

Codecell Contiguity in Optimal Fixed-Rate and Entropy-Constrained Network Scalar Quantizers *

Michelle Effros
Dept. of Electrical Engineering
California Institute of Technology
effros@caltech.edu

Dan Muresan
Dept. of Electrical Engineering
Stanford University
muresan@stanford.edu

Abstract

We consider the properties of optimal fixed-rate and entropy-constrained scalar quantizers for finite alphabet sources. In particular, we consider conditions under which the optimal scalar quantizer with contiguous codecells achieves performance no worse than the optimal scalar quantizer without the constraint of codecell contiguity. In addition to traditional scalar quantizers, we consider multi-resolution scalar quantizers and multiple description scalar quantizers and also look briefly at codes with decoder side information (Wyner-Ziv codes). While the conditions under which codecell contiguity is consistent with optimality in fixed-rate and entropy-constrained scalar quantization are quite broad, even with the squared error distortion measure, codecell contiguity in fixed-rate and entropy-constrained multi-resolution, multiple description, and Wyner-Ziv scalar quantization can preclude optimality for some sources.

1 Introduction

A scalar quantizer encoder defines a partition \mathcal{P} of the scalar source alphabet \mathcal{X} ; each set $c \in \mathcal{P}$ of the encoder's partition is a *codecell* describing all source symbols $x \in \mathcal{X}$ that map to a single binary description. Given an ordered, scalar alphabet, we say that encoder partition \mathcal{P} of \mathcal{X} has *contiguous* codecells if for any codecells $c, c' \in \mathcal{P}$ with $c \neq c'$, $x_i < x_j$ for some $x_i \in c$ and $x_j \in c'$ implies $x < x'$ for all $(x, x') \in c \times c'$. While it is common to think of the codecells of an optimal scalar quantizer as contiguous subsets (i.e., intervals) of the real line, in [1] György and Linder show that there exist finite-alphabet sources for which some rate-distortion points on the lower boundary (but not the convex hull) of the variable-rate scalar quantization operational distortion-rate bound cannot be achieved using an encoder with contiguous codecells.

*This material is based upon work partially supported by the NSF under Award No. CCR-9909026 and by the Caltech Lee Center for Advanced Networking.

We here consider the question of codecell contiguity for fixed-rate and entropy-constrained scalar quantization (SQ), multi-resolution scalar quantization (MRSQ),¹ and multiple description scalar quantization (MDSQ) with and without decoder side information. We focus on finite-alphabet sources on the real line.

2 Set-Up

Consider an ordered, finite, scalar alphabet $\mathcal{X} = \{x_1, \dots, x_N\} \subset \mathbb{R}$. Let $p[1] \cdots p[N]$ be an arbitrary probability mass function (pmf) on \mathcal{X} . We assume a non-negative distortion measure $d(x, \hat{x})$ and consider the properties of optimal fixed-rate and entropy-constrained SQ, MRSQ, and MDSQ for pmf $p[1] \cdots p[N]$ on alphabet $x_1 \cdots x_N$. For notational simplicity, we refer to scalar source alphabet $x_1 \cdots x_N$ by the symbol indices, writing $1 \cdots N$ to denote the full alphabet.

We define the encoders for SQ, MRSQ, and MDSQ using one or more partitions on $1 \cdots N$. Each partition divides the alphabet into non-overlapping codecells; each codecell describes all symbols given a particular binary description. We define an SQ encoder by a single partition \mathcal{P} ; each codecell $c \in \mathcal{P}$ describes a collection of source symbols given the same binary description. We define the encoder of an M -resolution MRSQ (M -RSQ) using M partitions $\mathcal{P}^M = (\mathcal{P}_1, \dots, \mathcal{P}_M)$; each codecell $c_i \in \mathcal{P}_i$ describes a collection of source symbols given the same binary description in resolutions $1 \cdots i$. We define the encoder of an M -description MDSQ (M -DSQ) using M partitions $\mathcal{P}^M = \mathcal{P}_1, \dots, \mathcal{P}_M$; each codecell $c_i \in \mathcal{P}_i$ describes a collection of source symbols given the same binary description in packet i .

In both fixed-rate and entropy-constrained coding, the expected distortions and rates of the optimal SQ, MRSQ, and MDSQ with a given partition or set of partitions are simple functions of the partitions. We next describe these functions. Given an encoder defined by partition \mathcal{P} , the expected distortion of the corresponding code is independent of the code type. In contrast, the expected rates vary with the code type. We begin with the distortion calculation.

The expected distortion for the optimal code defined by partition \mathcal{P} is

$$D(\mathcal{P}) = \sum_{C \in \mathcal{P}} \sum_{n \in C} p[n] d(x_n, \mu(C)).$$

Here $\mu(C)$ denotes the optimal codeword for codecell C . For example, when $d(x, \hat{x}) = (x - \hat{x})^2$, the optimal codeword $\mu(C)$ is the centroid $\mu(C) = \sum_{n \in C} p[n] x_n / p(C)$ of codecell C , where $p(C) = \sum_{n \in C} p[n]$ denotes the probability of codecell C .

