DECENTRALIZED DETECTION

John N. Tsitsiklis

ABSTRACT

Consider a set of sensors that receive observations from the environment and transmit finite-valued messages to a fusion center that makes a final decision on one out of M alternative hypotheses. The problem is to provide rules according to which the sensors should decide what to transmit, in order to optimize a measure of organizational performance. We overview the available theory for the Bayesian formulation, and provide an alternative approach to the Neyman–Pearson variant of the problem. We also discuss (i) computational issues, (ii) asymptotic results, (iii) generalizations to more complex organizations, and (iv) sequential problems.

1. INTRODUCTION

1.1 Motivation

Consider the following problem. There are two hypotheses on the state of the environment and each one of a set of sensors receives some relevant information.

Advances in Statistical Signal Processing, Volume 2, pages 297–344.

ISBN: 0-89232-907-6

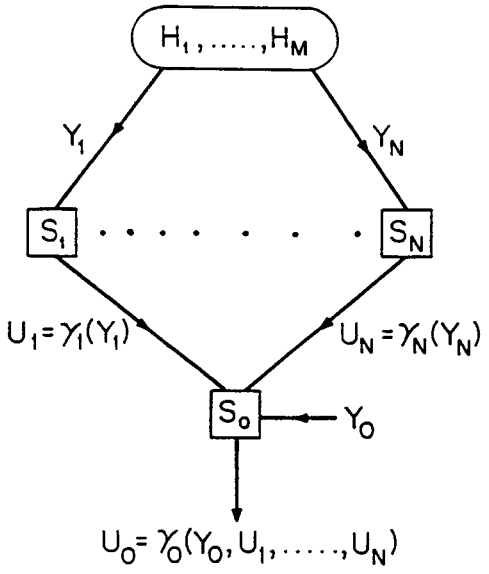

Figure 1.1.

In a *centralized* scheme, each sensor transmits all of its observations to a fusion center that solves a classical hypothesis testing problem and decides on one of the two hypotheses. A distinct alternative, introduced in [66], is a *decentralized* scheme whereby each sensor sends to the fusion center a summary of its own observations, in the form of a message taking values in a finite alphabet. Then, the fusion center makes a decision on the basis of the messages it has received (see Fig. 1.1). While the fusion center is again faced with a classical hypothesis testing problem (the messages being viewed as the fusion center's observations), a new and nontrivial problem arises: How should each sensor decide what message to send?

The above described centralized and decentralized schemes differ along several dimensions:

(a) *Performance:* In the decentralized scheme, the fusion center makes a decision on the basis of less information, which results in reduced performance.

(b) *Communication requirements:* The communication requirements of the decentralized scheme are typically much smaller than those of the centralized scheme. This is because each sensor only transmits a few bits, whereas the raw observations could be voluminous.

(c) *Off-line computation:* In the centralized scheme, the only off-line computation that is needed is the determination of a threshold to be used in a likelihood ratio test by the fusion center. The situation is much more complex with a

decentralized scheme because (as will be seen later) a set of thresholds must be computed for each sensor, and this computation is usually nontrivial.

(d) *On-line computation:* It will be seen in Section 2.3 that the online computational requirements of centralized and decentralized schemes are comparable.

Taking into account the above comparisons, we see that decentralized schemes offer the possibility for drastic reductions in communication requirements, at the expense of some performance reduction and of an increase in offline computational requirements. It turns out that the performance reduction is often negligible. Thus, if offline computation is relatively inexpensive, and online communication is severely constrained, a decentralized scheme is preferable.

A decentralized scheme is definitely worth considering in contexts involving geographically distributed sensors. Another application can be found in the context of failure detection whereby different sensors monitor different pieces of equipment and transmit small messages to a central monitoring station that makes the final decision whether a failure has occurred or not. A final context is provided by simple human decision-making organizations [7]. Here each one of a set of decision-makers receives some detailed information and transmits a summary (or a tentative decision) to a supervisor who makes a final decision. Given the limited abilities of humans to handle large amounts of information, such a scheme can be much preferable to having the supervisor receive all available raw data.

On the technical side, decentralized detection problems fall within the class of *team decision problems*. Team decision problems are easily solvable under certain linear–quadratic assumptions [34,47] but are generally intractable [45]. Decentralized detection problems have provided a new paradigm in team decision theory and have aroused considerable technical interest because, in contrast to the intractability of general team problems, definite progress can be made by exploiting their structure.

From another perspective, the subject matter of the theory of decentralized detection coincides with that of quantization theory [18,35,44], but there are some differences in emphasis. In quantization theory, the case of large quantizer alphabets is of particular interest, whereas most research in decentralized detection has focused on very small alphabets. Also, the most common performance criterion in decentralized detection is the probability of error, but this criterion has been avoided in the quantization literature for reasons of mathematical intractability.

This chapter is organized as follows. In the remainder of this section, we define our main model. Then, in Section 2, we introduce a Bayesian cost criterion. Under the assumption of conditional independence of the sensors' observations, we show that the sensors should decide what to transmit by performing a likelihood ratio test. We also discuss some computational issues and comment on the intractability of the problem when the conditional independence assumption is removed.

In Section 3, we consider the Neyman–Pearson variant of the problem, for the case of two hypotheses. Under the conditional independence assumption, the

optimality of likelihood ratio tests is again established. Special attention is paid here to the role of randomization because it introduces some interesting new facets to the problem.

In Section 4, the results of Sections 2 and 3 are extended to more general sensor configurations in which the sensors are connected in the form of a tree. We also comment on non–tree configurations.

