Distributed Structures, Sequential Optimization, and Quantization for Detection

Michael A. Lexa, *Student Member, IEEE*
and Don H. Johnson, *Fellow, IEEE*

***Abstract*—In the design of distributed quantization systems one inevitably confronts two types of constraints—those imposed by a distributed system's structure and those imposed by how the distributed system is optimized. Structural constraints are inherent properties of any distributed quantization system and are normally summarized by functional relationships defining the inputs and outputs of their component quantizers. The use of suboptimal optimization methods are often necessitated by the computational complexity encountered in distributed problems. This correspondence briefly explores the impact and interplay of these two types of constraints in the context of distributed quantization for detection. We introduce two structures that exploit inter-quantizer communications and that represent extremes in terms of their structural constraints. We then develop a sequential optimization scheme that maximizes the Kullback-Leibler divergence, takes advantage of statistical dependencies in the distributed system's output variables and leads to simple parameterizations of the component quantization rules. We present an illustrative example from which we draw insights into how these constraints influence the quantization boundaries and affect performance relative to lower and upper bounds.**

***Index Terms*—Quantization for detection, sequential optimization, Kullback-Leibler divergence, distributed detection.**

I. Introduction

Structure is often imposed on unconstrained vector quantization problems to reduce the encoding and decoding complexity [1]. In *distributed or decentralized* quantization systems, where a set of spatially separated quantizers act collectively to quantize an input vector [2], structure plays a similar role in that it often eases processing complexity; however, a distributed system's structure is an inherent property of the system and not one artificially imposed to reduce complexity. That said, structural constraints often make the joint optimization of a distributed quantization system's component quantizers difficult or even intractable. For example, determining the optimal decision (quantization) rules in a standard distributed detection problem (which can be thought of as a distributed quantization problem) is known to be NP-complete [3]. Thus, while a distributed quantizer's structure may *decrease* encoding/decoding complexity it may simultaneously *increase* the computational complexity of its optimization. This dichotomy frequently forces designers to use suboptimal methods to determine a distributed system's quantization rules. In these cases, a quantizer's performance (in comparison to its centralized counterpart) suffers from constraints imposed by a suboptimal optimization scheme, as well as those imposed by a distributed system's structure. This correspondence investigates the impact of these two types of constraints in the context of a distributed quantization for detection problem, where a distributed quantizer is designed such that a downstream detector is optimized [4, 5]. We consider two simple distributed quantization structures that sequentially process their inputs and share information among their constituent quantizers (Fig. 1). We then present a sequential optimization method that, like other sequential schemes [6], largely avoids the computational complexity of finding a globally optimal solution and that takes into account some of the statistical dependencies among the output variables. Because the structures considered here operate sequentially, there is a non-trivial interplay between their structural constraints and the constraints imposed by the sequential optimization.

In the *input-broadcast* system (Fig. 1), the quantizers sequentially broadcast their real-valued observation to all succeeding quantizers before quantizing their own observation; that is, the m^{th} quantizer broadcasts its data to all quantizers $k, k > m$ before quantizing its observation. In contrast, the quantizers in the *output-broadcast* system first quantize their observations and then broadcast the quantized output to all succeeding quantizers. Thus, any particular quantizer in the input-broadcast structure has access to all preceding observations, whereas a quantizer in the output-broadcast structure processes one observation and $m-1$ quantized outputs. In both structures, a detector receives all quantized outputs and ultimately (perhaps only after receiving a sequence of outputs) makes a decision about the hypothesized distribution of the inputs. Here, we restrict the outputs of the output-broadcast system to be binary so that the structures represent extremes in terms of the communication rate among the quantizers and so that the systems' structural constraints are accentuated. We recognize that an input-broadcast system centralizes all of the observations at the last quantizer, and hence could make an optimal centralized decision without a follow-on detector. We analyze the input-broadcast system primarily because it represents an ideal case where there is no communication rate constraint on the transmissions.

We optimize the distributed structures with respect to the Kullback-Leibler (KL) divergence which is well-known to be the optimal asymptotic exponential error rate for Neyman-Pearson type tests [7, p. 77]. Thus maximizing this quantity at the output of the distributed system, maximizes the potential asymptotic error decay rate of the detector.

II. Problem formulation

For ease of presentation, we restrict attention to structures with two quantizers. Let $\mathbf{X} = (X_1, X_2)$ denote the input data and $\mathbf{Y} = (Y_1, Y_2)$ the quantizers' outputs. Assume $\mathbf{X}$ is a real-valued random vector, and $\mathbf{Y}$ a binary-valued random vector: $\mathbf{X} \in \mathbb{R}^2, \mathbf{Y} \in \{0,1\} \times \{0,1\}$. Let $p_{\mathbf{X}}$ and $p_{\mathbf{Y}}$ denote the joint probability density function and the joint probability mass function of $\mathbf{X}$ and $\mathbf{Y}$, respectively. Let p_{X_m}, $m = 1, 2$, denote the marginal densities of $p_{\mathbf{X}}$, and p_{Y_m} $m = 1, 2$, the marginal probability mass functions of $p_{\mathbf{Y}}$. The realization $X_m = x_m$ represents the input to the m^{th} quantizer and $Y_m = y_m$ represents the corresponding output. The ordering of the quantizers is arbitrary, but it is assumed known and fixed. We shall concisely write the probability $\Pr(Y_1 = y_1, Y_2 = y_2)$ as $p(y_1, y_2)$.

A. Quantization rules

The outputs in both the input- and output-broadcast systems are described by the following mappings.