The expected rate of a fixed-rate SQ (FR-SQ) with encoder partition \mathcal{P} is

$$R^{\text{fr}}(\mathcal{P}) = \log |\mathcal{P}| \tag{1}$$

bits. We measure expected rate for variable-rate codes by the codecell entropies. The expected rate of an entropy-constrained SQ (EC-SQ) with encoder partition \mathcal{P} is

$$R^{\text{vr}}(\mathcal{P}) = - \sum_{C \in \mathcal{P}} p(C) \log p(C). \tag{2}$$

¹Multi-resolution codes are also commonly known as progressive transmission, successive-refinement, embedded, or scalable codes.

The rate calculation for an M -RSQ relies on the successive refinement in its partitions. Partition \mathcal{P} *refines* \mathcal{P}' (written $\mathcal{P}' \succeq \mathcal{P}$) if for any $C \in \mathcal{P}'$ there exists a collection of cells $c_1 \cdots c_k \in \mathcal{P}$ so that $\mathcal{P}(C) = \{c_1 \cdots c_k\}$ is a partition of C . Since M -RSQ descriptions are nested, the M -RSQ partitions are also nested. To have the same description in resolutions $1 \cdots i$, two symbols must have the same descriptions in resolutions $1 \cdots i - 1$. Thus in any M -RSQ, $\mathcal{P}_1 \succeq \mathcal{P}_2 \succeq \cdots \succeq \mathcal{P}_M$.

The i th-resolution rate of an M -RSQ is the expected number of bits in an incremental description of codecell $c_i \in \mathcal{P}_i$ given the codecells $c_1 \cdots c_{i-1}$ described in previous resolutions. For fixed-rate MRSQ (FR-MRSQ), the rate in resolution i equals

$$R^{\text{fr}}(\mathcal{P}_i|\mathcal{P}_{i-1}) = \max_{C \in \mathcal{P}_{i-1}} \log |\mathcal{P}_i(C)|.$$

For entropy-constrained MRSQ (EC-MRSQ) the rate in resolution i equals

$$R^{\text{vr}}(\mathcal{P}_i|\mathcal{P}_{i-1}) = - \sum_{C \in \mathcal{P}_{i-1}} p(C) \sum_{c \in \mathcal{P}_i(C)} p(c|C) \log p(c|C),$$

where $p(c|C) = p(c)/p(C)$. (In both cases, $\mathcal{P}_0 = \{1 \cdots N\}$.)

For MDSQ, we assume that each packet description must be uniquely decodable when only that packet is received, so we again use (1) and (2). The expected rate in resolution i of a fixed-rate MDSQ (FR-MDSQ) is $R^{\text{fr}}(\mathcal{P}_i)$. The expected rate in resolution i of an entropy-constrained MDSQ (EC-MDSQ) is $R^{\text{vr}}(\mathcal{P}_i)$.

Given this set-up and notation, we next consider the properties of optimal partitions for fixed-rate and entropy-constrained SQ, MRSQ, and MDSQ.

3 Scalar Quantization

Let $\mathcal{D}^{\text{fr}} = \{(R^{\text{fr}}(\mathcal{P}), D(\mathcal{P}))\}$ and $\mathcal{D}^{\text{vr}} = \{(R^{\text{vr}}(\mathcal{P}), D(\mathcal{P}))\}$ describe the sets of expected rate-distortion points achievable through FR-SQ and EC-SQ on pmf $\{p[n]\}_{n=1}^N$. An FR-SQ that achieves a point on the lower boundary $D_1^{\text{fr}}(R) = \inf_{\mathcal{P}: R^{\text{fr}}(\mathcal{P}) \leq R} D(\mathcal{P})$ of \mathcal{D}^{fr} is *optimal* in the sense that it has the lowest distortion among all FR-SQs satisfying some rate constraint. An EC-SQ that achieves a point on the lower boundary $D_1^{\text{vr}}(R) = \inf_{\mathcal{P}: R^{\text{vr}}(\mathcal{P}) \leq R} D(\mathcal{P})$ of \mathcal{D}^{vr} is *optimal* in the sense that it has the lowest distortion among all EC-SQs satisfying some rate constraint. While $D_1^{\text{fr}}(R)$ and $D_1^{\text{vr}}(R)$ are not convex in general and any point on the lower convex hull of $D_1^{\text{fr}}(R)$ and $D_1^{\text{vr}}(R)$ can be achieved through time-sharing, the above definitions for optimal FR-SQs and EC-SQs does not require performance on the convex hull of $D_1^{\text{fr}}(R)$ or $D_1^{\text{vr}}(R)$; time-sharing strategies, while low in complexity, are not strict *scalar* codes.

We begin by proving the optimality of codecell contiguity for FR-SQ and EC-SQ. While codecell contiguity for FR-SQ is widely known for common distortion measures like the squared-error distortion measure, the proof sets the stage for later arguments and we include it for completeness.