In Section 5, we consider the case where the number of sensors is large. Results are provided for the configuration of Figure 1.1. We also discuss the case of a tandem configuration and indicate that its performance is poor.

Finally, in Section 6, we comment on some sequential decentralized detection problems.

Our main objective is to provide a comprehensive review of already available results. We will also occasionally offer some new results and some extensions or more precise statements of old results.

1.2 The Main Model

In this section, we present our model for the case where all sensors transmit summary information to a fusion center. (More general models will be introduced in Sections 4–6.)

We consider a decentralized organization, illustrated in Fig. 1.1, whose basic parameters are the following:

(a) The number of hypotheses, M ($M \geq 2$).

(b) The number of peripheral sensors N ($N \geq 1$).

(c) The number D ($D \geq 2$) of distinct messages that each peripheral sensor can send.

In detail, the setup is as follows: There are M hypotheses $H_1, \ldots, H_M$, with positive prior probabilities $\Pr(H_1), \ldots, \Pr(H_M)$, respectively. We view the true hypothesis as a random variable H that takes the value H_j with probability $\Pr(H_j)$, $j = 1, \ldots, M$. There are $N+1$ sensors $S_0, S_1, \ldots, S_N$. Sensor S_0 plays a special role and will be called the *fusion center*. Each sensor S_i receives an observation Y_i that is a random variable[2,3] taking values in a set $\mathcal{Y}_i$. We assume that the joint probability distribution of $(Y_0, \ldots, Y_N)$, conditioned on H_j, is known for each j.

Each sensor S_i, $i \neq 0$, upon receiving a realization y_i of the random variable Y_i, evaluates a message $u_i = \gamma_i(y_i) \in \{1, \ldots, D\}$, and sends it to the fusion center.[4] Here, γ_i: $\mathcal{Y}_i \rightarrow \{1, \ldots, D\}$ is a function that will be referred to as the *decision rule* of sensor S_i, $i \neq 0$. The fusion center receives a realization y_0 of the random variable Y_0, together with the messages $u_1, \ldots, u_N$, and makes a final decision $u_0 = \gamma_0(y_0, u_1, \ldots, u_N) \in \{1, \ldots, M\}$. Here γ_0: $\mathcal{Y}_0 \times \{1, \ldots, D\}^N \rightarrow \{1, \ldots, M\}$ is a function that will be referred to as the *decision rule* of the fusion center or, alternatively, as the *fusion rule*.[5]

For $i = 0, 1, \ldots, N$, we use Γ_i to denote the set of all possible decision rules of sensor S_i. A collection $\gamma = (\gamma_0, \gamma_1, \ldots, \gamma_N)$ of decision rules will be referred to as a *strategy*. We let $\Gamma = \Gamma_0 \times \Gamma_1 \times \cdots \times \Gamma_N$, which is the set of all strategies.

For each sensor S_i, $i \neq 0$, we also consider *randomized* decision rules, which are constructed as follows. Let $\gamma_i^{(1)}, \ldots \gamma_i^{(K_i)}$ be some elements of Γ_i, where K_i is a positive integer. Suppose now that sensor S_i chooses a particular decision rule $\gamma_i^{(k)}$ and sends the corresponding message $\gamma_i^{(k)}(y_i)$, with each decision rule $\gamma_i^{(k)}$ having probability p_k of being chosen. (It is assumed that this random choice is statistically independent from the random variables $Y_0, Y_1, \ldots, Y_N$, and H.) Thus, a randomized decision rule corresponds to a random choice between (deterministic) decision rules. Randomized decision rules for the fusion center are defined similarly. Let $\overline{\Gamma}_i$ be the set of all randomized decision rules for sensor S_i. With a slight abuse of notation, we will sometimes use the functional notation $u_i = \gamma_i(y_i)$ to represent the message corresponding to a randomized decision rule $y_i \in \overline{\Gamma}_i$, even though u_i is not completely determined by y_i.

In the case where all sensors are to use randomization, there are two distinct alternatives:

(a) Each sensor i has a finite sequence $\{\gamma_i^{(k)} \mid k = 1, \ldots K_i\}$ of candidate deterministic decision rules (i.e., $\gamma_i^{(k)} \in \Gamma_i$) and uses decision rule $\gamma_i^{(k)}$ with probability $p_{i,k}$. Furthermore, the random selections of decision rules by different sensors are statistically *independent*. We let $\overline{\Gamma}$ be the set of all randomized strategies that are so obtained. Notice that $\overline{\Gamma} = \overline{\Gamma}_0 \times \cdots \times \overline{\Gamma}_N$.

(b) There is a finite sequence $\{\gamma^{(k)} \mid k = 1, \ldots, K\}$ of candidate deterministic strategies (i.e., $\gamma^{(k)} \in \Gamma$), and strategy $\gamma^{(k)} = (\gamma_0^{(k)}, \ldots, \gamma_N^{(k)})$ is used with probability p_k. Thus, in this case, the random selections of decision rules by the different sensors are dependent. Let Γ^* be the resulting set of randomized strategies.

For example, if $K_i = K = 2$ for each i, in case (a) each sensor chooses a decision rule by tossing its own coin, whereas in case (b) a single coin is tossed and all sensors choose their decision rule accordingly. We refer to these two cases as *independent* and *dependent* randomization, respectively. Notice that dependent randomization requires larger coordination between the sensors. However, such coordination can be carried offline; no additional online communication is needed.