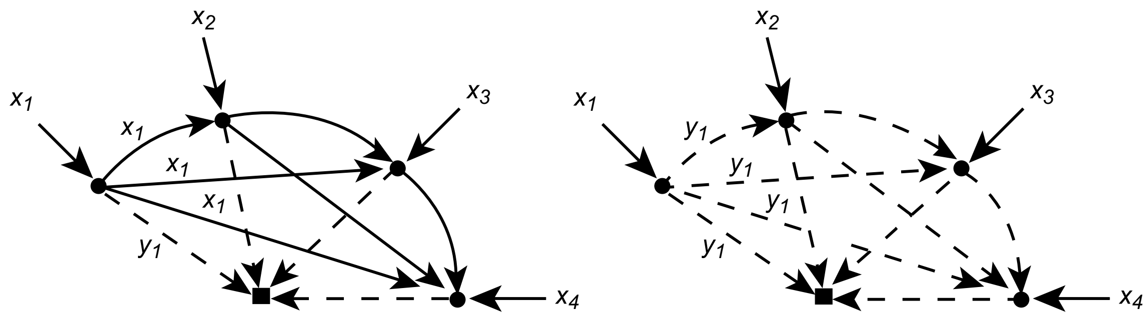
$$\left.\begin{aligned} y_1 &= \phi_1(x_1): \ \mathbb{R} \to \{0,1\}, \\ y_2 &= \phi_2(x_1, x_2): \ \mathbb{R}^2 \to \{0,1\} \end{aligned}\right\} \text{input-broadcast} \quad (1)$$

$$\left.\begin{aligned} y_1 &= \psi_1(x_1): \ \mathbb{R} \to \{0,1\}, \\ y_2 &= \psi_2(x_2, y_1): \ \mathbb{R} \times \{0,1\} \to \{0,1\} \end{aligned}\right\} \text{output-broadcast} \quad (2)$$

In the input-broadcast system, y_2 is a function of two real-valued variables, whereas in the output-broadcast system, y_2 is a function of one real-valued and one binary-valued variable.

Manuscript received March 4, 2006; revised September 27, 2006; revised again April 20, 2007. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Xiaodong Wang. This work was supported in part by the MIT Lincoln Laboratory Graduate Scholar Fellowship Program.

M. A. Lexa and D. H. Johnson are with the Department of Electrical and Computer Engineering, Rice University, Houston, TX 77251-1892 USA (email: amlexa@rice.edu; dhj@rice.edu).



(a) Input-broadcast (b) Output-broadcast

Fig. 1. Broadcast structures. In the input broadcast system, the real-valued observations x_m (solid lines) are broadcast to the other quantizers while the quantized outputs y_m (dashed lines) are transmitted to the detector. In the output-broadcast system, quantized outputs are broadcast to all succeeding quantizers and to the detector. The squares represent the detectors.

An alternative description of these quantization rules can be given in terms of binary partitions. For $A \subset \mathbb{R}^n$, we define a binary partition π to be a pair of disjoint sets $\{A, \bar{A}\}$ where $A \cup \bar{A} = \mathbb{R}^n$, $\bar{A}$ denoting the complement set. Let $\pi_1 = \{A_1, \bar{A}_1\}$ be a binary partition of $\mathbb{R}$ and $\pi_2 = \{A_2, \bar{A}_2\}$ a partition of $\mathbb{R}^2$. Then, for the input-broadcast system, the quantization rules are defined by

$$y_1 = \phi_1(x_1) = \begin{cases} 0 \text{ if } x_1 \in \bar{A}_1 \\ 1 \text{ if } x_1 \in A_1 \end{cases} \tag{3}$$

$$y_2 = \phi_2(x_1, x_2) = \begin{cases} 0 \text{ if } (x_1, x_2) \in \bar{A}_2 \\ 1 \text{ if } (x_1, x_2) \in A_2. \end{cases} \tag{4}$$

Defining the output-broadcast system in terms of binary partitions is more involved. Because y_1 can assume one of two values, the second quantizer has the freedom to use one partition when $y_1 = 0$ and an entirely different partition when $y_1 = 1$. Thus, the quantization rule ψ_2 consists of two binary partitions (of $\mathbb{R}$), not one. Letting $\nu_1 = \{B_1, \bar{B}_1\}$, $\nu_{2,0} = \{B_{2,0}, \bar{B}_{2,0}\}$, and $\nu_{2,1} = \{B_{2,1}, \bar{B}_{2,1}\}$ be binary partitions of $\mathbb{R}$, we have

$$y_1 = \psi_1(x_1) = \begin{cases} 0 \text{ if } x_1 \in \bar{B}_1 \\ 1 \text{ if } x_1 \in B_1 \end{cases} \tag{5}$$

$$y_2 = \psi_2(x_2, y_1 = i) = \begin{cases} 0 \text{ if } x_2 \in \bar{B}_{2,i} \\ 1 \text{ if } x_2 \in B_{2,i} \end{cases}, \ i = 0, 1. \tag{6}$$

The above functional relationships constitute to what we refer as the *structural constraints* of the input- and output-broadcast system. These constraints stipulate (restrict) the quantities on which the quantizers operate, and thus determine the links among the quantizers. Here, the only structural difference between the input- and output-broadcast systems is the communication rate between the quantizers. For the input-broadcast system the rate is infinite and for the output-broadcast system the rate is one bit/observation.

Broadcast structures have received little attention in the quantization for detection or distributed detection literature. While various authors have touched on various aspects of distributed quantization for *noncommunicative* structures [8, 9], none to our knowledge have considered broadcast structures. The vast majority of the distributed detection formulations also center on noncommunicative (parallel) structures (see [10] and references therein), however, there are two notable exceptions. First, Hashemi and Rhodes [11] investigated a two-stage distributed detection system that is essentially equivalent to the input-broadcast structure. Their system disallowed inter-quantizer communications, but processed temporal data in much the same way an input-broadcast system processes spatial data. Thus by interchanging the spatial and temporal domains, the structures' mathematical descriptions are nearly the same. However, unlike here, Hashemi and Rhodes used a Bayes criterion instead of the divergence and assumed that all observations (in both space and time) are statistically independent. Moreover, they optimized their system using a distinctly different approach than the one proposed here (see Section III for details). Second, serial architectures [10] are very similar to the output-broadcast structure; however, these systems take the output of the last quantizer as the decision of the system, whereas an output-broadcast structure bases its decision on *all* outputs. (This difference is most appreciable for systems with more than two quantizers.)