Lemma 1 *Given a pmf $p[1] \cdots p[N]$ and a distortion measure $d(x, \hat{x}) = \rho(|x - \hat{x}|)$ for some non-decreasing function $\rho: [0, \infty) \rightarrow [0, \infty)$, any point $(R, D_1^{\text{fr}}(R)) \in \mathcal{D}^{\text{fr}}$ is achievable by a fixed-rate SQ with contiguous codecells.*

Proof: We prove that for any partition $\mathcal{P} = \{c_1, \dots, c_K\}$ with optimal codewords $\mu(c_1) \leq \dots \leq \mu(c_K)$, there exists another partition \mathcal{P}^* with contiguous codecells $\{c_k^*\}_{k=1}^K$ that satisfies $\sum_{k=1}^K \sum_{n \in c_k^*} p[n] d(x_n, \mu(c_k)) \leq D(\mathcal{P})$. Since $(R, D_1^{\text{fr}}(R)) \in \mathcal{D}^{\text{fr}}$ implies, by definition of \mathcal{D}^{fr} , the existence of an encoder partition \mathcal{P} such that $\log |\mathcal{P}| \leq R$ and $D(\mathcal{P}) = D_1^{\text{fr}}(R)$, this observation gives the desired result.

For each $k, l \in 1 \dots K$, let

$$c'_{k,l} = \begin{cases} \{n : d(x_n, \mu(c_k)) < d(x_n, \mu(c_l))\} & \text{if } l < k \\ \{n : d(x_n, \mu(c_k)) \leq d(x_n, \mu(c_l))\} & \text{if } l \geq k. \end{cases}$$

Then $d(x, \hat{x}) = \rho(|x - \hat{x}|)$ and ρ non-decreasing imply that the $c'_{k,l}$ are half-lines. For example, if ρ is strictly increasing, then $c'_{k,l} = \{n : x_n > (\mu(c_k) + \mu(c_l))/2\}$ for each $l < k$ and $c'_{k,l} = \{n : x_n \leq (\mu(c_k) + \mu(c_l))/2\}$ for each $l \geq k$. The set $c_k^* = \bigcap_{l=1}^K c'_{k,l}$, describes all $n \in 1 \dots N$ for which $\mu(c_k)$ is the closest codeword, with ties broken in favor of the smallest codeword. Since the $c'_{k,l}$ are half-lines, each c_k^* must be an interval. Finally, $\mathcal{P}^* = \{c_k^*\}_{k=1}^K$ partitions $1 \dots N$ (every x_n has a unique closest codeword) and minimizes expected distortion, giving the desired result. \square

We make two modifications in generalizing Lemma 1 from FR-SQ to EC-SQ. First, we consider only points on the convex hull of $\mathcal{D}^{\text{vr}}(R)$. Second, we assume ρ is convex. The first constraint is practically motivated since the Lagrangian performance measure used for EC-SQ design can at best find the lower convex hull of $\mathcal{D}^{\text{vr}}(R)$; it is also theoretically motivated since there exist points on $\mathcal{D}^{\text{vr}}(R)$ (but not on its convex hull) that cannot be achieved with contiguous codecells [1]. The motivation for the second constraint is discussed after Theorem 1.

Theorem 1 *Given a pmf $p[1] \dots p[N]$ and a distortion measure $d(x, \hat{x}) = \rho(|x - \hat{x}|)$ where $\rho : [0, \infty) \rightarrow [0, \infty)$ is convex and non-decreasing, any point $(R, D_1^{\text{vr}}(R)) \in \mathcal{D}^{\text{vr}}$ on the lower convex hull of $D_1^{\text{vr}}(\cdot)$ is achievable by an EC-SQ with contiguous code cells.*

Proof: For any point $(R, D_1^{\text{vr}}(R)) \in \mathcal{D}^{\text{vr}}$, there exists a partition \mathcal{P} with $R^{\text{vr}}(\mathcal{P}) \leq R$ and $D(\mathcal{P}) \leq D_1^{\text{vr}}(R)$. We assume without loss of generality that $|\mathcal{P}| \leq N < \infty$ since any partition with more than N codecells must include empty codecells and empty codecells cannot improve EC-SQ performance. If $(R, D_1^{\text{vr}}(R)) \in \mathcal{D}^{\text{vr}}$ is on the lower convex hull of $D_1^{\text{vr}}(\cdot)$, then there exists a $\lambda > 0$ such that \mathcal{P} minimizes $D(\mathcal{P}) + \lambda R^{\text{vr}}(\mathcal{P})$ over all partitions on $1 \dots N$. Let $\mathcal{P} = \{c_1, \dots, c_K\}$ ($K \leq N$) with optimal codewords $\mu(c_1) \leq \dots \leq \mu(c_K)$. We next construct a contiguous codecell partition $\mathcal{P}^* = \{c_k^*\}_{k=1}^K$ that satisfies $\sum_{k=1}^K \sum_{n \in c_k^*} p[n] (d(x_n, \mu(c_k)) + \lambda r(c_k)) \leq D(\mathcal{P}) + \lambda R^{\text{vr}}(\mathcal{P})$.

