Big) \\
&= \mu\Big(B \backslash \Big(\bigcap_{j=1}^{l} H_{i,j} \backslash \bigcap_{j=1}^{l} H_{i,j}^{(n)}\Big)\Big) \\
&\quad - \mu\Big(\bigcap_{j=1}^{l} H_{i,j}^{(n)} \backslash S_i^{(n)}\Big) \\
&\geq \mu(B) - \mu\Big(\bigcap_{j=1}^{l} H_{i,j} \backslash \bigcap_{j=1}^{l} H_{i,j}^{(n)}\Big) - \Big(1 - \sum_{j=1}^{l} p_j^{(n)}\Big)
\end{aligned}
$$

where the first equality holds since $S_i^{(n)} \subset \bigcap_{j=1}^{l} H_{i,j}^{(n)}$, the second equality holds since $B \backslash \big(\bigcap_{j=1}^{l} H_{i,j}\big) = \emptyset$, and the last step follows from (16). Since $\lim_n h_{i,j}^{(n)} = h_{i,j}$, the dominated convergence theorem implies for each positive integer l

$$\liminf_{n \to \infty} \mu\Big(B \cap S_i^{(n)}\Big) \geq \mu(B) - \Big(1 - \sum_{j=1}^{l} p_j\Big).$$

Therefore, since $\mu(B) \geq \mu(B \cap S_i^{(n)})$ and $\sum_{j=1}^{\infty} p_j = 1$, (17) follows by letting $l \to \infty$.

Using (17), we next show that the codevectors of $\hat{Q}$ are well defined (i.e., $c_i \in \mathbb{R}^k$ for all i such that $\mu(\hat{S}_i) > 0$). Assume $\mu(\hat{S}_i) > 0$ (so that the interior of $\hat{S}_i$ is not empty) and let $B \subset \hat{S}_i$ be a closed ball with $\mu(B) > 0$. Suppose $c_i = \infty$. Then the partial distortion of the ith cell of Q_n has the lower bound

$$
\begin{aligned}
D_i(Q_n) &= \int_{S_i^{(n)}} \|x - c_i^{(n)}\|^2 \, d\mu(x) \\
&\geq \int_{B \cap S_i^{(n)}} \|x - c_i^{(n)}\|^2 \, d\mu(x) \\
&\geq \mu\Big(B \cap S_i^{(n)}\Big) \min_{x \in B} \|x - c_i^{(n)}\|^2.
\end{aligned}
$$

For all n large enough, we have $\mu(B \cap S_i^{(n)}) \geq \delta$ for some $\delta > 0$ by (17), and since we assumed $\lim_n c_i^{(n)} = c_i = \infty$

$$\lim_{n \to \infty} \min_{x \in B} \|x - c_i^{(n)}\| = \infty.$$

This would imply $\liminf_n D(Q_n) \geq \liminf_n D_i(Q_n) = \infty$, contradicting the assumption that $\liminf_n D(Q_n)$ is finite. Thus, $c_i \in \mathbb{R}^k$ for all i such that $\mu(\hat{S}_i) > 0$, and we conclude that $\hat{Q}$ is well defined.

It remains to show that $D(\hat{Q}) \leq \liminf_n D(Q_n)$. To prove this, we apply the method used in the proof of Theorem 1. Let

$$Y_i^{(n)}(X) \triangleq 1_{S_i^{(n)} \setminus \hat{S}_i}(X).$$

Then, as $n \to \infty$, for every $\epsilon > 0$ we have

$$P\left\{|Y_i^{(n)}(X)| > \epsilon\right\} \leq P\left\{Y_i^{(n)}(X) = 1\right\}$$
$$= \mu\left(S_i^{(n)}\right) - \mu\left(S_i^{(n)} \cap \hat{S}_i\right) \to 0 \qquad (18)$$

by (17). Thus, for all i, $Y_i^{(n)}(X) \to 0$ in probability as $n \to \infty$. Defining

$$\hat{Y}_i^{(n)}(X) \triangleq 1_{\hat{S}_i \setminus S_i^{(n)}}(X)$$

we similarly obtain that $\hat{Y}_i^{(n)}(X) \to 0$ in probability as $n \to \infty$ for all i. It follows that every subsequence of $\{Y_i^{(n)}, \hat{Y}_i^{(n)}\}$ has a subsequence, say $\{Y_i^{(n_t)}, \hat{Y}_i^{(n_t)}\}$, such that $\lim_t Y_i^{(n_t)} = 0$ and $\lim_t \hat{Y}_i^{(n_t)} = 0$ almost surely [12, Theorem 9.2.1]. Using Cantor's diagonal method, we can now choose an increasing sequence of positive integers $\{n_l\}$ such that for all $i \geq 1$, $\lim_l Y_i^{(n_l)} = 0$ and $\lim_l \hat{Y}_i^{(n_l)} = 0$ almost surely. In other words, for all i