B. Kullback-Leibler divergence

The KL divergence is a member of the class distance measures which quantify the "dissimilarity" between probability distributions. For two probability density functions $p_{\mathbf{X}}^{(0)}$ and $p_{\mathbf{X}}^{(1)}$, the KL divergence between $p_{\mathbf{X}}^{(1)}$ relative to $p_{\mathbf{X}}^{(0)}$ is defined as the expected value of the negative log-likelihood ratio with respect to $p_{\mathbf{X}}^{(0)}$,

$$\mathcal{D}(p_{\mathbf{X}}^{(0)} \| p_{\mathbf{X}}^{(1)}) := \mathcal{E}_0[-\log(L)] = \int p_{\mathbf{X}}^{(0)}(\mathbf{x}) \log \frac{p_{\mathbf{X}}^{(0)}(\mathbf{x})}{p_{\mathbf{X}}^{(1)}(\mathbf{x})} \, d\mathbf{x}, \tag{7}$$

where L denotes the likelihood ratio, $p_{\mathbf{X}}^{(1)}(\mathbf{x})/p_{\mathbf{X}}^{(0)}(\mathbf{x})$, and the choice of the logarithm's base is arbitrary. To ensure the existence of the integral, we assume that the two probability measures associated with $\mathbf{X}$ are absolutely continuous with respect to each other. When $p_{\mathbf{X}}^{(j)}, j = 0, 1$, are probability mass functions, the integral in (7) becomes a summation. Here, the relevance of the KL divergence stems from Stein's Lemma [7, p. 77], a well-known result that relates the divergence to the asymptotic performance of a detector. In words, Stein's Lemma says that an optimal Neyman-Pearson detector's error probability decays exponentially in the number of observations, with the asymptotic exponential decay rate equal to the divergence between the distributions characterizing the detector's inputs.

C. Problem setting

We assume the inputs $\mathbf{X}$ are distributed in one of two ways: $H_0 : \ \mathbf{X} \sim p_{\mathbf{X}}^{(0)}$, $H_1 : \ \mathbf{X} \sim p_{\mathbf{X}}^{(1)}$ and address the problem of maximizing the output divergence $\mathcal{D}(p_{\mathbf{Y}}^{(0)} \| p_{\mathbf{Y}}^{(1)})$ over the set of all quantization rules ϕ_1, ϕ_2 and ψ_2, ψ_2 for the input- and output-broadcast structures, respectively. We assume the input divergence is finite and that the joint distribution of $p_{\mathbf{X}}$ is completely known under both hypotheses. We are not directly concerned with the operation of the detector, but in the context of our problem, the divergence between the distributed systems' output distributions represent the detector's best asymptotic error decay rate. Therefore, $\mathcal{D}(p_{\mathbf{Y}}^{(0)} \| p_{\mathbf{Y}}^{(1)})$

governs how well the detector can assimilate quantizer outputs and optimally process them. In addition, our use of the KL divergence as the quantity to be optimized implies that the detector uses the quantizers' results in a particular way. Rather than making a decision each time the quantizers produce an output vector, the detector operates on a long sequence of quantizer outputs $\mathbf{Y}_l,\ l = 1, 2, \ldots$, before making a decision. Here, $\mathbf{Y}_l$ is the quantized output corresponding to the input $\mathbf{X}_l$ for $l = 1, 2, \ldots$. This scenario corresponds to situations when quantizer processing occurs frequently, and the detector can afford to assimilate several quantizer outputs before making its own. Note this mode of decision making is different than that of the standard distributed detection problem where each sensor's output is often considered a local decision on the hypothesis. We assume we have an *iid* sequence of input vectors $\mathbf{X}_l$ that allows for statistical dependencies (spatial dependencies) between $X_{1,l}$ and $X_{2,l}$ for a given l. Because we require the structures operate on one input vector at a time, the temporal index l is irrelevant and will be suppressed.

III. Structure optimizations

Because we can always factor the joint output distributions,

$$p^{(j)}(\mathbf{y}) = p^{(j)}(y_1)p^{(j)}(y_2|y_1),\ j = 0, 1 \tag{8}$$

we can write the output divergence as a sum of two component divergences

$$\mathcal{D}(p^{(0)}_{\mathbf{Y}} \| p^{(1)}_{\mathbf{Y}}) = \mathcal{D}(p^{(0)}_{Y_1} \| p^{(1)}_{Y_1}) + \mathcal{D}(p^{(0)}_{Y_2|Y_1} \| p^{(1)}_{Y_2|Y_1}). \tag{9}$$

The form of (9) suggests a natural, albeit suboptimal, sequential approach to maximize the joint output divergence; first maximize $\mathcal{D}(p^{(0)}_{Y_1} \| p^{(1)}_{Y_1})$ over ψ_1 (ϕ_1), then given the resulting quantization rule, maximize $\mathcal{D}(p^{(0)}_{Y_2|Y_1} \| p^{(1)}_{Y_2|Y_1})$ over ψ_2 (ϕ_2). This strategy treats the maximization of the joint divergence as a separable optimization over the quantization rules when in fact the divergence terms on the right-hand side of (9) are generally coupled. It is for this reason that this approach is suboptimal. It does, however, have the advantage of taking into account some of the statistical dependencies among the outputs.