For any $x \in [x_1, x_N]$ and $c \in c_1 \dots c_K$, let $j(x, c) = d(x, \mu(c)) + \lambda r(c)$. Then for each $k, l \in 1 \dots K$, let

$$c'_{k,l} = \begin{cases} \{n : j(x_n, c_k) < j(x_n, c_l)\} & \text{if } l < k \\ \{n : j(x_n, c_k) \leq j(x_n, c_l)\} & \text{if } l \geq k. \end{cases}$$

From [1, Lem. 1], if $\rho(x)$ is convex and non-decreasing in x , then $\rho(|x - \hat{x}_1|) - \rho(|x - \hat{x}_2|)$ is monotonic in x . Since $\rho(|x - \hat{x}_1|) - \rho(|x - \hat{x}_2|)$ is monotonic in x , $j(x, c_k) - j(x, c_l)$ is monotonic in x for each (k, l) . As a result, each non-empty $c'_{k,l}$ is a half line. The

set $c_k^* = \cap_{l=1}^K c'_{k,l}$, describes all $n \in 1 \cdots N$ for which $\mu(c_k)$ is the “closest” codeword by this modified nearest neighbor distortion measure, with ties broken in favor of the smallest codeword. Since the $c'_{k,l}$ are half-lines, each c_k^* must be an interval. Again $\mathcal{P}^* = \{c_k^*\}_{k=1}^K$ partitions $1 \cdots N$, and thus we have the desired result. \square

If ρ is non-decreasing and not convex, then codecell contiguity may preclude encoder optimality for some source distributions. In particular, guaranteeing a unique solution to $\rho(|x - a|) + \lambda r_1 = \rho(|x - b|) + \lambda r_2$ requires monotonicity of $\rho(|x - a|) - \rho(|x - b|)$ in x . We next show that monotonicity of $\rho(|x - a|) - \rho(|x - b|)$ in x for all a, b requires convexity of ρ .

For any fixed $a < b$, let

$$\begin{aligned} f(x) &= \rho(|x - a|) - \rho(|x - b|) \\ &= \begin{cases} \rho(a - x) - \rho(b - x) & \text{if } x < a \\ \rho(x - a) - \rho(b - x) & \text{if } a \leq x \leq b \\ \rho(x - a) - \rho(x - b) & \text{if } x > b. \end{cases} \end{aligned}$$

We next consider the conditions under which $f(x)$ is monotone. If ρ is non-decreasing, then $f(x)$ is non-decreasing on $[a, b]$. This leaves two possibilities: either $f(x)$ is constant on $[a, b]$ or $f(a) < f(b)$. The first case cannot be true for all $[a, b]$ (that is all codewords) unless $\rho(x)$ is linear in x , which gives a convex function. In the second case, monotonicity of $f(x)$ requires that $\rho(a - x) - \rho(b - x)$ is non-decreasing for all $x < a$ and that $\rho(x - a) - \rho(x - b)$ is non-decreasing for all $x > b$. The second condition gives $\rho(x - a) - \rho(x - b) \leq \rho(x + \Delta - a) - \rho(x + \Delta - b)$ or

$$\rho(x + \Delta - b) - \rho(x - b) \leq \rho(x + \Delta - a) - \rho(x - a)$$

for all $x > a$ and any $\Delta > 0$. Achieving this result for all a, b requires the convexity of ρ . Since, we can construct a pmf with optimal codewords at locations (a, b) for any a, b , guaranteeing the monotonicity of $\rho(|x - \mu(c_1)|) - \rho(|x - \mu(c_2)|)$ for all pmfs requires the convexity of $\rho(x)$ in x .

4 Multi-Resolution Scalar Quantization

Next, consider the properties of the partitions of optimal FR-MRSQ and EC-MRSQ. For any $R^M = R_1, \dots, R_M$ with $R_i \geq 0$ for all $i \in 1 \cdots M$, the rate-distortion points achievable by fixed-rate- R^M and entropy-constrained M -RSQs are

$$\begin{aligned} \mathcal{D}^{\text{fr}}(R^M) &= \{(R_i, D(\mathcal{P}_i))|_{i=1}^M : \mathcal{P}^M \in \mathcal{P}^{\text{fr}}(R^M)\} \\ \mathcal{D}^{\text{vr}} &= \{(R^{\text{vr}}(\mathcal{P}_i|\mathcal{P}_{i-1}), D(\mathcal{P}_i))|_{i=1}^M : \mathcal{P}^M \in \mathcal{P}^{\text{vr}}\} \end{aligned}$$

respectively, where

$$\begin{aligned} \mathcal{P}^{\text{fr}}(R^M) &= \{\mathcal{P}^M : \mathcal{P}_M \succeq \cdots \succeq \mathcal{P}_1 \wedge R^{\text{fr}}(\mathcal{P}_i|\mathcal{P}_{i-1}) \leq R_i \forall i\} \\ \mathcal{P}^{\text{vr}} &= \{\mathcal{P}^M : \mathcal{P}_M \succeq \cdots \succeq \mathcal{P}_1\}. \end{aligned}$$

An M -RSQ that achieves a point on a lower boundary of $\mathcal{D}^{\text{fr}}(R^M)$ (for any non-negative R^M) or \mathcal{D}^{vr} is in some sense optimal. (In fixed-rate- R^M coding, each lower

boundary describes the minimal value of some D_j subject to constraints on $\{D_i\}_{i \neq j}$. In entropy-constrained coding, each lower boundary describes the minimal value of some D_i or R_i subject to constraints on the remaining rates and distortions.)