It is seen that, in the case of independent randomization, we are still choosing probabilistically from a finite set of candidate strategies. [In particular, strategy $(\gamma_0^{(k_0)}, \ldots, \gamma_N^{(k_N)})$ is chosen with probability $\prod_{i=0}^{N} p_{i,k_i}$.] Thus,

$$\overline{\Gamma} \subset \Gamma^*. \tag{1.1}$$

Notice that, for $i \neq 0$, once a decision rule $\gamma_i \in \Gamma_i$ is fixed, the value u_i of the message of sensor S_i can be viewed as a realization of a random variable U_i defined by $U_i = \gamma_i(Y_i)$. Clearly, the probability distribution of U_i depends on the distribution

of Y_i, as well as on the choice of the decision rule γ_i. Similarly, once a strategy $\gamma \in \Gamma$ is fixed, the decision of the fusion center becomes a random variable U_0 defined by $U_0 = \gamma_0(U_1 \ldots, U_N, Y_0)$. The same comment applies to the case of randomized decision rules as well.

1.3 Notation

We use $\Pr(A)$ to denote the probability of an event A. If X is a random variable and A is an event, the notation $\Pr(A \mid X)$ stands for the conditional probability of A, given X. It is a random variable, since it is a function of the random variable X. We use the notation $\Pr(A \mid X = x)$ to denote the value of the random variable $\Pr(A \mid X)$ when the realization of X is equal to x. We will occasionally use the simpler notation $\Pr(A \mid x)$ whenever the nature of the random variable X is clear from the context. Similarly, if Z is another random variable, we will use the notation $E[Z]$ (respectively, $E[Z \mid X]$, $E[Z \mid X = x]$) to denote its expectation (respectively, its conditional expectation, a realization of its conditional expectation). While $E[Z \mid X]$ is not uniquely defined, different versions of $E[Z \mid X]$ are equal with probability 1. In our subsequent use of the notation $E[Z \mid X]$, it will be assumed that a particular version has been chosen. The same comment applies to $\Pr(A \mid X)$.

2. BAYESIAN FORMULATION

In this section, we assume that the organizational objective is the minimization of the expectation of a cost function. A precise formulation of the problem is provided in Section 2.1. After presenting an auxiliary result (Section 2.2), we characterize optimal solutions for the case where the sensors' observations are conditionally independent given any hypothesis (Section 2.3). This characterization is exploited in Section 2.4, in order to suggest some computational approaches to the solution of the problem. Finally, in Section 2.5, we discuss the intractability of the problem when the conditional independence assumption is relaxed.

2.1 Problem Formulation

In the most general Bayesian formulation, we are given a cost function C: $\{1, \ldots, M\} \times \{1, \ldots, D\}^N \times \{H_1, \ldots, H_M\} \to \mathcal{R}$, with $C(u_0, u_1, \ldots, u_N, H_j)$ representing the cost associated to a fusion center decision u_0, and messages $u_1, \ldots, u_N$, when H_j is the true hypothesis. For any strategy $\gamma \in \Gamma^*$, its cost $J(\gamma)$ is defined by

$$J(\gamma) = E[C(U_0, U_1, \ldots, U_N, H)],$$

where $U_0, \ldots, U_N$, are the random variables defined by $U_i = \gamma_i(Y_i)$, $i \neq 0$, and $U_0 = \gamma_0(U_1, \ldots, U_N, Y_0)$. An equivalent expression for $J(\gamma)$, in which the dependence on γ is more apparent, is

$$J(\gamma) = \sum_{j=1}^{M} \Pr(H_j) E[C(\gamma_0(Y_0, \gamma_1(Y_1), \ldots, \gamma_N(Y_N)), \gamma_1(Y_1), \ldots, \gamma_N(Y_N), H_j) \mid H_j].$$

We wish to find an *optimal strategy*, that is, a strategy that minimizes $J(\gamma)$ over all $\gamma \in \Gamma^*$.

Notice that if $\gamma \in \Gamma^*$ is a randomized strategy that uses the deterministic strategies $\gamma^{(1)}, \ldots, \gamma^{(K)} \in \Gamma$ with respective probabilities $p_1, \ldots, p_K$, then

$$J(\gamma) = \sum_{k=1}^{K} p_k J(\gamma^{(k)}). \tag{2.1}$$

Using the fact that $\Gamma \subset \overline{\Gamma} \subset \Gamma^*$ [cf. Eq. (1.1)], and Eq. (2.1), the following result is easily obtained:

Proposition 2.1

$$\inf_{\gamma \in \Gamma^*} J(\gamma) = \inf_{\gamma \in \overline{\Gamma}} J(\gamma) = \inf_{\gamma \in \Gamma} J(\gamma). \tag{2.2}$$

Furthermore, if one of the minimization problems in Eq. (2.2) has a solution, so do the other two.

As a consequence of Proposition 2.1, only deterministic strategies will be considered in the rest of this section.

There are two special cases of interest regarding the choice of the cost function:

CASE A: The function C depends explicitly only on the variable u_0. That is, the performance of the system is judged on the basis of the fusion center's decision.

EXAMPLE 2.1 (*Probability of error criterion*). Let $C(u_0, u_1, \ldots, u_M, H_j)$ be equal to 1 when $u_0 \neq j$, and zero otherwise. Thus, we get unit penalty when the fusion center chooses an incorrect hypothesis, and therefore the objective is the minimization of the error probability.