For broadcast structures with small numbers of quantizers, we can increase the complexity of the above approach, improve its performance, but still maintain tractability. Rather than sequentially maximizing the divergences on the right-hand side of (9) with the goal of determining the optimizing partitions, we can sequentially maximize them with the intent of only *parameterizing* a class of partitions. Then, if such parameterizations exist, we can jointly optimize the output divergence $\mathcal{D}(p^{(0)}_{\mathbf{Y}} \| p^{(1)}_{\mathbf{Y}})$ over the set of all valid parameters. Thus, this alternative approach first attempts to define a parametric class of partitions, and then searches for the best partition (quantization rule) within that class. By jointly optimizing over the parameters, we partially mitigate the effect of initially treating the maximization of the joint divergence as a separable optimization.

Note that this approach generally does not yield person-by-person optimal (PBPO) solutions[1] which are often sought in distributed detection problems [10, 11]. Using a PBPO approach, one would determine ψ_1 (ϕ_1) by maximizing the *joint* divergence $\mathcal{D}(p^{(0)}_{\mathbf{Y}} \| p^{(1)}_{\mathbf{Y}})$ while holding ψ_2 (ϕ_2) fixed, and likewise determine ψ_2 (ϕ_2) while holding ψ_1 (ϕ_1) fixed. Thus, a PBPO approach does not parameterize the observation space through sequentially maximizing the marginal and conditional divergences in (9). A PBPO approach is not used here because it suffers from essentially the same intractability as finding the global solution [12].

[1] A set of quantization rules is person-by-person optimal if the overall system performance cannot be improved by adjusting a single quantization rule while all other rules are held fixed. Person-by-person optimality is a necessary (but not a sufficient) condition for global optimality.

A. Output-broadcast

We begin with the first quantizer in isolation and ask what type of binary partition of the outcome space of X_1 maximizes $\mathcal{D}(p^{(0)}_{Y_1} \| p^{(1)}_{Y_1})$. Tsitsiklis showed in [13, Proposition 4.1] that a likelihood ratio partition, that is a partition defined by thresholding the likelihood ratio, is the maximizing partition,

$$\nu_1 : B_1 = \{x_1 \in \mathbb{R} | L(x_1) > \tau_1\}, \quad L(x_1) = \frac{p^{(1)}(x_1)}{p^{(0)}(x_1)}. \tag{10}$$

This means that ν_1 is simply parameterized by the threshold τ_1.

We next consider maximizing $\mathcal{D}(p^{(0)}_{Y_2|Y_1} \| p^{(1)}_{Y_2|Y_1})$ to derive a parameterization the second quantizer's quantization rule while holding ψ_1 fixed. By definition, [7], $\mathcal{D}(p^{(0)}_{Y_2|Y_1} \| p^{(1)}_{Y_2|Y_1})$ equals,

$$\begin{aligned} \sum_{y_1} p^{(0)}(y_1) \left[\sum_{y_2} p^{(0)}(y_2|y_1) \log \frac{p^{(0)}(y_2|y_1)}{p^{(1)}(y_2|y_1)} \right] = \\ p^{(0)}(Y_1 = 0)\, \mathcal{D}(p^{(0)}_{Y_2|Y_1=0} \| p^{(1)}_{Y_2|Y_1=0}) \\ + p^{(0)}(Y_1 = 1)\, \mathcal{D}(p^{(1)}_{Y_2|Y_1=1} \| p^{(1)}_{Y_2|Y_1=1}) \end{aligned} \tag{11}$$

We see from (11) that the conditional divergence is an average of two component divergences between the Bernoulli distributions $p^{(j)}(y_2|y_1 = i),\ i, j = 0, 1$. Note that $\mathcal{D}(p^{(0)}_{Y_2|Y_1=0} \| p^{(1)}_{Y_2|Y_1=0})$ is only conditioned on the event $\{Y_1 = 0\}$ and $\mathcal{D}(p^{(1)}_{Y_2|Y_1=1} \| p^{(1)}_{Y_2|Y_1=1})$ only on $\{Y_1 = 1\}$. Therefore, $\mathcal{D}(p^{(0)}_{Y_2|Y_1=0} \| p^{(1)}_{Y_2|Y_1=0})$ is only associated with the partition that is used when $y_1 = 0$, and $\mathcal{D}(p^{(1)}_{Y_2|Y_1=1} \| p^{(1)}_{Y_2|Y_1=1})$ is only associated with the partition that is used when $y_1 = 1$. We can thus apply Tsitsiklis' likelihood ratio partitioning result to each component divergence separately and conclude that $\nu_{2,0}$ and $\nu_{2,1}$ can each be parameterized by a single threshold, with the partition sets

$$\nu_{2,i} : B_{2,i} = \{x_2 \in \mathbb{R} | L(x_2|y_1 = i) > \tau_{2,i}\}, \tag{12}$$

$$L(x_2|y_1 = i) = \frac{p^{(1)}(x_2|y_1 = i)}{p^{(0)}(x_2|y_1 = i)},\ i = 0, 1. \tag{13}$$

Summarizing, our sequential optimization approach transforms the maximization of the joint output divergence over a general set of quantization rules ψ_1, ψ_2 into a maximization of over just three threshold parameters $\tau_1,\ \tau_{2,0},\ \tau_{2,1}$. When the likelihood ratios are all monotonic, each partition can be described by a single threshold on an *input observation*, as opposed to a threshold on the likelihood ratio.