While the lower boundaries of $\mathcal{D}^{\text{fr}}(R^M)$ and \mathcal{D}^{vr} are not convex in general, optimal M -RSQ design algorithms focus on achieving points on the lower convex hull of $\mathcal{D}^{\text{fr}}(R^M)$ and \mathcal{D}^{vr} . For any $\mathcal{P}_*^M \in \mathcal{P}^{\text{fr}}(R^M)$ that achieves a point on the lower convex hull of $\mathcal{D}^{\text{fr}}(R^M)$, there exists a positive vector $\nu^M = \{\nu_i\}_{i=1}^M$ for which \mathcal{P}_*^M minimizes the Lagrangian

$$J^{\text{fr}}(\mathcal{P}^M, R^M, \nu^M) = \sum_{i=1}^M \nu_i D(\mathcal{P}_i)$$

over all $\mathcal{P}^M \in \mathcal{P}^{\text{fr}}(R^M)$. Similarly, for any $\mathcal{P}_*^M \in \mathcal{P}^{\text{vr}}$ that achieves a point on the lower convex hull of \mathcal{D}^{vr} there exist positive vectors $\nu^M = \{\nu_i\}_{i=1}^M$ and $\lambda^M = \{\lambda_i\}_{i=1}^M$ such that $\mathcal{P}_*^M \in \mathcal{P}^{\text{vr}}$ minimizes the Lagrangian

$$J^{\text{vr}}(\mathcal{P}^M, \nu^M, \lambda^M) = \sum_{i=1}^M [\nu_i D(\mathcal{P}_i) + \lambda_i R^{\text{vr}}(\mathcal{P}_i | \mathcal{P}_{i-1})]$$

over all $\mathcal{P}^M \in \mathcal{P}^{\text{vr}}$.

We next prove the existence of pmfs for which the constraint of codecell contiguity in \mathcal{P}_1 precludes optimality even for points on the lower convex hull of $\mathcal{D}^{\text{fr}}(R^M)$ with $d(x, \hat{x}) = \rho(|x - \hat{x}|)$ and ρ convex. One such example follows.

Example 1: Let $d(x, \hat{x}) = (x - \hat{x})^2$, and consider pmf $\{1/8, 1/8, 3/8, 3/8\}$ on alphabet $\{20, 40, 60, 140\}$. Then for $M = 2$ there exists a point on the lower convex hull of $\mathcal{D}^{\text{fr}}((1, 1))$ that cannot be achieved with codecell contiguity. In particular, in order to minimize D_1 with incremental rates $R_1 = 1$ and $R_2 = 1$ and distortion $D_2 = 0$, we must use non-contiguous codecells at resolution 1. Equivalently, for all ν_1, ν_2 such that $\nu_1/\nu_2 < .02695$, optimality requires non-contiguous codecells at resolution 1. The optimal partitions (in terms of the symbol alphabet rather than the symbol indices) are $\mathcal{P}_1 = \{\{20, 60\}, \{40, 140\}\}$ and $\mathcal{P}_2 = \{\{20\}, \{60\}, \{40\}, \{140\}\}$. Figure 1(a) shows the corresponding optimal 2RSQ codebook.

Intuitively, requiring incremental rates of 1 and 1 leads to a binary tree, and using a value of ν_2 high enough to force D_2 to zero forces each source symbol to occupy a distinct leaf at depth 2 in that tree. The only task, then, is to find the partition \mathcal{P}_1 that minimizes $D(\mathcal{P}_1)$ over all $\mathcal{P}_1 = \{c_1, c_2\}$ with $|c_1| = |c_2| = 2$. Here clustering a low-probability symbol with each high probability symbol leads to a lower expected distortion than could be achieved if the closer, more probable symbols were grouped together. This clustering choice would never be necessary in a single-resolution code since the constraint $|c_1| = |c_2| = 2$ does not arise there.

The same problem can arise for entropy-constrained MRSQs achieving performance on the lower convex hull of \mathcal{D}^{vr} , as shown in Example 2.

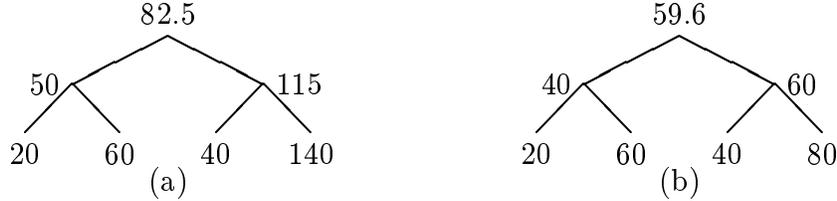


Figure 1: (a) The optimal fixed-rate 2RSQ codebook for the scenario described in Example 1 requires non-contiguous codecells at resolution 1. The codebook’s resolution-1 partition is $\mathcal{P}_1 = \{\{20, 60\}, \{40, 140\}\}$. (b) The optimal entropy-constrained 2RSQ codebook for the scenario described in Example 2 requires non-contiguous codecells in \mathcal{P}_1 . Here $\mathcal{P}_1 = \{\{20, 60\}, \{40, 80\}\}$.