CASE B: Suppose that $D = M$. In this case, we may wish to interpret each sensor's message as a local decision regarding the nature of the true hypothesis. Then, with a suitable choice of the cost function C, we can penalize incorrect decisions by the fusion center as well as incorrect local decisions by the sensors. As an extreme case, the function C might be independent of u_0, in which case the fusion center becomes irrelevant to the problem and we only need to optimize

with respect to $\gamma_1, \ldots, \gamma_N$; we are then interested only in the quality of the local decisions. The following example is drawn from [66]:

EXAMPLE 2.2 Suppose that $N = M = D = 2$ and that the cost function C is given by

$$C(u_0, u_1, u_2, H_j) = \begin{cases} 0, & \text{if } u_1 = u_2 = j, \\ k, & \text{if } u_1 \neq j \text{ and } u_2 \neq j, \\ 1, & \text{otherwise.} \end{cases} \tag{2.3}$$

Here, k is a real number larger than 2. This cost function penalizes wrong local decisions and imposes a disproportionately large penalty when both are incorrect.

REMARKS

1. A characteristic property of team decision problems (the Bayesian decentralized detection problem being a special case) is that optimization is carried out over a set of functions (in our case, the set Γ). This makes team problems very difficult to solve computationally [45], unless one can guarantee that an optimal strategy can be found within a highly structured set of strategies admitting a finite-dimensional parametrization. It will be seen shortly that this is the case for our problem, under certain assumptions.

2. There is a variant of the problem in which the fusion rule γ_0 is fixed a priori and the minimization is only carried out with respect to $\gamma_1, \ldots, \gamma_N$. It is easily shown that this variant is a special case of our model, for a suitable choice of the function C.

3. When $M = D$, it is very tempting to interpret the messages u_i as tentative local decisions. Sometimes the cost function C might be chosen so as to enforce this interpretation, as discussed in Case B. On the other hand, this interpretation is incorrect in general. For example, when the objective is the minimization of the fusion center's error probability, it may be that the optimal sensor messages are very poor when viewed as local decisions.

2.2 Preliminaries

In this section, we consider a particular centralized Bayesian problem and derive its solution. This result will be used in Section 2.3, as well as in Section 4.

Let there be M hypotheses $H_1, \ldots, H_M$, with respective prior probabilities $\Pr(H_1), \ldots, \Pr(H_M)$, and let X be a random variable, taking values in a set $\mathcal{X}$, with known conditional distribution given each hypothesis. Let D be some positive integer, and let Δ be the set of all functions $\delta: \mathcal{X} \to \{1, \ldots, D\}$. (Consistently with our earlier terminology, we can call such functions decision rules.)

Proposition 2.2 Let Z be a random variable taking values in a set Z and assume that, conditioned on any hypothesis, Z is independent of X. Let F: $\{1, \ldots, D\} \times Z \times \{H_1, \ldots, H_M\} \to \mathcal{R}$ be a given cost function. Let δ^* be an element of Δ. Then, δ^* minimizes $E[F(\delta(X), Z, H)]$ over all $\delta \in \Delta$ if and only if

$$\delta^*(X) = \arg \min_{d=1,\ldots,D} \sum_{j=1}^{M} a(H_j, d)\ Pr(H_j \mid X), \quad \text{w.p.1,} \tag{2.4}$$

where

$$a(H_j, d) = E[F(d, Z, H_j) \mid H_j], \qquad \forall \quad j, d.$$

Proof. The minimization of $E[F(\delta(X), Z, H)]$ over all $\delta \in \Delta$ is equivalent to requiring that $\delta(X)$ minimize $E[F(d, Z, H) \mid X]$, over all $d \in \{1, \ldots, D\}$, with probability 1. The expression being minimized can be rewritten as

$$E[E[F(d, Z, H) \mid H, X] \mid X],$$

which, by the conditional independence of X and Z, is equal to

$$E[E[F(d, Z, H) \mid H] \mid X] = \sum_{j=1}^{M} E[E[F(d, Z, H_j) \mid H_j] \Pr(H_j \mid X). \qquad \text{Q.E.D.}$$

A decision rule δ^* satisfying Eq. (2.4) is almost completely described by the coefficients $a(H_j, d)$, but there are two degrees of freedom. First, if the action of δ^* is modified on a subset of $\mathcal{X}$ of zero probability measure, we still have a decision rule associated to the same set of coefficients $a(H_j, d)$. Clearly, such modifications cannot have any consequences on the expected cost. The second degree of freedom arises when there are ties in the minimization of Eq. (2.4). In this case, the value of $\delta^*(X)$ can be any element of $\{1, \ldots, D\}$ that attains the minimum in Eq. (2.4). An arbitrary tie-breaking rule (e.g., choose the smallest minimizing d) can be used here.

2.3 Optimality Conditions: The Conditionally Independent Case

Throughout this subsection, the following assumption will be in effect:

Assumption 2.1. The random variables $Y_0, \ldots, Y_N$ are conditionally independent, given any hypothesis.

This assumption is satisfied in problems of detecting a known signal, when each sensor's observation is contaminated with independent noise. It is violated if the

sensors' noises are dependent or in problems of detecting an unknown signal in noise. Thus, Assumption 2.1 can be fairly restrictive. Nevertheless, it results to a considerable simplification of the problem and much less progress can be made without it, as will be seen in Section 2.5.

The following result characterizes the outcome of the minimization with respect to the decision rule of a single sensor, when the other decision rules are held fixed.

Proposition 2.3. Let Assumption 2.1 hold.

