Broadcast Detection Structures with Applications to Sensor Networks

Michael A. Lexa[†]* and Don H. Johnson[‡]
Department of Electrical and Computer Engineering
Rice University, Houston, TX 77251-1892
amlexa@rice.edu, dhj@rice.edu
(713) 348-2256[†], (713) 348-4956[‡]
(713) 348-6196 (fax)

Abstract

Data broadcasting is potentially an effective and efficient way to share information in wireless sensor networks. Broadcasts offer energy savings over multiple, directed transmissions, and they provide a vehicle to exploit the statistical dependencies often present in distributed data. In this paper, we examine two broadcast structures in the context of a distributed detection problem whose inputs are statistically dependent. Specifically, we develop a suboptimal approach to maximize the Kullback-Leibler divergence over a set of binary quantization rules. Our approach not only leads to simple parameterizations of the quantization rules in terms of likelihood ratio thresholds, but also provides insight into the inherent constraints distributed structures impose. We then present two examples in detail and compare the performance of the broadcast structures to that of a centralized system and a noncooperative system. These examples suggest that in situations where the detection problem is difficult (small input divergence), broadcasting solitary bits (or even nothing at all) may be nearly as effective as broadcasting real-valued observations.

EDICS: SEN-COLB

Broadcast Detection Structures with Applications to Sensor Networks

I. Introduction

It is often the case in wireless sensor networks that sensed data are statistically dependent in both time and space [1]. This fact, coupled with the challenge of processing data "in-network", promotes inter-sensor communications as a way to exploit these statistical dependencies. In particular, broadcasting is an enticing capability because information can be shared among several sensors for roughly the same energy cost as that of a single directed transmission. In this paper, we analyze two broadcasting structures in the context of a binary-hypothesis distributed detection problem. We optimize the structures so as to maximize the Kullback-Leibler (KL) divergence between the distributions characterizing the distributed system's outputs, or equivalently, a final decision maker's inputs (Fig. 1). Here, we assume the sensor observations are statistically dependent in space but independent and identically distributed (*iid*) in time. Because the divergence is well-known to be the optimal asymptotic exponential error rate for Neyman-Pearson type tests [2, p. 77], optimizing it at the output of the distributed system maximizes the potential asymptotic error decay rate of the final decision maker. Thus, a key concept in this approach is to view the distributed structure as a single system acting collectively, instead of as M individual sensors. Stated in a slightly different way, we think of the structure as a single preprocessor to the final decision maker (Fig. 1). From this perspective then this work is most like, and can be viewed as an extension of, problems in quantization for detection [3, 4] [5, Chap. 4], where quantizers are designed such that downstream detectors are optimized.

We consider two extreme cases of broadcast structures. In the *input-broadcast* structure (Fig. 2), each sensor broadcasts its real-valued observation to all succeeding sensors; that is, after the m^{th} sensor collects its observation, it broadcasts its data to all sensors $k, k > m$. In contrast, in the *output-broadcast* system each sensor processes its observation and broadcasts its result, a solidary bit, to all succeeding sensors. Thus, any sensor in the input-broadcast structure has access to all preceding observations, whereas a sensor in the output-broadcast structure processes one real-valued observation and $m-1$ binary-valued outputs. In both cases, a final decision maker receives the binary outputs from all the sensors and ultimately decides which hypothesis is true.

We recognize that an input-broadcast structure centralizes all of the observations at the M^{th} sensor,

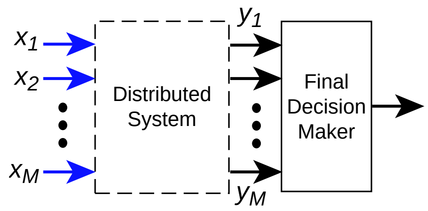
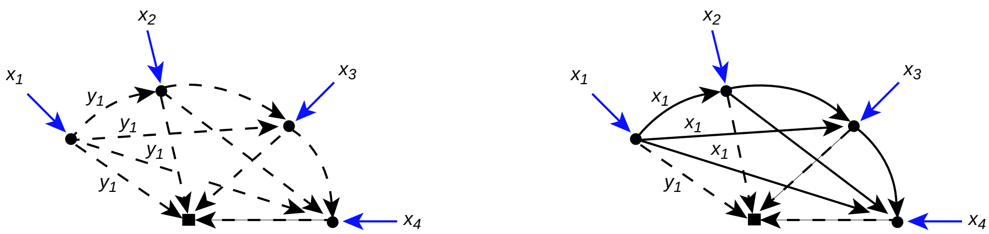


Fig. 1. A distributed detection system viewed as a preprocessor to a final decision maker. Note that we illustrate the distributed system as a single box to emphasize its collective functionality.



(a) Output-broadcast

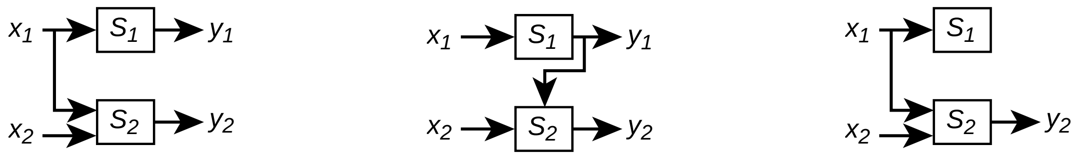
(b) Input-broadcast

Fig. 2. Local cooperative broadcast structures with four sensors. In the output-broadcast system, binary outputs (dashed lines) are broadcast to all succeeding sensors and to the final decision maker. In the input broadcast system, the real-valued observations x_m (solid lines) are broadcast to the other sensors while the binary outputs y_m are transmitted to the final decision maker. The squares represent the final decision makers. Note that in the output-broadcast structure, the final decision maker receives the outputs y_m as part of the inter-sensor broadcasts, but in the input-broadcast structure, each output must be individually transmitted.

and thus it could make an optimal centralized decision, rendering the final decision maker unnecessary. For our purposes here, however, we are not concerned with this scenario. We analyze the input-broadcast system primarily because it represents the extreme or ideal case where there is no rate (bandwidth) constraint on the broadcasts.

Broadcast structures have received little attention in the distributed detection literature and none to our knowledge have dealt with dependent observations. The vast majority of the problem formulations assume parallel or serial architectures (see e.g. [6, 7]), and while other structures have been proposed and studied [8], nearly all results assume statistically independent observations (under each hypothesis). We note that Swaszek and Willett [9] examined a broadcast detection structure with the novel criterion of having the sensors reach consensus, and while their results do not directly pertain to our analysis, their work serves as added context to our problem.