Example 2: Consider $M = 2$, $d(x, \hat{x}) = (x - \hat{x})^2$, pmf $\{1/8, 3/8, 1/8, 3/8\}$, and alphabet $\{20, 40, 60, 80\}$. If $\lambda_2 + 31.36\nu_1 < \lambda_1 < \lambda_2 + 5951.16\nu_1$ and ν_2 is sufficiently large to force D_2 to 0, then achieving the optimal performance requires non-contiguous codecells at resolution 1. The optimal partitions (in terms of the symbol alphabet not the symbol indices) are $\mathcal{P}_1 = \{\{20, 60\}, \{40, 80\}\}$ and $\mathcal{P}_2 = \{\{20\}, \{60\}, \{40\}, \{80\}\}$. The resolution-1 codewords are $\mu(\{20, 60\}) = 40$, $\mu(\{40, 80\}) = 60$, $R^{\text{vr}}(\mathcal{P}_1) = .1414$, and $R^{\text{vr}}(\mathcal{P}_2|\mathcal{P}_1) = 1$. Figure 1(b) shows the optimal 2RSQ codebook.

This example treats another case where ν_2 is chosen sufficiently large to force D_2 to 0. Here, the tree structure need not be binary; nonetheless, clustering together two high-probability and two low-probability events yields a better rate-distortion tradeoff for the given parameter values than any alternative clustering at resolution 1. Again, the constraint that each source symbol must occupy a separate leaf at the second resolution of the tree plays a critical role; without a competing constraint in another resolution, the need for non-contiguous codecells would not arise here.

Theorem 2 *Requiring codecell contiguity in partition \mathcal{P}_1 of a fixed-rate or entropy-constrained M -RSQ with $M > 1$ precludes optimality for some finite-alphabet sources, even when $d(x, \hat{x}) = \rho(|x - \hat{x}|)$ with $\rho : [0, \infty) \rightarrow [0, \infty)$ non-decreasing and convex.*

The examples given above prove the theorem. Since the squared error distortion measure is popular for practical coding applications, Examples 1 and 2 treat the case where $d(x, \hat{x}) = (x - \hat{x})^2$. Similar examples exist for other distortion measures [2].

There exist distortion measures for which the assumption of codecell contiguity in the finest partition (\mathcal{P}_M in an M -RSQ) rather than the coarsest partition (\mathcal{P}_1 in Theorem 2) does not preclude optimality. Theorem 3 proves this result for FR-MRSQ with $d(x, \hat{x}) = (x - \hat{x})^2$. The same argument extends immediately to EC-MRSQ and fixed-rate and entropy-constrained MDSQ and restricted MDSQ [2].

Theorem 3 *Given pmf $p[1] \cdots p[N]$ and distortion measure $d(x, \hat{x}) = (x - \hat{x})^2$, any point $(R^M, D^M) \in \mathcal{D}^{\text{fr}}(R^M)$ that sits on the lower convex hull of $\mathcal{D}^{\text{fr}}(R^M)$ is achievable by a fixed-rate MRSQ with contiguous code cells in \mathcal{P}_M .*

Proof: For any point $(R^M, D^M) \in \mathcal{D}^{\text{fr}}(R^M)$ on the lower convex hull of $\mathcal{D}^{\text{fr}}(R^M)$, there exist partitions $\mathcal{P}^M \in \mathcal{PP}^{\text{fr}}(R^M)$ with $D(\mathcal{P}_i) = D_i$ for each $i \in 1 \cdots M$. Since (R^M, D^M) is on the lower convex hull of $\mathcal{D}^{\text{fr}}(R^M)$, there exists a non-negative vector ν^M for which \mathcal{P}^M minimizes $J^{\text{fr}}(\mathcal{P}^M, R^m, \nu^M)$ over all $\mathcal{P}^M \in \mathcal{PP}^{\text{fr}}(R^M)$. Label the codecells of each partition in \mathcal{P}^M as $\mathcal{P}_i = \{c_{i,1}, \dots, c_{i,K(i)}\}$ ($K(i) \leq 2^{\sum_{j=1}^i R_j}$) with optimal codewords $\mu(c_{i,1}), \dots, \mu(c_{i,K(i)})$ for each $i \in 1 \cdots M$. Without loss of generality, we can index the codecells of \mathcal{P}_M so that $\mu(c_{M,1}) \leq \mu(c_{M,2}) \leq \dots \leq \mu(c_{M,K(M)})$. Since \mathcal{P}_{i+1} refines \mathcal{P}_i for each $i \in 1 \cdots M-1$, let $a(i, k)$ denote the index of the resolution- i codecell from which codecell $c_{M,k}$ descends. (Here $i \in 1 \cdots M$ and $a(M, k) = k$ for all $k \in 1 \cdots K(M)$.) We next construct partitions $\mathcal{P}^{\star M} \in \mathcal{PP}^{\text{fr}}(R^M)$ so that $\mathcal{P}^{\star M}$ has contiguous codecells and the expected Lagrangian performance based on $\mathcal{P}^{\star M}$ is no worse than that based on \mathcal{P}^M .