(a) Fix some $i \neq 0$ and suppose that $\gamma_j \in \Gamma_j$ has been fixed for all $j \neq i$. Then γ_i minimizes $J(\gamma)$ over the set Γ_i if and only if

$$\gamma_i(Y_i) = \arg \min_{d=1,\ldots,D} \sum_{j=1}^{M} \Pr(H_j \mid Y_i) a_i(H_j,d), \quad \text{w.p.1,} \tag{2.5}$$

where

$$a_i(H_j, d) = E[C(\gamma_0(Y_0, U_1, \ldots, U_{i-1}, d, U_{i+1}, \ldots, U_N),$$

$$U_1, \ldots, U_{i-1}, d, U_{i+1}, \ldots, U_N, H_j) \mid H_j], \tag{2.6}$$

and where each U_i, $i \neq 0$, is the random variable defined by $U_i = \gamma_i(Y_i)$.

(b) Suppose that $\gamma_1, \ldots, \gamma_N$ have been fixed. Then γ_0 minimizes $J(\gamma)$ over the set Γ_0 if and only if

$$\gamma_0(Y_0, U_1, \ldots, U_N) = \arg \min_{d=1,\ldots,M} \sum_{j=1}^{M} \Pr(H_j \mid Y_0, U_1, \ldots, U_N)$$

$$\times C(d, U_1, \ldots, U_N, H_j), \quad \text{w.p.1.} \tag{2.7}$$

Proof. (a) Notice that we are concerned with the minimization of

$$E[C(\gamma_0(Y_0, U_1, \ldots, U_{i-1}, \gamma_i(Y_i), U_{i+1}, \ldots, U_N),$$

$$U_1, \ldots, U_{i-1}, \gamma_i(Y_i), U_{i+1}, \ldots, U_N, H_j)]$$

with respect to $\gamma_i \in \Gamma_i$. This is of the form considered in Proposition 2.2, provided that we identify X with Y_i, Z with $(Y_0, U_1, \ldots, U_{i-1}, U_{i+1}, \ldots, U_N)$, and $F(d, Z, H_j)$ with $C(\gamma_0(Y_0, U_1, \ldots, U_{i-1}, d, U_{i+1}, \ldots, U_N), U_1, \ldots, U_{i-1}, d, U_{i+1}, \ldots, U_N, H_j)$. The result then follows from Proposition 2.2.

(b) This is obvious because

$$\min_{\gamma_0 \in \Gamma_0} J(\gamma_0(Y_0, \gamma_1, \ldots, \gamma_N)$$

$$= E[\min_{d=1,\ldots,M} E[C(d, U_1, \ldots, U_N, H) \mid Y_0, U_1, \ldots, U_N]]$$

$$= E[\min_{d=1,\ldots,M} \sum_{j=1}^{M} \Pr[H_j \mid Y_0, U_1, \ldots, U_N) C(d, U_1, \ldots, U_N, H_j)] \qquad \text{Q.E.D.}$$

Notice that if a strategy γ is optimal, then it must also be *person-by-person optimal.* That is, each γ_i must minimize $J(\gamma)$ when the decision rules of all other sensors S_j, $j \neq i$, have been fixed to γ_j. We thus obtain the following.

Corollary 2.1. Let Assumption 2.1 hold. If $\gamma \in \Gamma$ is an optimal strategy, then Eqs. (2.5)–(2.7) hold.

Equations (2.5)–(2.7) are necessary conditions for optimality but do not lead directly to an optimal strategy. This is because the right-hand side of Eq. (2.6) involves the random variables U_k, $k \neq i$, whose distribution depends on the coefficients $a_k(H_j, d)$. Thus, for any fixed choice of γ_0, Eq. (2.6) can be viewed as a system of *NMD* (nonlinear) equations in as many unknowns.

The structure of an optimal fusion rule [cf. Eq. (2.7)] is hardly surprising. For example, if $C(u_0, u_1, \ldots, u_N, H_j)$ equals 0 when $u_0 = j$, and equals 1 otherwise (i.e., if we are minimizing the probability of error), then Eq. (2.7) simplifies to

$$\gamma_0(Y_0, U_1, \ldots, U_N) = \arg \max_{d=1,\ldots,M} \Pr(H_j \mid Y_0, U_1, \ldots, U_N), \qquad \text{w.p.1.} \qquad (2.8)$$

This is the classical MAP (maximum a posteriori probability) rule for hypothesis testing by a sensor that has access to measurements $Y_0, U_1, \ldots, U_N$.

We now assume that each set $\mathcal{Y}_i$ is a Euclidean space and that the probability distribution of each Y_i can be described by a probability density function, under each hypothesis. We can then proceed to obtain an alternative representation of optimal decision rules.[6]

Let $f_{Y_i|H}: \mathcal{Y}_i \times \{1, \ldots, M\} \to [0,\infty)$ be the probability density function of Y_i, conditioned on H. In particular, $f_{Y_i|H}(y_i \mid H_j)$ is the probability density of Y_i, evaluated at some $y_i \in \mathcal{Y}_i$, conditioned on $H = H_j$. Using Bayes's rule, we have, for $i \neq 0$,

$$\Pr(H_j \mid Y_i) = \frac{f_{Y_i|H}(Y_i \mid H_j) \Pr(H_j)}{\sum_{k=1}^{M} f_{Y_i|H}(Y_i \mid H_k) \Pr(H_k)}, \qquad \text{w.p.1.}$$

We use this formula to replace $\Pr(H_j \mid Y_i)$ in Eq. (2.5). We obtain

$$\gamma_i(Y_i) = \arg \min_{d=1,\ldots,D} \sum_{j=1}^{M} \frac{f_{Y_i|H}(Y_i \mid H_j) \Pr(H_j)}{\sum_{k=1}^{M} f_{Y_i|H}(Y_i \mid H_k) \Pr(H_k)} a_i(H_j, d), \quad \text{w.p.1.} \tag{2.9}$$