In 1985, Tsitsiklis and Athans [10] showed that determining the optimal decision rules for a parallel



(a) Input-broadcast (b) Output-broadcast (c) Modified input-broadcast

Fig. 3. Block diagrams of broadcast structures when $M = 2$. A performance comparison of 3(c) to 3(a) highlights the point that the outputs should not be considered as preliminary decisions to hypothesis test, but as elements to an optimized joint output distribution.

structure with dependent inputs is NP-complete. Largely because of this result, much of the early work in distributed detection assumes that the input observations are statistically independent (under each hypothesis). More recently, though, researchers have begun to tackle problems with dependent observations. Blum and his coauthors [11, 12] have made substantial progress in some weak signal cases, and Zhu *et al.* [13] proposed an iterative algorithm that approximates optimal decision rules for the parallel structure with a fixed fusion rule. However, the fact still remains that analysis has proven extremely difficult and surprisingly complex even for the "simplest" dependence problems (two sensor, parallel structure, Gaussian inputs, and binary outputs) [14]. For this reason, we restrict attention here to broadcast structures with only two sensors (Fig. 3). Fortunately, the restriction to two sensors is not as problematic as it might seem because our analysis can be easily extended, at least theoretically, to larger broadcast structures.

II. Problem formulation

Let $\mathbf{X} = (X_1, X_2)$ denote the input data and $\mathbf{Y} = (Y_1, Y_2)$ the sensors' outputs. Assume $\mathbf{X}$ is a continuous real-valued random vector, and $\mathbf{Y}$ a discrete binary-valued random vector: $\mathbf{X} \in \mathbb{R}^2, \mathbf{Y} \in \{0,1\}^2$. 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 a marginal density of $p_{\mathbf{X}}$, and p_{Y_m} a marginal probability mass function of $p_{\mathbf{Y}}$. In our problem, the realization $X_m = x_m$ represents the input to the m^{th} sensor and $Y_m = y_m$ represents the corresponding output (Fig. 3). We shall concisely write the probability $\Pr(Y_1 = y_1, Y_2 = y_2)$ as $p(y_1, y_2)$, and we shall denote the m^{th} sensor by S_m.

A. Quantization rules

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

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

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

Note that 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. 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 $\{A, \overline{A}\}$ where $A \cup \overline{A} = \mathbb{R}^n$, $\overline{A}$ denoting the complement set. Let $\pi_1 = \{A_1, \overline{A}_1\}$ be a binary partition of $\mathbb{R}$ and $\pi_2 = \{A_2, \overline{A}_2\}$ a partition of $\mathbb{R}^2$. Then, for the input-broadcast system, the quantization rules are defined by

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

$$y_2 = \phi_2(x_1,x_2) = \begin{cases} 0 \text{ if } (x_1,x_2) \in \overline{A}_2 \\ 1 \text{ if } (x_1,x_2) \in A_2. \end{cases} \tag{2}$$

Defining the output-broadcast system in terms of binary partitions is more involved. The discrete nature of y_1 allows S_2 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 $\nu_1 = \{B_1, \overline{B}_1\}$, $\nu_{2,0} = \{B_{2,0}, \overline{B}_{2,0}\}$, and $\nu_{2,1} = \{B_{2,1}, \overline{B}_{2,1}\}$ be binary partitions of $\mathbb{R}$, we have

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

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

B. KL divergence

For two probability density functions $p_{\mathbf{X}}^{(0)}$ and $p_{\mathbf{X}}^{(1)}$, the KL divergence between $p_{\mathbf{X}}^{(0)}$ relative to $p_{\mathbf{X}}^{(1)}$ is defined as the expected value of the negative log-likelihood-ratio with respect to $p_{\mathbf{X}}^{(0)}$,

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

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 becomes a summation.

The KL divergence is a member of the class of Ali-Silvey [15] distance measures which quantify the "dissimilarity" between probability distributions. The divergence is always non-negative and equals zero only when the distributions are the same (to within a set of measure zero). The relevance of the KL divergence in this problem largely stems from Stein's Lemma [2, 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 statement

In this paper, we assume the inputs $\mathbf{X}$ are distributed in one of two ways,

$$\begin{aligned} H_0 : &\quad \mathbf{X} \sim p_{\mathbf{X}}^{(0)} \\ H_1 : &\quad \mathbf{X} \sim p_{\mathbf{X}}^{(1)}. \end{aligned}$$

and address the problem of maximizing the output divergence $\mathcal{D}(p_{\mathbf{Y}}^{(0)} \| p_{\mathbf{Y}}^{(1)})$ (assuming the input divergence is finite) over the quantization rules ψ_1, ψ_2 and ϕ_2, ϕ_2 associated with the input- and output-broadcast structures. We are not directly concerned with the operation of the final decision maker, but in the context of our problem, the divergence between the distributed systems' output distributions represent the decision maker's best asymptotic error decay rate. Therefore, $\mathcal{D}(p_{\mathbf{Y}}^{(0)} \| p_{\mathbf{Y}}^{(1)})$ governs how well the decision maker can assimilate sensor outputs and optimally process them. Moreover, our use of the KL divergence as the quantity to be optimized implies that the decision maker uses the sensors' results in a particular way. Rather than making a decision each time the sensors produce an output vector, the decision maker operates on a long sequence of sensor outputs $\mathbf{Y}_l$, $l \in \mathbb{N}$, before making the final decision. This scenario corresponds to situations when sensor processing occurs frequently, and the decision maker can afford to assimilate several sensor outputs before making its own.

We assume we have an *iid* sequence of input vectors $\mathbf{X}_l$ but allow 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 dropped for the remainder of the paper.

With this approach, our primary intent is to develop an analytical method (tool) to optimize and study distributed broadcast structures in some insightful way.

III. Divergence versus local decisions

The majority of the distributed detection literature is heavily biased towards the perspective that each sensor produces a local decision, and that the final decision maker fuses these decisions and produces a final decision for every set of local decisions it receives (see e.g. [7, 16, 17]. Here, we instead study detection structures that do not produce local decisions, and final decision makers that operate on data collected over longer time periods. To illustrate how our results will differ from those based on analyzing local decisions, consider the modified input-broadcast structure shown in Fig. 3(c). In this case, S_1 simply forwards its observation to S_2 without sending an output to the final decision maker. This is in contrast to the structure in Fig. 3(a) where both sensors transmit their outputs to the final decision maker. We now ask: Which structure has the largest output divergence?