$$\lim_{l \to \infty} 1_{S_i^{(n_l)} \setminus \hat{S}_i}(x) = 0 \quad \mu\text{-a.e.}$$

and

$$\lim_{l \to \infty} 1_{\hat{S}_i \setminus S_i^{(n_l)}}(x) = 0 \quad \mu\text{-a.e.}$$

which is equivalent to

$$\lim_{l \to \infty} 1_{S_i^{(n_l)}}(x) = 1_{\hat{S}_i}(x) \quad \mu\text{-a.e.} \qquad (19)$$

Since $Q_{n_l}(x) = c_i^{(n_l)}$ if $x \in S_i^{(n_l)}$, $\hat{Q}(x) = c_i$ if $x \in \hat{S}_i$, and $\lim_l c_i^{(n_l)} = c_i$, (19) implies

$$\lim_{l \to \infty} Q_{n_l}(x) = \hat{Q}(x) \quad \mu\text{-a.e.}$$

Using Fatou's lemma as in (15), we conclude that $D(\hat{Q}) \leq \liminf_l D(Q_{n_l})$, which completes the proof. $\square$

ACKNOWLEDGMENT

REFERENCES

[1] S. P. Lloyd, "Least squared quantization in PCM," Bell Labs., unpublished memorandum, 1957. Also, *IEEE Trans. Inform. Theory*, vol. IT-28, pp. 129–137, Mar., 1982.

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

[3] R. M. Gray and D. L. Neuhoff, "Quantization," *IEEE Trans. Inform. Theory*, vol. 44, pp. 2325–2383, Oct. 1998.

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

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

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

[7] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1990.

[8] A. György and T. Linder, "On the structure of optimal entropy-constrained scalar quantizers," *IEEE Trans. Inform. Theory*, vol. 48, pp. 416–427, Feb. 2002.

[9] M. Effros and D. Muresan, "Codecell contiguity in optimal fixed-rate and entropy-constrained network scalar quantization," in *Proc. Data Compression Conf., DCC 2002*, 2002, pp. 312–321.

[10] A. György, T. Linder, P. A. Chou, and B. J. Betts, "Do optimal entropy-constrained quantizers have a finite or infinite number of codewords?," preprint, June 2002. [Online]. Available: http://magenta.mast.queensu.ca/~linder/psfiles/GyLiChBe02.ps, under revision, to be published.

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

[12] R. M. Dudley, *Real Analysis and Probability*. New York: Chapman & Hall, 1989.

On the Stability of Distributed Sequence Adaptation for Cellular Asynchronous DS-CDMA Systems

Chi Wan Sung, *Member, IEEE,* and Kin Kwong Leung

Abstract—In this correspondence, we consider the sequence adaptation problem for cellular asynchronous code-division multiple-access (CDMA) systems. A game-theoretic approach is used to investigate the stability issues of distributed adaptation algorithms. It is shown that the Nash equilibrium may not exist for cellular CDMA systems if the traditional interference measure is used. In turn we propose a new interference measure which ensures system stability.

Index Terms—Cellular systems, code-division multiple access (CDMA), distributed algorithm, Nash equilibrium, signature sequence adaptation.

I. INTRODUCTION

Mitigating interference is of paramount importance in the design of a code-division multiple-access (CDMA) system. Much research has been conducted in designing multiuser detection techniques to suppress interference for a given set of signature sequences. Recently, there has been interest in managing interference from the transmitters' side. Optimum signature sequence sets for synchronous CDMA systems were characterized in [6], [10], [11]. These sets may be constructed by iterative methods [5], [9].

For practical applications, it is helpful to devise distributed algorithms that allow each user to adapt his own signature sequence based on local information. Rapajic and Vucetic suggested that each user iteratively replaces his signature sequence by the weight vector obtained at his minimum mean-squared error (MMSE) receiver [4]. This is called the MMSE algorithm, and its properties are investigated in [1], [7]–[9], [12]. Signature sequence adaptation in the presence of multipath was considered in [3].