The denominator in Eq. (2.9) does not depend on j, d, and is positive with probability 1. Thus, Eq. (2.9) is equivalent to

$$\gamma_i(Y_i) = \arg \min_{d=1,\ldots,D} \sum_{j=1}^{M} f_{Y_i|H}(Y_i \mid H_j) \Pr(H_j) a_i(H_j, d)$$

$$= \arg \min_{d=1,\ldots,D} \sum_{j=1}^{M} f_{Y_i|H}(Y_i \mid H_j)\, b_i(H_j, d), \quad \text{w.p.1,} \tag{2.10}$$

where

$$b_i(H_j, d) = a_i(H_j, d) \Pr(H_j). \tag{2.11}$$

We can make a similar argument for the fusion center to obtain

$$\gamma_0(Y_0, u_1, \ldots, u_N) = \arg \min_{d=1,\ldots,M} \sum_{j=1}^{M} f_{Y_0|H}(Y_0 \mid H_j)$$

$$\prod_{i=1}^{N} \Pr(\gamma_i(Y_i) = u_i \mid H_j) \Pr(H_j) C(d, u_1, \ldots, u_N, H_j)$$

$$= \arg \min_{d=1,\ldots,D} \sum_{j=1}^{M} f_{Y_0|H}(Y_0 \mid H_j)\, b_0(H_j, d; u_1, \ldots, u_N), \quad \text{w.p.1,} \tag{2.12}$$

where

$$b_0(H_j, d; u_1, \ldots, u_N) = \Pr(H_j) C(d, u_1, \ldots, u_N, H_j) \prod_{i=1}^{N} \Pr(\gamma_i(Y_i) = u_i \mid H_j). \tag{2.13}$$

REMARKS

1. (*On-line computational requirements.*) Consider the case where the objective is the minimization of the fusion center's error probability. In the decentralized scheme, each sensor S_i, $i \neq 0$, has to compute the likelihoods $f_{Y_i|H}(Y_i \mid H_j)$, $j = 1, \ldots, M$, and perform the algebraic operations prescribed by Eq. (2.10) in order to choose

the value of its message. In a centralized scheme where the fusion center is to make an optimal decision based on all available information, the sensors S_i, $i \neq 0$, could compute the likelihoods $f_{Y_i|H}(Y_i \mid H_j)$, $j = 1, \ldots, M$, and transmit them to the fusion center. (This is because these likelihoods are well-known to be sufficient statistics.) Thus, the online computational requirements at the peripheral sensors are comparable for the centralized and decentralized schemes. A similar comment applies to the fusion center, assuming that the coefficients $b_0(H_j, d, u_1, \ldots, u_N)$ of Eq. (2.12) have been precomputed and are available, say, in the form of a table. This justifies the statement in Section 1 that the centralized and decentralized schemes are comparable regarding online computation.

2. (*The shape of the decision regions at the peripheral sensors.*) Let us call the vector with components $f_{Y_i|H}(Y_i \mid H_j)$, $j = 1, \ldots M$, the likelihood vector of sensor S_i. Equation (2.10) shows that, for $i \neq 0$, this vector is a sufficient statistic for sensor S_i. Furthermore, the M-dimensional space of all likelihood vectors is partitioned into D regions, and a message $u_i = d$ is sent if the vector of likelihoods belongs to the dth region. According to Eq. (2.10) each region is determined by a set of linear inequalities and is therefore polyhedral. This structure is identical to the structure of optimal decision rules in classical *M-ary hypothesis testing.*

3. (*The shape of the decision regions at the fusion center.*) Equation (2.12) shows that the M-dimensional space of all likelihood vectors for sensor S_0 is partitioned into M polyhedral regions, defined by a set of linear inequalities. However, the coefficients $b_0(H_j, d; u_1, \ldots, u_N)$ of these linear inequalities depend on the vector $(u_1, \ldots, u_N)$ of received messages. We thus, have a total of D^N partitions, and the fusion center makes its decision according to the partition corresponding to the received vector of messages. In the special case where the fusion center does not receive any observations Y_0 of its own, the above-mentioned partitions are trivial and the fusion rule is simply a function from $\{1, \ldots, D\}^N \rightarrow \{1, \ldots, M\}$. As such, it can be represented by a precomputable table.

4. (*The fusion rule for the probability of error criterion.*) Suppose that the error criterion is the probability of error. The structure of the fusion rule, as discussed in Remark 3, is determined by a set of coefflcients $b_0(H_j, d; u_1, \ldots, u_N)$. Such a description seems more complicated than the simple statement that the fusion rule should be the MAP rule of Eq. (2.8). In fact, the latter characterization is much more convenient for theoretical purposes. However, when it comes to algorithmic solution, it seems that the computation of the coefficients $b_0(H_j, d; u_1, \ldots, u_N)$ cannot be avoided.

5. Let us consider the variant of the problem whereby the fusion rule γ_0 is a priori fixed and we wish to optimize with respect to $\gamma_1, \ldots, \gamma_N$. It is easily seen that part (a) of Proposition 2.3 remains valid. Thus, the optimal decision rules of the peripheral sensors are again of the form (2.5) and the system of equations (2.6) is still valid.

2.3.1 *The Case of Binary Hypotheses*

We now consider in more detail the case where $M = 2$. Our final objective is to show that optimal strategies can be described by a set of thresholds against which certain likelihood ratios are to be compared.