B. Input-broadcast

For the input-broadcast structure, the parameterization of ϕ_1 is exactly the same as that for the output-broadcast structure. Therefore, we know that π_1 is a likelihood ratio partition parameterized by a single threshold η_1,

$$\pi_1 : A_1 = \{x_1 \in \mathbb{R} | L(x_1) > \eta_1\}, \quad L(x_1) = \frac{p^{(1)}(x_1)}{p^{(0)}(x_1)}. \tag{14}$$

For the maximization of $\mathcal{D}(p^{(0)}_{Y_2|Y_1} \| p^{(1)}_{Y_2|Y_1})$ over ϕ_2, we begin by considering $p^{(j)}(y_2{=}0|y_1{=}0)$ expressed as

$$\int_{\mathbb{R}^2} p^{(j)}(y_2 = 0|y_1 = 0, \mathbf{x})\, p^{(j)}(\mathbf{x}|y_1 = 0)\ d\mathbf{x}, \tag{15}$$

where $\mathbf{x} = (x_1, x_2)$. The first term in the integrand is the probability of the event $\{Y_2{=}0\}$ conditioned on the joint event $\{Y_1{=}0, X_1{=}x_1, X_2{=}x_2\}$. The second term in the integrand of (15) is the joint density of (X_1, X_2) conditioned on $\{Y_1 = 0\}$. From (3) and the properties of conditional densities, we have

$$p^{(j)}(\mathbf{x}|y_1 = 0) = p^{(j)}\ \ \mathbf{x}|(x_1, x_2) \in \bar{A}_1 \times \mathbb{R}$$
$$= \begin{cases} \dfrac{p^{(j)}(\mathbf{x})}{\Pr\ \ \bar{A}_1 \times \mathbb{R}; H_j}, & \text{if } (x_1, x_2) \in \bar{A}_1 \times \mathbb{R} \\ 0 \text{ otherwise.} \end{cases}$$

Thus, $p^{(j)}(\mathbf{x}|y_1 = 0)$ only has support on $\bar{A}_1 \times \mathbb{R}$. For example, if $\bar{A}_1$ is the semi-infinite interval $(-\infty, 1]$, then $p^{(j)}(\mathbf{x}|y_1 = 0)$ only has support on the semi-infinite plane bounded by the vertical line $x_1 = 1$. The integral in (15) therefore reduces to

$$\int_{\bar{E}_{2,y_1=0}} p^{(j)}\ \ \mathbf{x}|(x_1, x_2) \in \bar{A}_1 \times \mathbb{R}\quad d\mathbf{x}, \tag{16}$$

where $\bar{E}_{2,y_1=0} = (\bar{A}_1 \times \mathbb{R}) \cap \bar{A}_2$ and $\bar{E}_{2,y_1=0} \subset \bar{A}_2$. We can similarly write

$$p^{(j)}(y_2{=}0|y_1{=}1) = \int_{\bar{E}_{2,y_1=1}} p^{(j)}\ \ \mathbf{x}|(x_1, x_2) \in A_1 \times \mathbb{R}\quad d\mathbf{x}, \tag{17}$$

where $\bar{E}_{2,y_1=1} = (A_1 \times \mathbb{R}) \cap \bar{A}_2$ and $\bar{E}_{2,y_1=1} \subset \bar{A}_2$. Recalling S_2's quantization rule for the input structure (4), expressions (16) and (17) imply that $\bar{A}_2 = \bar{E}_{2,y_1=0} \cup \bar{E}_{2,y_1=1}$. This means that the partition π_2 is composed of four disjoint sets whose pairwise unions equal A_2 and $\bar{A}_2$: $\pi_2 = \{E_{2,y_1=0} \cup E_{2,y_1=1}, \bar{E}_{2,y_1=0} \cup \bar{E}_{2,y_1=1}\}$. Moreover, because A_1 and $\bar{A}_1$ are disjoint, $\bar{E}_{2,y_1=0}$ and $\bar{E}_{2,y_1=1}$ are also disjoint. Thus, the two component divergences appearing in the expression of the conditional divergence (11) can be independently optimized in much the same way as they were in the output-broadcast system. We can maximize $\mathcal{D}(p^{(0)}_{Y_2|Y_1=0} \| p^{(1)}_{Y_2|Y_1=0})$ over $E_{2,y_1=0}$ and maximize $\mathcal{D}(p^{(0)}_{Y_2|Y_1=1} \| p^{(1)}_{Y_2|Y_1=1})$ over $E_{2,y_1=1}$. Again, by applying Tsitsiklis' likelihood ratio partitioning result, we conclude that π_2 can be parameterized by two threshold parameters $\eta_{2,0}$ and $\eta_{2,1}$, i.e.

$$\pi_2 : A_2 = E_{2,y_1=0} \cup E_{2,y_1=1}, \text{ where} \tag{18}$$
$$E_{2,y_1=i} = \{\mathbf{x} \in \bar{A}_1 \times \mathbb{R} | L(\mathbf{x}|y_1 = i) > \eta_{2,i}\}, \tag{19}$$
$$L(\mathbf{x}|y_1 = i) = \frac{p^{(1)}(\mathbf{x}|y_1 = i)}{p^{(0)}(\mathbf{x}|y_1 = i)},\ i = 0, 1. \tag{20}$$

Therefore, like the output-broadcast structure, we transformed the maximization of $\mathcal{D}(p^{(0)}_{\mathbf{Y}} \| p^{(1)}_{\mathbf{Y}})$ over ϕ_1, ϕ_2 into a maximization over three threshold parameters: η_1, $\eta_{2,0}$, $\eta_{2,1}$. In contrast to the output-broadcast structure, the partitioning sets of the second quantizer $(A_2, \bar{A}_2)$ are subsets of $\mathbb{R}^2$ not $\mathbb{R}^1$. The extra dimension adds a degree of freedom to the partitioning sets (leading to improved performance), but generally makes it more difficult to relate the likelihood ratio thresholds directly to the observations.

Note that the partitions π_1 and π_2 in (14) and (18) look very much like the PBPO partition Hashemi and Rhodes derived in [11] for their two-stage distributed detection system. But, as pointed out in [14], their results are incorrect for a PBPO approach. We re-emphasize that our approach is not PBPO.