For any $x \in [x_1, x_N]$ and any $c_{M,k}$ with $k \in 1 \cdots K(M)$, let

$$j(x, c_{M,k}) = \sum_{i=1}^M \nu_i d(x, \mu(c_{i,a(i,k)})).$$

Then $j(x, c_{M,k})$ describes the Lagrangian performance of the unique M -resolution description with codecell $c_{M,k}$ in resolution M . For each $k, l \in 1 \cdots K$, let

$$c'_{k,l} = \begin{cases} \{n : j(x_n, c_{M,k}) < j(x_n, c_{M,l})\} & \text{if } l < k \\ \{n : j(x_n, c_{M,k}) \leq j(x_n, c_{M,l})\} & \text{if } l \geq k. \end{cases}$$

Since $d(x, \hat{x}) = (x - \hat{x})^2$, the difference $j(x, c_{M,k}) - j(x, c_{M,l})$ is linear in x for all $k \neq l$. (The quadratic terms for k and l , which are weighted by the same constants ν_1, \dots, ν_M , cancel.) As a result, each non-empty $c'_{k,l}$ is a half line, and the set $c^{\star}_{M,k} = \cap_{l=1}^K c'_{k,l}$, describes all $n \in 1 \cdots N$ for which $\{\mu(c^{\star}_{i,a(i,k)})\}_{i=1}^M$ is the M -resolution reproduction with the largest Lagrangian performance. (Ties are broken in favor of the smallest resolution- M codeword.)

Since the $c'_{k,l}$ are half-lines, each $c^{\star}_{M,k}$ must be an interval. The partition $\mathcal{P}^{\star M} = \{c^{\star}_{M,k}\}_{k=1}^M$, together with the ancestor relationships described by $a(\cdot, \cdot)$, describe a partition $\mathcal{P}^{\star M} \in \mathcal{PP}^{\text{fr}}(R^M)$ with $\mathcal{P}_i^{\star} = \{c^{\star}_{i,k}\}_{k=1}^{K(i)}$ and $c^{\star}_{i,k} = \cup_{k' \in 1 \cdots K(M): a(i,k')=k} c^{\star}_{M,k'}$. (Here $|\mathcal{P}_i^{\star}| = |\mathcal{P}_i|$ for all i by construction.) Since $\mathcal{P}^{\star M}$ minimizes the Lagrangian performance with respect to the given codebook,

$$\sum_{k=1}^{K(M)} \sum_{n \in c^{\star}_{M,k}} p[n] \sum_{i=1}^M \nu_i d(x_n, \mu(c_{i,a(i,k)})) \leq J^{\text{fr}}(\mathcal{P}^M, R^m, \nu^M),$$

giving the desired result. \square

5 Multiple Description Scalar Quantization

The M partitions defining an M -DSQ induce $2^M - M - 1$ non-trivial *intersection partitions*. For any $\mathcal{M} \subseteq \{1 \cdots M\}$, the intersection partition for $\{\mathcal{P}_i\}_{i \in \mathcal{M}}$ is defined

as $\mathcal{P}_{\mathcal{M}} = \{\cap_{i \in \mathcal{M}} c_i : c_i \in \mathcal{P}_i\}$. We here assume that the packet-loss scenarios corresponding to all of these in tersection partitions occur with nonzero probability and thus that all corresponding distortions are constrained.

The optimal expected rates and distortions of an MDSQ are the rates associated with the M defining partitions and the distortions associated with the 2^M partitions achievable through the 2^M possible packet-loss scenarios. (These 2^M partitions include the M defining partitions, the $2^M - M - 1$ non-trivial intersection partitions, and the single trivial partition $\mathcal{P}_{\phi} = \mathcal{X}$ used when all M packets are lost. As a result, the rate-distortion points achievable by fixed-rate- R^M MDSQ are

$$\mathcal{D}^{\text{fr}}(R^M) = \left\{ \left(R^{\text{fr}}(\mathcal{P}_i) \Big|_{i=1}^M, D(\mathcal{P}_{\mathcal{M}}) \Big|_{\mathcal{M} \subseteq \{1 \dots M\}} \right) : R^{\text{fr}}(\mathcal{P}_i) \leq R_i \forall i \in 1 \dots M \right\},$$

and the rate-distortion points achievable by variable-rate MDSQ are

$$\mathcal{D}^{\text{vr}} = \left\{ \left(R^{\text{vr}}(\mathcal{P}_i) \Big|_{i=1}^M, D(\mathcal{P}_{\mathcal{M}}) \Big|_{\mathcal{M} \subseteq \{1 \dots M\}} \right) \right\}.$$

An M -DSQ that achieves a point on a lower boundary of $\mathcal{D}^{\text{fr}}(R^M)$ (for any non-negative R^M) or \mathcal{D}^{vr} is in some sense optimal.