For each i and j, let $\mathcal{P}_{i|H_j}$ be a probability measure on $\mathcal{Y}_i$ that describes the conditional distribution of Y_i given H_j. We define L_i: $\mathcal{Y}_i \to [0,\infty]$ as the (generalized) *likelihood ratio* of $\mathcal{P}_{i|H_2}$ with respect to $\mathcal{P}_{i|H_1}$.[7] In particular, if $\mathcal{Y}_i$ is a Euclidean space and if the conditional distribution of Y_i is described by a probability density function $f_{Y_i|H}$, then L_i is given by

$$L_i(Y_i) = \frac{f_{Y_i|H}(Y_i \mid H_2)}{f_{Y_i|H}(Y_i \mid H_1)} \qquad \text{w.p.1.} \tag{2.14}$$

We now define the class of decision rules that can be parametrized by a set of thresholds. In this definition, randomized decision rules are also considered because they will be of interest in Section 3.

Definition 2.1. (a) Let $i \neq 0$. A decision rule $\gamma_i \in \overline{\Gamma}_i$ is called a *monotone threshold rule* if there exist thresholds $t_1, \ldots, t_{D-1}$ satisfying $0 \leq t_1 \leq t_2 \leq \cdots \leq t_D \leq \infty$ and with the following property. Let $I_1 = [0, t_1]$, $I_D = [t_{D-1}, \infty]$, and $I_d = [t_{d-1}, t_d]$, $d = 2, \ldots, D-1$. Then,

$$\gamma_i(y_i) = d \qquad \text{only if} \qquad y_i \in I_d.$$

(b) Let $i \neq 0$. A decision rule $\gamma_i \in \overline{\Gamma}_i$ is called a *threshold rule* if there exists a permutation mapping σ: $\{1, \ldots, D\} \to \{1, \ldots, D\}$ such that $\sigma \circ \gamma_i$ is a monotone threshold rule.

(c) A decision rule $\gamma_0 \in \overline{\Gamma}_0$ is called a *threshold rule* if for every $u \in \{1, \ldots, D\}^N$ there exists some $t(u) \in [0,\infty]$ and a permutation σ^u: $\{1, 2\} \to \{1, 2\}$ such that

$$\gamma_0(y_0, u) = \begin{cases} \sigma^u(1), & \text{if } L_0(y_0) < t(u), \\ \sigma^u(2), & \text{if } L_0(y_0) > t(u). \end{cases}$$

(d) A decision rule $\gamma_0 \in \overline{\Gamma}_0$ is called a *monotone threshold rule* if (i) it is a threshold rule, (ii) σ^u is the identity permutation for every $u \in \{1, \ldots, D\}^N$ and (iii) $t(u)$ is a nonincreasing function of u.

(e) We say that a strategy $\gamma = (\gamma_0, \ldots, \gamma_N) \in \overline{\Gamma}$ is a *(monotone) threshold strategy* if each γ_i is a (monotone) threshold rule.

Proposition 2.4. Let $M = 2$. Suppose that Assumption 2.1 holds and that there exists an optimal strategy. Then, there exists an optimal strategy that is a deterministic threshold strategy. Furthermore, if the cost criterion is the fusion

center's probability of error, then there exists an optimal strategy that is a deterministic monotone threshold strategy.

Proof. Let γ be an optimal strategy. Equation (2.5) yields, for $i \neq 0$,

$$\gamma_i(Y_i) = \arg \min_{d=1,\ldots,D} [\Pr(H_1 | Y_i) a_i(H_1, d) + \Pr(H_2 | Y_i) a_i(H_2, d)], \qquad \text{w.p.1},$$

Using Bayes's rule, we have

$$\frac{\Pr(H_2 | Y_i)}{\Pr(H_1 | Y_i)} = L_i(Y_i) \frac{\Pr(H_2)}{\Pr(H_1)}, \qquad \text{w.p.1},$$

and it follows that

$$\gamma_i(Y_i) = \begin{cases} \arg \min_{d=1,\ldots,D} \left[a_i(H_1, d) + L_i(Y_i) a_i(H_2, d) \frac{\Pr(H_2)}{\Pr(H_1)} \right], & \text{if } L_i(Y_i) < \infty, \\ \arg \min_{d=1,\ldots,D} a_i(H_2, d) \frac{\Pr(H_2)}{\Pr(H_1)}, & \text{if } L_i(Y_i) = \infty, \end{cases} \qquad \text{w.p.1.} \tag{2.15}$$

While γ_i itself is not necessarily a threshold rule [e.g., if $a_i(H_j, d) = 0$ for all j, d, then the function γ_i could be anything], it is easily seen (Figure 2.1) that there exists a deterministic threshold rule $\bar{\gamma}_i$ that satisfies Eq. (2.15). Notice that Eq. (2.15) is equivalent to Eq. (2.5), which is a necessary and sufficient condition for γ_i to minimize $J(\gamma)$ while the other decision rules γ_j, $j \neq i$, are held fixed. This shows that replacing γ_i by $\bar{\gamma}_i$ cannot increase the value of $J(\gamma)$. This argument shows that the decision rules of all peripheral sensors can be replaced by deterministic threshold rules, without losing optimality. A similar argument leads to the same conclusion for γ_0.