If one believes that maximizing $\mathcal{D}(p^{(0)}_{\mathbf{Y}} \| p^{(1)}_{\mathbf{Y}})$ yields an optimal local decision, then one would probably argue that incorporating y_1 would not provide any new information because in both cases, S_2 has access to x_1 *and* x_2. Hence, the two structures ought to be able to achieve the same performance in terms of divergence. In fact, continuing this line of argument, if S_2 is designed optimally, then the structure in Fig. 3(a) could never outperform the structure in Fig. 3(c) because the final decision maker needs only consider y_2 as knowing y_1 would not offer any new information.

Admittedly this argument sounds convincing; however, by an application of the log-sum inequality [2, p. 15-16], it can be shown that a structure which incorporates both y_1 and y_2 always outperforms (except in exceptional circumstances) a structure using only y_1. (We provide a proof of this result in [18].) For example, when the inputs are correlated Gaussian random variables with $\rho = 0.9$, and the sensors are each designed as minimum probability of error detectors, $\mathcal{D}(p^{(0)}_{Y_1} \| p^{(1)}_{Y_1}) = 0.3226$ but $\mathcal{D}(p^{(0)}_{Y_1,Y_2} \| p^{(1)}_{Y_1,Y_2}) = 0.3402$. We therefore conclude that when maximizing the output divergence the sensors' binary outputs do not represent local decisions: if they did, the structure in Fig. 3(a) could never outperform the one in

Fig. 3(c). Instead, the outputs are best viewed as random variables whose joint distributions (under the two hypotheses) are as dissimilar as possible under given constraints.

IV. STRUCTURE OPTIMIZATIONS

The problem of maximizing the joint output divergence over the sensors' quantization rules is that of optimally partitioning the observation (input) space into disjoint sets. In this section, we derive a suboptimal partitioning method for the input- and output-broadcast structures depicted in Fig. 3. Finding the globally optimal partitioning is computationally intractible, even for a small number of sensors; in fact, finding the optimal quantization rules for a two-sensor, non-cooperative structure with discrete inputs is NP-complete [6, 10]. Consequently, we believe that the global optimization of a *cooperative* structure with *continuous* inputs can only be more burdensome. Therefore, a suboptimal approach is an attractive alternative in this setting.

Because we can always factor the joint output density as

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

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

$$\mathcal{D}(p_{\mathbf{Y}}^{(0)} \| p_{\mathbf{Y}}^{(1)}) = \mathcal{D}(p_{Y_1}^{(0)} \| p_{Y_1}^{(1)}) + \mathcal{D}(p_{Y_2|Y_1}^{(0)} \| p_{Y_2|Y_1}^{(1)}). \tag{3}$$

The form of (3) immediately suggests a natural way to maximize the joint output divergence over ψ_1, ψ_2 (or ϕ_1, ϕ_2); first maximize $\mathcal{D}(p_{Y_1}^{(0)} \| p_{Y_1}^{(1)})$ over ψ_1 (ϕ_1), then given the resulting quantization rule for S_1, maximize $\mathcal{D}(p_{Y_2|Y_1}^{(0)} \| p_{Y_2|Y_1}^{(1)})$ over ψ_2 (ϕ_2). This strategy is reminiscent of the person-by-person optimization technique sometimes employed in the distributed detection literature [7, 19], and it certainly constrains the problem sufficiently to avoid the complexity of jointly optimizing the partitioning sets (quantization rules). But we can do better by slightly relaxing the constraints of this strategy. The basic idea is to maximize the component divergences as described above, but instead of deriving specific partitions from each maximization, we parameterize them. Thus, we first derive a parametric class of partitions, and then jointly optimize the parameters to maximize the total output divergence $\mathcal{D}(p_{\mathbf{Y}}^{(0)} \| p_{\mathbf{Y}}^{(1)})$. This strategy is less restrictive than separately maximizing each term in (3), but more restrictive that the true joint optimization. We develop this strategy below for both the input- and output-broadcast structures.

A. *Output-broadcast*

We begin with the first sensor in isolation and ask what type of binary partition of the outcome space of X_1 maximizes $\mathcal{D}(p_{Y_1}^{(0)} \| p_{Y_1}^{(1)})$. Tsitsiklis showed in [20, Proposition 4.1] that a likelihood ratio partition, that is a partition defined by thresholding the likelihood ratio, is the maximizing partition,

$$\nu_1 : B_1 = \{x_1 \in \mathbb{R} | L(x_1) > \tau_1\}, \quad L(x_1) = \frac{p^{(1)}(x_1)}{p^{(0)}(x_1)}. \tag{4}$$

This means that ν_1 is simply parameterized by the threshold τ_1. Tsitsiklis proved this result in a broad study about centralized M-level quantizers, but it is applicable here because we consider the maximization of each of the component divergence terms in (3) separately, and therefore essentially treat the joint optimization of $\mathcal{D}(p_{\mathbf{Y}}^{(0)} \| p_{\mathbf{Y}}^{(1)})$ as multiple centralized optimizations. Independently of Tsitsiklis, Warren and Willett [4] proved a similar result in a distributed detection setting for a standard parallel architecture with independent inputs.

Next, we consider the maximization of $\mathcal{D}(p_{Y_2|Y_1}^{(0)} \| p_{Y_2|Y_1}^{(1)})$ in order to derive a parameterization the second sensor's quantization rule. Recall that ψ_2 is defined by two partitions $\nu_{2,0}$ and $\nu_{2,0}$, and thus we need to parameterize each in some manner.