IV. Gaussian Example

Consider the distributed hypothesis test, $H_0 : \mathbf{X} \sim \mathcal{N}(\mathbf{0}, \Sigma)$, $H_1 : \mathbf{X} \sim \mathcal{N}(\mathbf{m}, \Sigma)$, where under both hypotheses the inputs X_1, X_2 each have unit variance and are correlated with correlation coefficient ρ.

A. Output-broadcast

For this example, it is well-known that the likelihood ratios defining the partitioning sets in (10) and (12) are monotonic. We can therefore re-parameterize ψ_1 and ψ_2 using three thresholds to which the observations can be directly compared; denoted here by $\xi_1, \xi_{2,0}$, and $\xi_{2,1}$. Setting $\rho = 0.9$, we jointly optimize $\mathcal{D}(p^{(0)}_{\mathbf{Y}} \| p^{(1)}_{\mathbf{Y}})$ over $\xi_1, \xi_{2,0}$, and $\xi_{2,1}$ using standard numerical methods. Fig. 2 depicts the resulting overall partition.

B. Input-broadcast

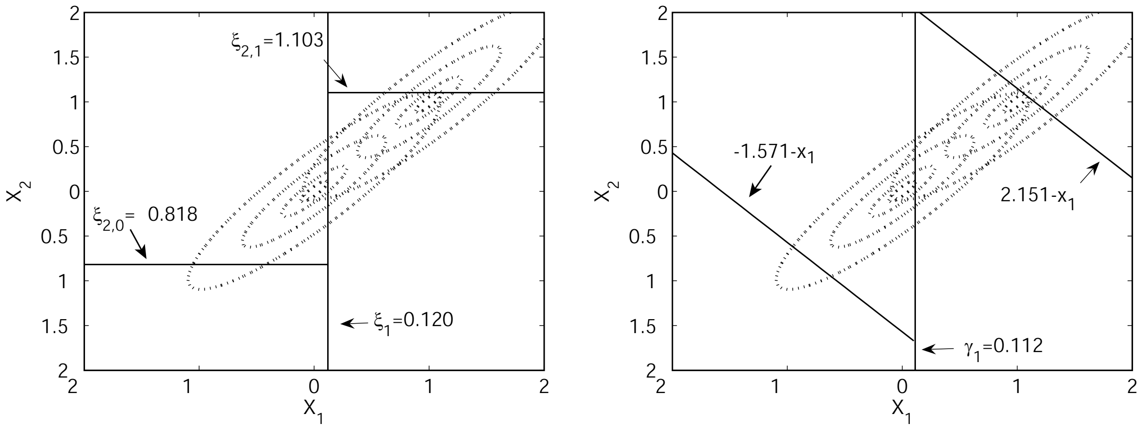
As with the output-broadcast structure, the likelihood ratio that characterizes the first quantizer's quantization rule is monotonic. Therefore, the threshold parameter η_1 can be replaced by another threshold parameter γ_1 which can be directly compared to the observation X_1. To relate the second quantizer's threshold parameters to thresholds on the observations, we consider the conditional likelihood ratios appearing in (20). For $y_1 = 0$, we have

$$p^{(j)}(\mathbf{x}|y_1 = 0) = \frac{I_{\bar{A}_1 \times \mathbb{R}}}{2\pi c_j |\Sigma|^{1/2}} \exp\ \ -\frac{1}{2}\boldsymbol{\delta}'\Sigma^{-1}\boldsymbol{\delta} \tag{21}$$

where I is the indicator function, $c_j = \Pr(\bar{A}_1 \times \mathbb{R}; H_j)$, $\Sigma = \begin{matrix} 1 & \rho \\ \rho & 1 \end{matrix}$, $\boldsymbol{\delta} = \mathbf{x}$ under H_0, and $\boldsymbol{\delta} = \mathbf{x} - \mathbf{m}$ under H_1. Forming the likelihood ratio and simplifying the inequality in (19) yields $E_{2,y_1=0} = \ \ \mathbf{x} \in \bar{A}_1 \times \mathbb{R} |\ x_2 > b_0 - x_1$, where $b_0 = 1 + (1 + \rho) \ln (c_0\eta_{2,0}/c_1)$. Thus, $E_{2,y_1=0}$ is defined by a line in the left half plane (left of the partitioning line $x_1 = \gamma_1$) with slope equal to -1 and y-intercept equal to b_0. Because the slope is constant, we are able to parameterize $E_{2,y_1=0}$ by b_0. The companion partitioning set $E_{2,y_1=1}$ similarly can be characterized by a y-intercept. In particular, we have $E_{2,y_1=1} = \ \ \mathbf{x} \in A_1 \times \mathbb{R} |\ x_2 > b_1 - x_1$, where $b_1 = 1 + (1 + \rho) \ln (c'_0\eta_{2,1}/c'_1)$ and $c'_j = \Pr(A_1 \times \mathbb{R}; H_j)$. Hence, the likelihood ratio threshold parameterization of the input-broadcast quantization rules is equivalent to a parameterization with one threshold parameter γ_1 and two y-intercept parameters b_0, b_1. With $\mathbf{m} = (1, 1)$ and $\rho = 0.9$, we jointly solve for the maximizing parameters (again using standard numerical techniques) and show the resulting overall partition in Fig. 2.