Again, the lower boundaries of $\mathcal{D}^{\text{fr}}(R^M)$ and \mathcal{D}^{vr} are not convex in general, but optimal M -DSQ design algorithms, like optimal M -RSQ design algorithms, focus on achieving points on the lower convex hull of $\mathcal{D}^{\text{fr}}(R^M)$ and \mathcal{D}^{vr} . For any \mathcal{P}_{\star}^M that meets the R^M rate constraint and achieves a point on the lower convex hull of $\mathcal{D}^{\text{fr}}(R^M)$, there exists a positive vector $\underline{\nu} = \{\nu_{\mathcal{M}}\}_{\mathcal{M} \subseteq \{1 \dots M\}}$ for which \mathcal{P}_{\star}^M minimizes the Lagrangian

$$J^{\text{fr}}(\mathcal{P}^M, R^M, \underline{\nu}) = \sum_{\mathcal{M} \subseteq \{1 \dots M\}} \nu_{\mathcal{M}} D(\mathcal{P}_{\mathcal{M}})$$

over all \mathcal{P}^M with $R^{\text{fr}}(\mathcal{P}_i) \leq R_i$ for all $i \in \{1, \dots, M\}$. Similarly, for any \mathcal{P}_{\star}^M that achieves a point on the lower convex hull of \mathcal{D}^{vr} there exist positive vectors $\underline{\nu} = \{\nu_{\mathcal{M}}\}_{\mathcal{M} \subseteq \{1 \dots M\}}$ and $\lambda^M = \{\lambda_i\}_{i=1}^M$ such that \mathcal{P}_{\star}^M minimizes the Lagrangian

$$J^{\text{vr}}(\mathcal{P}^M, R^M, \underline{\nu}) = \sum_{\mathcal{M} \subseteq \{1 \dots M\}} \nu_{\mathcal{M}} D(\mathcal{P}_{\mathcal{M}}) + \sum_{i=1}^M \lambda_i R^{\text{vr}}(\mathcal{P}_i)$$

over all \mathcal{P}^M .

Notice that MRSQ can be considered as a special case of MDSQ with packet-loss scenarios restricted to $\mathcal{M} \in \{\{1\}, \{1, 2\}, \dots, \{1 \dots M\}\}$. As a result, it is not surprising that MDSQ and other restricted versions of MDSQ exhibit codecell contiguity properties similar to those observed in MRSQ. In particular, the above observations that the assumption of codecell contiguity precludes optimality in FR-MRSQ and EC-MRSQ for some source distributions and some Lagrangian parameters generalizes to MDSQ and restricted MDSQ as well. We can, however, guarantee codecell contiguity in $\mathcal{P}_{\{1 \dots M\}}$ for $d(x, \hat{x}) = (x - \hat{x})^2$. Details appear in [2].

6 Side-Information Codes

We conclude with a brief look at side-information codes. In this case, there exist very simple examples in which we cannot achieve optimality using codes with contiguous

codecells. This problem can occur even in FR-SQ with $d(x, \hat{x}) = \rho(|x - \hat{x}|)$ for a well-behaved function ρ . The following example illustrates the problem.

Example 3: Consider $M = 2$, $d(x, \hat{x}) = (x - \hat{x})^2$, pmf $\{1/4, 1/4, 1/4, 1/4\}$, and alphabet $\{20, 40, 60, 80\}$. Suppose that the side information takes on values from $\{1, 2\}$ with $\Pr(1|X = 20) = \Pr(1|X = 40) = \Pr(2|X = 60) = \Pr(2|X = 80) = .999$. The side information is available only to the decoder. Then the optimal fixed-rate-1 SQ with decoder side information requires non-contiguous codecells. In particular, $\mathcal{P} = \{\{20, 60\}, \{40, 80\}\}$. Here the side information can be viewed as the output of a random code with contiguous codecells. The encoder can best add to that information using non-contiguous codecells.

7 Summary and Conclusions

This work treats the properties of optimal codes for fixed-rate and entropy-constrained SQ, MRSQ, and MDSQ with and without decoder side information. In particular, we derive a family of conditions under which we can guarantee the existence of an optimal code with contiguous codecells. We also demonstrate that for some code types there exist pmfs for which some points on the lower convex hull of the achievable rate-distortion region cannot be achieved with an encoder with contiguous codecells. For codes without decoder side information, it is interesting to notice that in the examples considered to date, most (or all) of the points on the lower convex hull can be achieved with contiguous codecells. Nonetheless, the existence of pmfs for which codecell contiguity precludes optimality at even one rate implies that code designs that restrict their attention to codes with contiguous codecells cannot guarantee optimality for all pmfs. Thus optimal fixed-rate and entropy-constrained SQ design algorithms that use distortion measures satisfying the conditions of Lemma 1 and Theorem 1, respectively, and restrict their attention to codes with contiguous codecells (e.g., [3, 2]) guarantee globally optimal design. In contrast fixed- and entropy-constrained MRSQ, MDSQ, and Wyner-Ziv code design algorithms that restrict their attention to codes with contiguous codecells [2] cannot guarantee globally optimal design for all pmfs.

References

- [1] A. György and T. Linder. On the structure of optimal entropy-constrained scalar quantizers. Preprint, Dec. 2000.
- [2] D. Muresan and M. Effros. Quantization as histogram segmentation: globally optimal scalar quantizer design in network systems. 2001. In preparation.
- [3] J. D. Bruce. *Optimum Quantization*. PhD thesis, M.I.T., Cambridge, May 1964.