Assume the first sensor's quantization rule is now fixed. By definition, $\mathcal{D}(p_{Y_2|Y_1}^{(0)} \| p_{Y_2|Y_1}^{(1)})$ equals,

$$\begin{aligned} \sum_{y_1} p^{(0)}(y_1) \left[\sum_{y_2} p^{(0)}(y_2|y_1) \log \frac{p^{(0)}(y_2|y_1)}{p^{(1)}(y_2|y_1)} \right] = \\ p^{(0)}(Y_1 = 0)\, \mathcal{D}(p_{Y_2|Y_1=0}^{(0)} \| p_{Y_2|Y_1=0}^{(1)}) \\ + p^{(0)}(Y_1 = 1)\, \mathcal{D}(p_{Y_2|Y_1=1}^{(1)} \| p_{Y_2|Y_1=1}^{(1)}) \end{aligned} \tag{5}$$

Because the probabilities $p^{(j)}(y_2|y_1 = i)$, $i, j = 0, 1$, define four Bernoulli distributions, we see from (5) that the conditional divergence is an average of two component Bernoulli divergences, $\mathcal{D}(p_{Y_2|Y_1=0}^{(0)} \| p_{Y_2|Y_1=0}^{(1)})$ and $\mathcal{D}(p_{Y_2|Y_1=1}^{(1)} \| p_{Y_2|Y_1=1}^{(1)})$. Note that the first Bernoulli divergence only involves conditioning on the event $\{Y_1 = 0\}$ and the second only on the event $\{Y_1 = 1\}$. Therefore, $\mathcal{D}(p_{Y_2|Y_1=0}^{(0)} \| p_{Y_2|Y_1=0}^{(1)})$ is only associated with the partition that is used when $y_1 = 0$, and $\mathcal{D}(p_{Y_2|Y_1=1}^{(1)} \| p_{Y_2|Y_1=1}^{(1)})$ is only associated with the partition that is used when $y_1 = 1$. This means that we can apply the above 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 defined by,

$$\nu_{2,i} : B_{2,i} = \{x_2 \in \mathbb{R} | L(x_2|y_1 = i) > \tau_{2,i}\}, \tag{6}$$

$$L(x_2|y_1 = i) = \frac{p^{(1)}(x_2|y_1 = i)}{p^{(0)}(x_2|y_1 = i)}, \ i = 0, 1. \tag{7}$$

In summary, our suboptimal approach transforms the maximization of the joint output divergence over a general set of quantization rules for the output-broadcast structure into a maximization of over just three threshold parameters, which collectively, define a parametric class of partitions over the entire observation

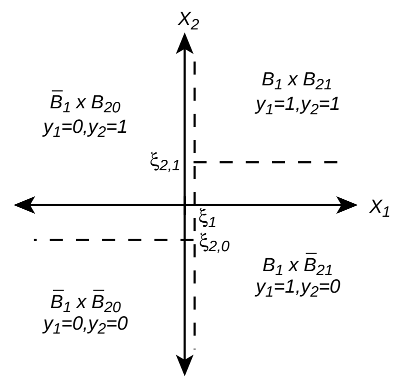


Fig. 4. Parameterized partition of the outcome space of $\mathbf{X}$ for the output-broadcast structure (with two sensors). The parameter ξ_1 defines the quantization rule for S_1, and $\xi_{2,0}$ and $\xi_{2,1}$ define the quantization rule for S_2. Each region is a Cartesian product of the partitioning sets associated with the quantization rules. Here, we assume all likelihood ratios are monotonic.

space.

$$\max_{\psi_1,\psi_2} \mathcal{D}(p^{(0)}_{\mathbf{Y}} \| p^{(1)}_{\mathbf{Y}}) \rightarrow \max_{\tau_1,\,\tau_{2,0},\,\tau_{2,1}} \mathcal{D}(p^{(0)}_{\mathbf{Y}} \| p^{(1)}_{\mathbf{Y}})$$

In the special case where the likelihood ratios are all monotonic, each partition is described by a single threshold on the appropriate input observation, and the overall partition of the observation space resembles that shown in Fig. 4. Therefore, our approach simplifies the harder, joint optimization problem and allows easier comparison with other structures.

B. Input-broadcast

For the input-broadcast structure, the parameterization of S_1's quantization rule for the input structure is exactly the same as that for the output structure. Therefore, we know that π_1 is a likelihood ratio partition parameterized by a single threshold η_1,

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

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

$$\int_{\mathbb{R}^2} p^{(j)}(y_2 = 0 | y_1 = 0, \mathbf{x})\, p^{(j)}(\mathbf{x} | y_1 = 0)\ d\mathbf{x}, \tag{8}$$

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 $\{Y_1 = 0, X_1 = x_1, X_2 = x_2\}$. Because y_2 is completely determined by x_1 and x_2 in the input-broadcast structure, $p^{(j)}(y_2 = 0|y_1 = 0, \mathbf{x})$ equals one when $(x_1, x_2) \in \overline{E}_{2,y_1=0}$, where $\overline{E}_{2,y_1=0}$ is some subset of $\overline{A}_2$, and equals zero otherwise.

The second term in the integrand of (8) is the joint density of (X_1, X_2) conditioned on $\{Y_1 = 0\}$. From (1) and the properties of conditional densities, we have

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

Thus, $p^{(j)}(\mathbf{x}|y_1 = 0)$ only has support on $\overline{A}_1 \times \mathbb{R}$. For example, if $\overline{A}_1$ 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 $x_1 = 1$. Therefore, we conclude that the integral in (8) reduces to

$$\int_{\overline{E}_{2,y_1=0}} p^{(j)}\big(\mathbf{x}|(x_1, x_2) \in \overline{A}_1 \times \mathbb{R}\big)\ d\mathbf{x}. \tag{9}$$

Note that because the integrand in (9) only has support on $\overline{A}_1 \times \mathbb{R}$, we must have that $\overline{E}_{2,y_1=0}$ is a subset of $\overline{A}_1 \times \mathbb{R}$ (to within a set of measure zero).

We can similarly write $p^{(j)}(y_2 = 0|y_1 = 1)$ as

$$\int_{\overline{E}_{2,y_1=1}} p^{(j)}\big(\mathbf{x}|(x_1, x_2) \in A_1 \times \mathbb{R}\big)\ d\mathbf{x}, \tag{10}$$

where $\overline{E}_{2,y_1=1} \subset \overline{A}_2$ and $\overline{E}_{2,y_1=1} \subset A_1 \times \mathbb{R}$.