C. Discussion

Most visible in Fig. 2 is the impact of the infinite rate on the link between the first and second quantizers in the input-broadcast system. The structural freedom afforded by the infinite rate permits the partitioning boundaries of ϕ_2 to have nonzero slopes, whereas the partitioning boundaries of ψ_2 are constrained to have zero slope. The misalignment of the boundaries across $\phi_1's$ $(\psi_1's)$ vertical cut is, however, a direct by-product of the sequential optimization and not caused by structural constraints. It reflects the fact that the second quantizer is optimized with respect to the conditional divergence $\mathcal{D}(p^{(0)}_{Y_2|Y_1} \| p^{(1)}_{Y_2|Y_1})$ instead of, say, the marginal divergence $\mathcal{D}(p^{(0)}_{Y_2} \| p^{(1)}_{Y_2})$ which, incidentally, would force the alignment of the boundaries. Moreover, note that the *number* of vertical boundaries in Fig. 2 is also a by-product of the sequential optimization. Maximizing $\mathcal{D}(p^{(0)}_{Y_1} \| p_{Y_1})$ to determine ϕ_1 and ψ_1 results in a likelihood ratio thresholding rule, and because the likelihood ratio is monotonic in this case, there is only a *single* vertical boundary. The structural constraints only insist that the boundaries be orthogonal to the axes; they do not restrict the number of boundaries. If one jointly optimizes the quantizers and eliminates the constraints imposed by the sequential optimization, very different partitions emerge; in particular, partitions that do *not* derive from thresholding likelihood ratios [2]. In addition, the ordering of the sequential optimization has a definite effect on the partition. If ϕ_2 (ψ_2) were optimized first, the vertical boundaries in Fig. 2 would become horizontal and ϕ_1 (ψ_1) would be composed of two component quantization rules, much like ϕ_2 (ψ_2) is described in Section II. In this example, the ordering makes no difference because of the Gaussian distribution's symmetry, but



(a) Output-broadcast (b) Input-broadcast

Fig. 2. Optimized partitions of the observation space with Gaussian inputs when $\rho = 0.9$ and $\mathbf{m} = (1, 1)$. The elliptical curves are contours of equiprobability for the underlining input distributions. Because of an output-broadcast structure's structural constraints, each partitioning line must be orthogonal to the axes. In comparison, the input-broadcast system's structural constraints are weaker and therefore the partitioning lines of the second quantizer can have a nonzero slope. Both partitions suffer from the effects of the sequential optimization.

in general, different orderings yield different performances. How one should order the quantizers in any given situation is largely an open problem [15].

The left panel of Fig. 3 shows the output divergence of the input- and output-broadcast systems as a function of ρ, along with two upper and one lower bounds. It follows from the divergence's invariance property [7, pp. 18-22] that the input divergence $\mathcal{D}(p_{\mathbf{X}}^{(0)} \| p_{\mathbf{X}}^{(1)})$ serves as an upper bound to any transformation (quantization) of the inputs, and here represents the optimal asymptotic error decay rate of a detector that has direct access to both observations X_1 and X_2. The divergence of an optimal centralized likelihood ratio quantizer (LRQ) [13] also serves as an upper bound because such a system is free from any structural or optimization constraints, the presence of which would only decrease the divergence. The optimal LRQ's divergence, unlike the input divergence, takes into account the inherent loss in divergence due to quantization and represents the decay rate of a detector that operates on the output of an optimal centralized quantizer. Thus, any performance gap between a LRQ's benchmark and that of the input- or output broadcast systems is only caused by structural constraints and suboptimal optimization. The lower bound shown in Fig. 2 represents a worst-case scenario in terms of both structural constraints and optimization technique. Specifically, we consider a noncommunicative system in which there is no inter-quantizer communications (most stringent structure) and an optimization technique that only locally maximizes the divergence. That is, each quantization rule is found by optimizing the divergence between the marginal output distributions, $\mathcal{D}(p_{Y_m}^{(0)} \| p_{Y_m}^{(1)}), m = 1, 2$. Note this optimization approach completely ignores statistical dependencies and hence the bound is constant for all values of ρ.

To better judge performance, we normalize the divergences by the optimal centralized LRQ's divergence and plot the resulting curves in the right hand panel of Fig. 3. Because we normalize by the optimal *centralized* quantizer, the graph shows the percentage loss in terms of divergence due *only* to structural constraints and those caused by sequentially optimizing. On the one hand, these plots indicate that the performance gain is greatest when the correlation between the observations is negative. That is, the divergence would increase the most if one were to increase the communication rate between the quantizers (or equivalently change a distributed system's structure) when the correlation is negative. On the other hand, the percentage loss is *least* when the observations are positively correlated. This means that for, say $\rho = 0.9$, it is nearly as easy to discriminate between the output distributions $p_{\mathbf{Y}}^{(j)}$ when only one bit is communicated than when a real-valued observation is communicated. As one would expect, the input-broadcast system never performs worse than the output-broadcast system because the latter is a special case of the former (assuming that the input-broadcast system's second quantizer has knowledge of ϕ_1). Also of interest, is the performance at $\rho = 1$ because this case represents the situation where the quantizers have a common input. The presence of a common input causes the input- and output-broadcast structures to coincide (again assuming that the second quantizer has knowledge of the first quantizer's quantization rule), and therefore they achieve the same performance at this point. Furthermore, because of the monotonicity of the likelihood ratios, both structures achieve the same performance as an optimal centralized LRQ at $\rho = 1$.

V. Concluding Remarks

Designers of distributed vector quantization systems will, more often than not, face choices and make tradeoffs concerning a system's structure and the methods with which to optimize it. It is evident from the above example that the degree of impact can vary greatly depending on the circumstances. Thus, understanding how both structure and optimization techniques constrain the overall problem in any given situation is fundamental to making prudent and effective design choices. The following summarizes our observations.

- With finite rate communication links, structural constraints *only* imply that the partition boundaries are orthogonal to the axes.
- Sequential optimization leads to likelihood ratio thresholding quantization rules (parameterizations) that reduce complexity, but automatically fix the number and locations of the partitioning boundaries. Note that the globally optimal quantization rules are not, in general, likelihood ratio thresholding rules [2].
- Sequential optimization creates a sensitivity to the order in which the component quantizers are optimized.
- A common input dissolves the structural difference between the input- and output-broadcast systems.