Recalling S_2's quantization rule for the input structure (2), expressions (9) and (10) imply that $\overline{A}_2 = \overline{E}_{2,y_1=0} \cup \overline{E}_{2,y_1=1}$. This means that the partition π_2 is composed of four disjoint sets whose pairwise unions equal A_2 and $\overline{A}_2$,

$$
\begin{aligned}
\pi_2 &= \{A_2, \overline{A}_2\} \\
&= \{E_{2,y_1=0} \cup E_{2,y_1=1}, \overline{E}_{2,y_1=0} \cup \overline{E}_{2,y_1=1}\}.
\end{aligned}
$$

Moreover, because A_1 and $\overline{A}_1$ are disjoint, $\overline{E}_{2,y_1=0}$ and $\overline{E}_{2,y_1=1}$ are also disjoint. Thus, the two Bernoulli divergences appearing in the expression of the conditional divergence (5) can be independently optimized in much the same way as they were in the output-broadcast system. We can maximize $\mathcal{D}(p^{(0)}_{Y_2|\mathbf{Y}_1=0} \| p^{(1)}_{Y_2|\mathbf{Y}_1=0})$ over $E_{2,y_1=0}$ and maximize $\mathcal{D}(p^{(0)}_{Y_2|\mathbf{Y}_1=1} \| p^{(1)}_{Y_2|\mathbf{Y}_1=1})$ 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

$\eta_{2,1}$. That is, we have that $\pi_2 : A_2 = E_{2,y_1=0} \cup E_{2,y_1=1}$ where

$$E_{2,y_1=i} = \{\mathbf{x} \in \overline{A}_1 \times \mathbb{R} | L(\mathbf{x}|y_1 = i) > \eta_{2,i}\}, \tag{11}$$

$$L(\mathbf{x}|y_1 = i) = \frac{p^{(1)}(\mathbf{x}|y_1 = i)}{p^{(0)}(\mathbf{x}|y_1 = i)},\ i = 0, 1. \tag{12}$$

Therefore, like the output-broadcast structure, we transformed the maximization of $\mathcal{D}(p_{\mathbf{Y}}^{(0)} \| p_{\mathbf{Y}}^{(1)})$ 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 sensor ($A_2, \overline{A}_2$) are subsets of $\mathbb{R}^2$ not $\mathbb{R}^1$. The extra dimension adds a degree of freedom to the partitioning sets but generally makes it more difficult to relate the likelihood ratio thresholds directly to the observations.

We conclude this section by emphasizing that the overall partition of the observation space, induced by the quantization rules, for both the input- and output-broadcast structures are constrained by two factors:

Structural constraints. The distributed nature of broadcast systems restrict the dimension of the space that is quantized by each sensor. In the output-broadcast structure, each sensor only sees one component of the observation vector $\mathbf{X}$. Thus, all quantization rules are restricted to be partitions of only $\mathbb{R}$; ψ_1 must partition of the outcome space of X_1, and ψ_2 must partition of the outcome space of X_2. Consequently, for the output-broadcast system all partitioning lines in the space of $\mathbf{X}$ are perpendicular to the axes. The transmission of y_1 to the second sensor does, however, add a degree of freedom in the second sensor's quantization rule. In Fig. 4, this extra degree of freedom manifests itself as a split threshold in the horizontal direction.

In the input-broadcast structure, only S_1 faces structural constraints. As in the output-broadcast structure, its quantization rule can only partition the real line. On the other hand, the second sensor has full freedom to partition $\mathbb{R}^2$. The transmission of x_1 to S_2 centralizes the data and eliminates any structural constraints.

Constraints imposed by optimization technique. The sequential nature of our optimization technique leads to suboptimal parameterizations of the observation space simply because the component divergences are coupled and we do not jointly maximize them. Thus, the performance gap between an optimal centralize quantizer [1]and the input- and output-broadcast systems is caused by the constraints imposed by the distributed structure of the systems and the suboptimal method with which we optimized them.

[1]Here we mean a 4-level vector quantizer optimized to maximize the output KL divergence.

V. GAUSSIAN EXAMPLE

Consider the distributed hypothesis test,

$$H_0: \quad \mathbf{X} \sim \mathcal{N}(\mathbf{0}, \Sigma)$$

$$H_1: \quad \mathbf{X} \sim \mathcal{N}(\mathbf{m}, \Sigma)$$

where the inputs X_1, X_2 each have unit variance and are correlated with a correlation coefficient ρ. These hypotheses could model the situation of testing between two random signal vectors or determining the presence or absence of a deterministic signal in additive noise.

A. Output-broadcast

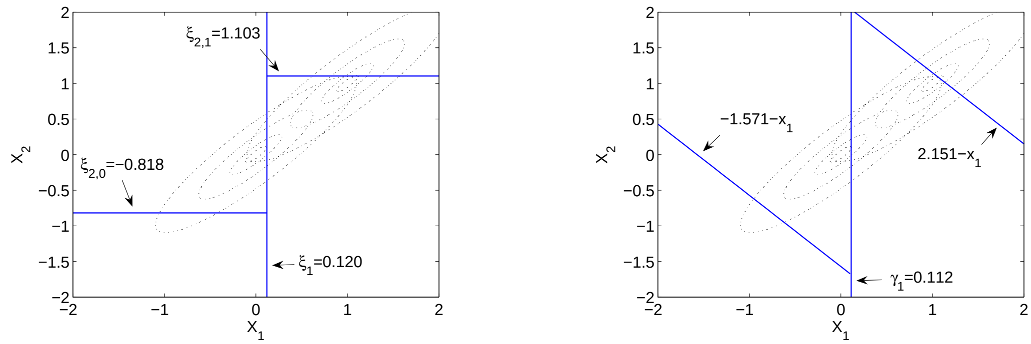
For this example, it is easily verified that the likelihood ratios defining the partitioning sets (see (4) and (6)) are monotonic. Therefore, we can re-parameterize ψ_1 and ψ_2 by three thresholds to which the observations can be directly compared; denoted here by ξ_1, $\xi_{2,0}$, and $\xi_{2,1}$. Thus, instead of jointly maximizing $\mathcal{D}(p_{\mathbf{Y}}^{(0)} \| p_{\mathbf{Y}}^{(1)})$ over $\tau_1, \tau_{2,0}, \tau_{2,1}$, we jointly maximize over $\xi_1, \xi_{2,0}, \xi_{2,1}$. This means that the conditional probabilities found within the expression of the output divergence are simply computed by integrating over rectangular regions (Fig. 4).

Setting $\rho = 0.9$, we solve this problem using standard numerical methods. We depict the overall partition induced by the quantization rules in Fig. 5(a). Again, note that all of the partitioning lines are perpendicular to the axes because of the structural constraints imposed by the output-broadcast system. The joint output divergence equals 0.4421 which is 84% of the total discrimination information (divergence) initially present in the input observations. Surprisingly, this means that when $\rho = 0.9$ it is as nearly as easy to discriminate between the multivariate *binary* distributions $p_{\mathbf{Y}}^{(j)}$ as it is to discriminate between the two *real-valued* distributions $p_{\mathbf{X}}^{(j)}$. This is not always the case, but this example makes the point that heavily quantizing an observation does not automatically imply an enormous loss in detection performance.

When X_1 and X_2 are independent, $\xi_{2,0}$ and $\xi_{2,1}$ are equal, and the overall partition of the observation space becomes a simple rectangular partition. This fact implies that in this case the transmission of y_1 to S_2 is not necessary and cannot increase the divergence because S_2 uses the same partition to produce its output regardless of y_1's value.

B. Input-broadcast

As in the output-broadcast structure, the likelihood ratio that characterizes the first sensor's quantization rule is monotonic. Therefore, the threshold parameter η_1 can be replaced by another threshold parameter



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

Fig. 5. Optimized partitions of the observation space with Gaussian inputs when $\rho = 0.9$ and $\mathbf{m} = \mathbf{1}$. In each plot, the elliptical curves are contours of equiprobability for the underlining input distributions. For the output-broadcast structure, each threshold must be orthogonal to the axes; whereas, for the input structure, the partitioning lines of S_2 have slope negative one. Despite the fact that, in this example, the output structure can only approximate the partition of the input structure, we see that the resulting divergences only differ by about 2.8%.

γ_1 which can be directly compared to the observation X_1.

To relate the second sensor's threshold parameters to thresholds on the observations, we consider the conditional likelihood ratios appearing in (12). For $y_1 = 0$, we have

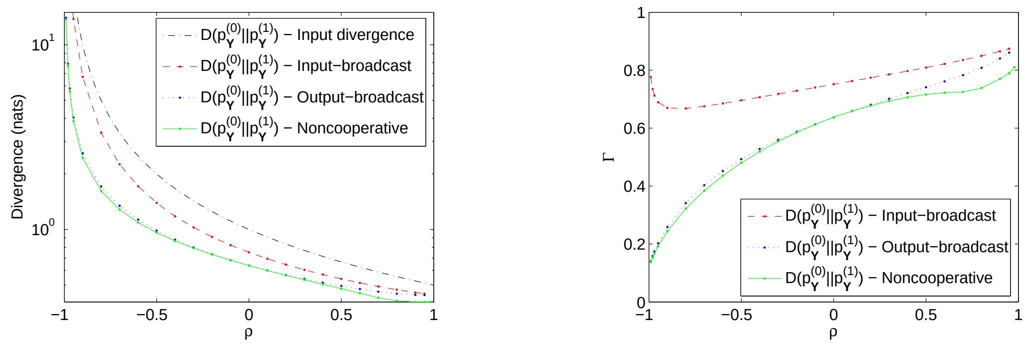
$$p^{(j)}(\mathbf{x}|y_1 = 0) = \frac{I_{\overline{A}_1 \times \mathbb{R}}}{2\pi c_j |\Sigma|^{-1/2}} \exp\left(-\frac{1}{2}\boldsymbol{\delta}'\Sigma^{-1}\boldsymbol{\delta}\right)$$

where I is the indicator function, $c_j = \Pr(\overline{A}_1 \times \mathbb{R}; H_j)$, $\Sigma = \left(\begin{smallmatrix} 1 & \rho \\ \rho & 1 \end{smallmatrix}\right)$, $\boldsymbol{\delta} = \mathbf{x}$ under H_0, and $\boldsymbol{\delta} = \mathbf{x} - \mathbf{m}$ under H_1. Forming the likelihood ratio and simplifying the inequality in (11) yields the following partitioning set:

$$E_{2,y_1=0} = \left\{\mathbf{x} \in \overline{A}_1 \times \mathbb{R} \,|\, x_2 > b_0 - x_1\right\},$$

where $b_0 = \ln(c_0\eta_{2,0}/c_1) + (1+\rho)^{-1}$. 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 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 $E_{2,y_1=1} = \left\{\mathbf{x} \in A_1 \times \mathbb{R} \,|\, x_2 > b_1 - x_1\right\}$, where $b_1 = \ln(c_0'\eta_{2,1}/c_1') + (1+\rho)^{-1}$ 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 $\mathbf{m} = \mathbf{1}$ and $\rho = 0.9$, we jointly solve for the maximizing parameters (again using standard numerically techniques) and show the resulting overall partition in Fig. 5(b). The total output divergence



(a) (b)

Fig. 6. Gaussian example. Panel (a) shows the output divergence of the input- and output-broadcast structures as a function of the correlation coefficient on a semi-log plot and compares the results to the input divergence and to the performance of a noncooperative system. Panel (b) shows the ratio of the input and output divergences, $\Gamma = \mathcal{D}(p_{\mathbf{Y}}^{(0)} \| p_{\mathbf{Y}}^{(1)}) / \mathcal{D}(p_{\mathbf{X}}^{(0)} \| p_{\mathbf{X}}^{(1)})$ for the various structures. Because of the invariance property of the divergence, Γ is a normalized quantity [2, p. 18].

equals 0.4554, which equates to 86% of the input divergence.

C. Discussion

Fig. 5 clearly illustrates the impact system structure has on the partitioning of the observation space. Having both input observations available at S_2 allows the partitioning lines of ϕ_2 to be two-dimensional compared to the one-dimensional partitioning lines in the output-broadcast structure. Because of structural constraints, the output-broadcast system simply cannot mimic the input-broadcast structure's partitions, and thus an optimized output-broadcast structure can never outperform an optimized input-broadcast structure. In fact, it is easy to see that the input-broadcast structure can theoretically perform no worse than the output-broadcast structure because the latter is a special case of the former.

That said, the actual difference in performance is problem dependent. For the specific case $\rho = 0.9$, the performance difference is minimal. Relative to the input divergence, the input-broadcast structure preserves 86% of the discrimination information compared to 84% for the output structure; a difference of only 2%. However, for different input distributions (say, for different values of ρ), the situation can be very different. From Fig. 6, we see that the difference between the output divergences becomes more pronounced for negative correlation values. For example, for $\rho = -0.9$, the input structure preserves 77% of the input divergence, while the output structure only preserves 14%; a difference of 63%!

Fig. 6 also compares the input- and output-broadcast structures to a noncooperative system in which there is no inter-sensor communications and whose performance is optimized according to our suboptimal approach. A noncooperative system's performance represents a lower bound on the performance of the broadcast systems because it is a more severely constrained system. Indeed, the quantization rules for the noncooperative system are completely characterized by simple thresholds (one per sensor), resulting in a simple quadrature partition.