One shortfall of our sequential optimization scheme is that it becomes computationally expensive when jointly optimizing a large number of partitioning *parameters* because the number of parameters over which to maximize grows exponentially with the number of quantizers. In these cases, a more feasible approach would be to

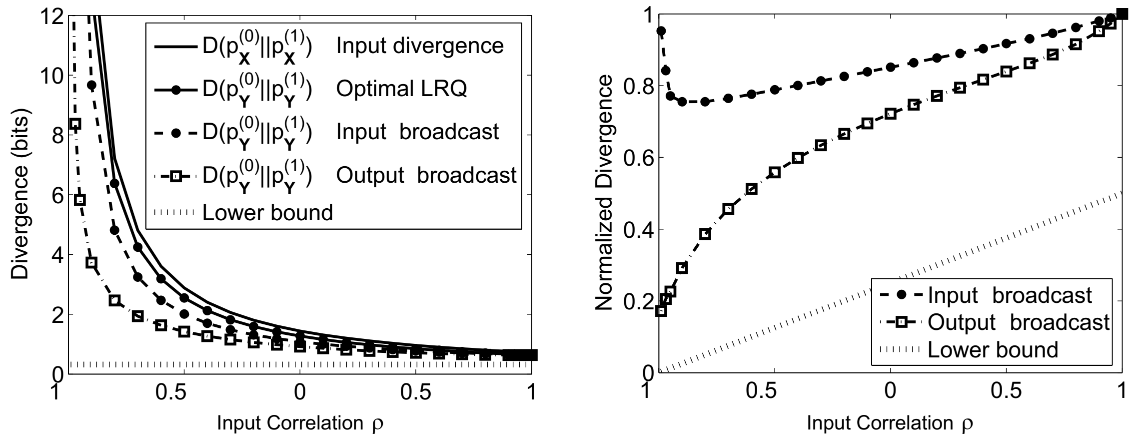


Fig. 3. The panel on the left shows the output divergence of the input- and output-broadcast structures as a function of the correlation coefficient. The performance of these structures are upper bounded by the input divergence and the divergence of an optimal centralized likelihood ratio quantizer (LRQ). The lower bound represents the performance of a noncommunicative structure whose component quantizers are optimized individually. The panel on the right shows the performance curves normalized by the LRQ's performance.

optimize a broadcast structure as was first suggested in Section III: sequentially maximize the divergences on the right-hand side of (9) to determine each quantizer's quantization rule instead of sequentially maximizing them to find a parametric class of partitions. This optimization boils down to a sequence of maximizations that each only involve a single parameter.

Acknowledgment

The authors would like to thank the reviewers for their helpful and constructive comments.

References

[1] A. Gersho and R. M. Gray, *Vector Quantization and Signal Compression*. Kluwer Academic Publishers, 1992.

[2] M. A. Lexa and D. H. Johnson, "Joint optimization of distributed quantization system for classification," in *Proc. IEEE Data Compression Conf.*, J. Storer and M. Marcellin, Eds. IEEE Computer Society, Mar 2007, pp. 363–372.

[3] J. Tsitsiklis and M. Athans, "On the complexity of decentralized decision making and detection problems," *IEEE Trans. Auto. Control*, vol. AC-30, no. 5, pp. 440–446, May 1985.

[4] H. V. Poor and J. B. Thomas, "Applications of Ali-Silvey distance measures in the design of generalized quantizers for binary decision systems," *IEEE Trans. on Communications*, vol. 25, no. 9, pp. 893–900, Sep 1977.

[5] S. Kassam, *Signal Detection in Non-Gaussian Noise*. Springer-Verlag, 1988.

[6] R. Balasubramanian, C. A. Bouman, and J. P. Allebach, "Sequential scalar quantization of vectors: An analysis," *IEEE Trans. Image Processing*, vol. 4, no. 9, pp. 1282–1295, Sep 1995.

[7] S. Kullback, *Information Theory and Statistics*. New York: Wiley, 1959.

[8] T. Flynn and R. M. Gray, "Encoding of correlated observations," *IEEE Trans. Info. Th.*, vol. 33, no. 6, pp. 773–787, Nov 1987.

[9] M. Longo, T. D. Lookabaugh, and R. M. Gray, "Quantization for decentralized hypothesis testing under communication constraints," *IEEE Trans. Info. Th.*, vol. 36, no. 2, pp. 241–255, Mar 1990.

[10] R. Viswanathan and P. Varshney, "Distributed detection with multiple sensors: Part I-Fundamentals," *Proc. of the IEEE*, vol. 85, no. 1, pp. 54–63, Jan 1997.

[11] H. Hashemi and I. Rhodes, "Decentralized sequential detection," *IEEE Trans. Info. Th.*, vol. 35, no. 3, pp. 509–520, May 1989.

[12] V. Veeravalli, T. Basar, and H. Poor, "Decentralized sequential detection with a fusion center performing the sequential test," *IEEE Trans. Info. Th.*, vol. 39, no. 2, pp. 433–442, Mar 1993.

[13] J. Tsitsiklis, "Extremal properties of likelihood-ratio quantizers," *IEEE Trans. Comm.*, vol. 41, no. 4, pp. 550–558, Apr 1993.

[14] V. Veeravalli, "Comments on "Decentralized Sequential Detection"," vol. 38, no. 4, pp. 1428–1429, Jul 1992.

[15] J. Papastavrou and M. Athans, "On optimal distributed decision architectures in a hypothesis testing environment," *IEEE Trans. on Automatic Control*, vol. 8, no. 37, pp. 1154–1169, Aug 1992.