From Fig. 6, it is clear that the performance of the output-broadcast and the noncooperative structures are nearly equal for negative and moderately positive correlation coefficients. For highly positively correlated observations, the gap between the noncooperative and output-broadcast systems widens. However, for $\rho = 1$ it is easily seen that the input-broadcast and the noncooperative systems degenerate into the same structure, and therefore their optimal performance should theoretically be identical. Furthermore, because the output system's performance is pinned between the input-broadcast and noncooperative systems, all three structures should have the same optimal performance at $\rho = 1$. In this example, the output-broadcast system's performance does indeed approach that of the input-broadcast structure as $\rho \to 1$, but the noncooperative system's performance does not. The reason being, at least qualitatively, is that our maximization approach allows the optimal input- and output-broadcast partitions to coincide at $\rho = 1$, but constrains the noncooperative partition from doing the same.

VI. Laplace Example

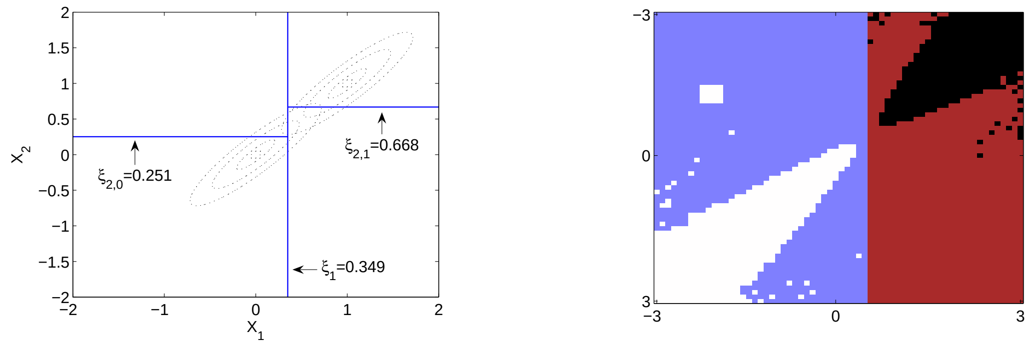
This example considers the same hypothesis test as above, but now the observations are characterized by a symmetric bivariate Laplace distribution $\mathcal{L}(\mathbf{m}, \Sigma)$ with density function [21]

$$p(\mathbf{x}) = \frac{1}{\pi\sqrt{|\Sigma|}}\; K_0\big(\sqrt{2(\mathbf{x}-\mathbf{m})'\Sigma^{-1}(\mathbf{x}-\mathbf{m})}\,\big),\;\; \mathbf{x} \neq \mathbf{0},$$

where $K_0(\cdot)$ is the modified Bessel function of the second kind [22, p. 374]. Because the bivariate Laplace density is elliptically contoured like the Gaussian density, one would expect the re-parameterization of the partitions to be similar to the Gaussian case, but this supposition is not true in general.

A. Output-broadcast

With Laplace observations, the likelihood ratio defining the partitioning set of S_1 (see (4)) is monotonic. Therefore, like the Gaussian case, we can re-parameterize the partition ν_1 using ξ_1. The conditional likelihood ratios defining $\nu_{2,0}$ and $\nu_{2,1}$ are not, however, monotonic for all values of ρ. (Counterexamples can easily be found.) Through experiments, however, we found that for most values of ρ the conditional likelihood ratios are either monotonic or are nearly monotonic. Thus, we adopt the same re-parameterization as before where the likelihood ratio thresholds $\tau_{2,0}, \tau_{2,1}$ are replaced by two other



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

Fig. 7. Optimized partitions of the observation space with Laplace inputs when $\rho = 0.9$ and $\mathbf{m} = \mathbf{1}$. The partition for the output-broadcast system is completely described by the three parameters $\xi_1, \xi_{2,0}$ and $\xi_{2,1}$ which represent thresholds on the observations. The dotted elliptical curves are contours of equiprobability for the underlining input distributions. The partition in 7(b) was determined through simulated annealing. The different shaded regions represent the partitioning sets $E_{2,y_1=i}, \overline{E}_{2,y_1=i}$, $i = 0, 1$, described in Section IV-B.

thresholds $\xi_{2,0}, \xi_{2,1}$ which are directly compared to the observations. When $\rho = 0.9$ and $\mathbf{m} = \mathbf{1}$, this scheme results in an output divergence equaling 0.6819. The optimized partition of the observation space is shown in Fig. 7(a).

B. Input-broadcast

In this case, it is evident that the likelihood ratios formed from the conditional Laplace densities do not yield any analytic parameterizations of the observation space. Thus, we choose to use simulated annealing [23] to find approximate solutions to our optimization problems. In brief, simulated annealing is a well-known stochastic relaxation technique used to solve combinatorial problems. Its application to the present problem requires a tessellation of the real plane, and for this reason, it only produces approximate solutions. However, our implementation adheres to the suboptimal approach described in Section IV-B. Therefore, the results are approximate solutions to our proposed approach, not approximate solutions to the globally optimal answer. A detailed overview of the algorithm and a short description of our implementation can be found in [23] and [18], respectively.

Fig. 8 summarizes our results for all values of the correlation coefficient when $\mathbf{m} = \mathbf{1}$. With $\rho = 0.9$, the total output divergence equals 0.7151, which equates to 92% of the input divergence. The plots show the same general behavior as in the Gaussian example; however, in this example, the output-broadcast system's performance does not approach that of the input-broadcast system as $\rho \to 1$. As in the last

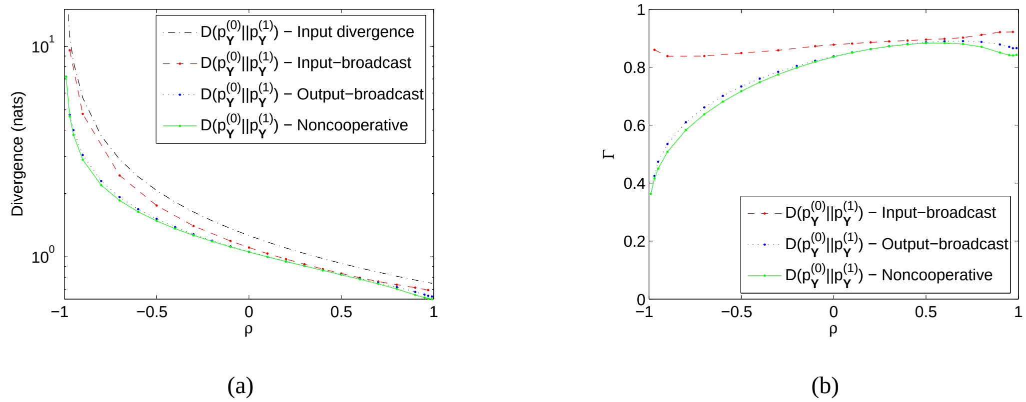


Fig. 8. Laplace example. Panel (a) shows the output divergence of the input- and output-broadcast structures in comparison to the input divergence and the performance of a noncooperative system as a function of the correlation coefficient. Panel (b) shows the ratio of the input and output divergences (Γ) for the different structures.

example, this behavior is caused by the partitioning constraints imposed by our maximization approach.

VII. Summary and Conclusion

Motivated by the potential benefits of broadcasting data in sensor networks, we analyzed two different broadcast structures with statistically dependent input observations. We developed a suboptimal approach to maximize the KL divergence over the sensors' quantization rules that results in simple parameterizations of the sensor's quantization rules in terms of likelihood ratio thresholds. This approach not only avoids the computational complexity (for small numbers of sensors - see the last paragraph) of finding the joint solution, but also illuminates the role structural constraints play in distributed detection systems.

We presented two examples in detail that demonstrate our approach and compare the performance of the input- and output-broadcast structures to the performance of a centralized system (equal to the input divergence) and a noncooperative system. The results show that in situations where the detection problem is difficult (small input divergence), broadcasting *bits or even nothing at all* may be nearly as effective as broadcasting *real-valued* observations. Such behavior may have far-reaching implications for sensor networks. For example, the output-broadcast structure may offer a significant energy savings (for each sensor) over the input-broadcast structure without sacrificing much in performance in these situations. From another viewpoint, the tradeoff between the energy expended in cooperative processing and the effectiveness of the resulting signal processing architecture is problem-dependent. In both examples, highly positively correlated data can be processed with little energy spent. However, when negatively correlated, much more energy must be expended to sustain a reasonable level of processing performance (witness the large performance gap between the input- and output-broadcast system for negatively correlated data).

One shortfall of our optimization method is that it becomes computationally expensive for large numbers of sensors, that is, the number of parameters over which to jointly maximize grows exponentially with M, the number of sensors. For our purposes here, scalability was not an issue since our focus was on a basic understanding of broadcast detection structures. However, in situations where larger numbers of sensors need to be considered, a more appropriate approach would be to optimize the broadcast structure as was first suggested in Section IV: successively maximize the conditional divergences to determine each sensor's quantization rule. The advantage of this alternate method is that the maximizations over the quantization rules all separate into maximizations over just one parameter, each of which is easily solved. The disadvantage, of course, is that the number of maximizations grows exponentially with M. In comparison to optimizing each sensor's marginal output divergence (for which the number of maximizations grows linearly with M), maximizing the conditional divergences has the advantage of taking into account some of the statistical dependencies between observations, whereas optimizing marginal divergences does not.

Finally, we believe extension of this work to larger broadcast structures will provide little new insight. More worthwhile directions perhaps include development methods that back away from the assumption that the input distributions are completely known and new theoretical analysis on the impact of structure on distributed signal processing tasks.

References

[1] S. Kumar, F. Zhao, and D. Shepherd, "Collaborative signal and information processing in microsensor networks," *IEEE Signal Processing Mag.*, vol. 19, no. 2, pp. 13–14, Mar 2002.

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

[3] H. Poor and J. 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.

[4] D. Warren and P. Willett, "Optimum quantization for detector fusion: some proofs, examples, and pathology," *Jour. Franklin Institute*, vol. 336, pp. 323–359, 1999.

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

[6] J. Tsitsiklis, *Advances in Statistical Signal Processing*. JAI Press, 1993, vol. 2, ch. Decentralized Detection, pp. 297–344.

[7] 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.

[8] Z. Tang, K. Pattipati, and D. Kleinman, "Optimization of detection networks: Part II-Tree structures," *IEEE Trans. on Systems, Man, and Cybernetics*, vol. 23, no. 1, pp. 211–221, Jan/Feb 1993.

[9] P. Swaszek and P. Willett, "Parley as an approach to distributed detection," *IEEE Trans. Aerospace and Electronic Systems*, vol. 31, no. 1, pp. 447–457, Jan 1995.

[10] 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.

[11] R. Blum and S. Kassam, "Optimum distributed detection of weak signals in dependent sensors," *IEEE Trans. Info. Th.*, vol. 38, no. 3, pp. 1066–1079, May 1992.

[12] R. Blum, "Necessary conditions for optimum distributed sensor detectors under the Neyman-Pearson criterion," *IEEE Trans. Info. Th.*, vol. 42, no. 3, pp. 990–994, May 1996.

[13] Y. Zhu, R. Blum, Z.-Q. Luo, and K. Wong, "Unexpected properties and optimum-distributed sensor detectors for dependent observation cases," *IEEE Trans. Automatic Control*, vol. 45, no. 1, pp. 62–72, Jan 2000.

[14] P. Willett, P. Swaszek, and R. Blum, "The good, bad, and ugly: Distributed detection of a known signal in dependent gaussian noise," *IEEE Trans. Signal Processing*, vol. 48, no. 12, pp. 3266–3279, Dec 2000.

[15] S. Ali and S. Silvey, "A general class of coefficients of divergence of one distribution from another," *J. of the Royal Statistical Society*, vol. 28, no. 1, pp. 131–142, 1966.

[16] I. Hoballah and P. "Distributed Bayesian signal detection," *IEEE Trans. Info. Th.*, vol. 35, no. 5, pp. 995–1000, Sep 1989.

[17] R. Viswanathan, S. Thomopoulos, and R. Tumuluri, "Optimal serial distributed decision fusion," *IEEE Trans. on Aerospace and Electronic Systems*, vol. 24, no. 4, pp. 366–376, Jul 1988.

[18] M. Lexa. (2006, Mar.) Broadcast detection structures with applications to sensor networks: Appendix. [Online]. Available: http://www.ece.rice.edu/~amlexa/broadcast_structures2006/broadcastappendix.pdf

[19] S. Alhakeem and P. Varshney, "A unified approach to the design of decentralized detection systems," *IEEE Trans. Aerospace and Electronic Systems*, vol. 31, no. 1, pp. 9–19, Jan 1995.

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

[21] T. Kozubowski and K. Podgórski, "Asymmetric Laplace laws and modeling financial data," *Math. and Comput. Modelling*, vol. 34, pp. 1003–1021, Nov 2001.

[22] M. Abramowitz and I. Stegun, Eds., *Handbook of Mathematical Functions*. U.S. National Bureau of Standards, 1964.

[23] P. van Laarhoven and E. Aarts, *Simulated Annealing: Theory and Applications*. D. Reidel Publishing Co., 1987.