Consider now the case where the error criterion is the fusion center's probability of error. Suppose that $\gamma \in \Gamma$ is an optimal strategy that is a deterministic threshold strategy. Then, for each $i \neq 0$, we can replace γ_i by $\bar{\gamma}_i = \sigma_i \circ \gamma_i$, where σ_i is a permutation mapping that makes γ_i monotone, without changing the information conveyed to the fusion center. Thus, the performance of the fusion center will be unaffected, provided that γ_0 is replaced by $\bar{\gamma}_0$, where $\bar{\gamma}_0(y_0, u_1, \ldots, u_N) = \gamma_0[y_o, \sigma_1^{-1}(u_1), \ldots, \sigma_N^{-1}(u_N)]$. The monotonicity of $\bar{\gamma}_i$, $i \neq 0$, implies that $\Pr(U_i = u_i | H_2)/\Pr(U_i = u_i | H_1)$ is nondecreasing in u_i. Now, recall that the MAP rule is optimal for the fusion center. Thus, the fusion center makes a decision by comparing

$$L_0(Y_0) \prod_{i=1}^{N} \frac{\Pr(U_i | H_2)}{\Pr(U_i | H_1)}$$

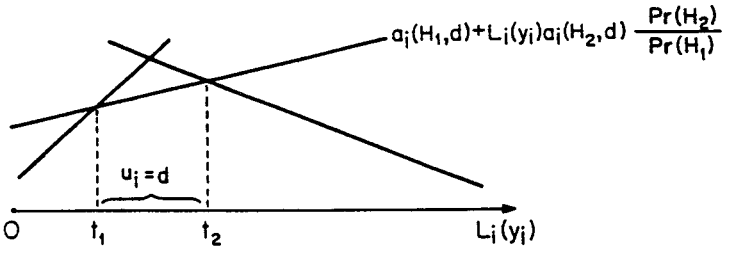

Figure 2.1.

to a threshold t. Equivalently, $L_0(Y_0)$ is compared to a threshold

$$t \prod_{i=1}^{N} \frac{\Pr(U_i|H_1)}{\Pr(U_i|H_2)},$$

which is nonincreasing in $U_1, \ldots, U_N$. This establishes the existence of an optimal strategy within the set of deterministic monotone threshold strategies. Q.E.D.

The converse of Proposition 2.4 is not true, in general. Namely, there may exist optimal strategies that are not threshold strategies. This can happen, for example, if the observation Y_i of some peripheral sensor is so poor that the fusion center always disregards the message U_i. In that case, the decision rule γ_i can be arbitrarily chosen without affecting the performance of the system. However, we have the following result.

Proposition 2.5. Let $M = 2$ and suppose that Assumption 2.1 holds. Suppose that γ is an optimal strategy under the error probability criterion. Then, either γ is a threshold strategy (possibly, after being modified on a set of zero probability measure), or the alphabet size of some peripheral sensor can be reduced without increasing the value of the optimal cost.

Proof (Outline). Fix an optimal strategy γ and consider sensor S_i, $i \neq 0$. If γ_i violates the definition of a threshold rule on a set of positive probability, then it is clear from Eq. (2.15) and Figure 2.1 that there exist some d, d', with $d \neq d'$ such that $a_i(H_1, d) = a_i(H_1, d')$ and $a_i(H_2, d) = a_i(H_2, d')$. If we now remove d' from the set $\{1, \ldots, D\}$ and let sensor S_i transmit d in its place, Eq. (2.15) will still hold, and optimality is retained. The argument for the case $i = 0$ is straightforward and is omitted. Q.E.D.

As discussed earlier, Eq. (2.6) is really a system of equations that must be satisfied by an optimal strategy. In the case of binary hypotheses, it is straightforward to rewrite (2.6) as a system of equations involving the thresholds of the different sensors. These equations are necessary conditions for optimality that must be satisfied by any optimal threshold strategy. It would be very desirable if sufficient conditions for optimality were available, but no such conditions have been proved so far. Even worse, it is conceivable that a threshold strategy could be person-by-person optimal [and thus satisfy Eqs. (2.5)–(2.7)] without being "locally optimal." This is because person-by-person optimality means that the cost cannot be reduced by small perturbations of the thresholds of any individual sensor, but implies nothing regarding the simultaneous perturbation of several thresholds.

2.3.2 *Identical Sensors*

An interesting special case arises when the random variables Y_i, $i \neq 0$, are identically distributed, given each hypothesis, and the cost function C is symmetric in the variables $u_1, \ldots, u_N$. In this case, symmetry considerations lead to the conjecture that we can constrain the peripheral sensors to use identical decision rules without increasing the optimal cost. This conjecture is false even for the simplest case where $M = D = N = 2$ and the fusion center receives no information of its own, as the following counterexample from [66] shows. Consider Example 2.2 of Section 2.1. If k is extremely large [cf. Eq. (2.3)], then the two sensors should ensure that their messages are different with high probability, and this is possible only if they use different thresholds. This counterexample depends heavily on a somewhat artificial aspect of the cost function. A less artificial counterexample has been given in [59] in which the cost criterion is the probability of error by the fusion center.

Numerical experimentation suggests that, for the case where $M = 2$ and for the error probability criterion, the restriction $\gamma_1 = \cdots = \gamma_N$ often results in little or no loss of optimality. By enforcing this restriction, the problem simplifies considerably because the number of independent parameters is reduced. Furthermore, under this restriction, the random variables $U_1, \ldots, U_N$ become identically distributed. In the special case where $D = 2$ and the fusion center receives no observations of its own, the number K, defined as the cardinality of the set $\{i \mid U_i = 1\}$, becomes a sufficient statistic for the fusion center. Thus an optimal fusion rule can be restricted to have the form $\gamma_0(U_1, \ldots, U_N) = 1$ if and only if $K \leq k^*$, where k^* is a threshold value. (Such decision rules are often referred to as "k^*-out-of-N.")

An explanation for the little loss of performance caused by the equal-threshold restriction will be provided in Section 5.1. In addition, [38] provides some interesting upper bounds on the deterioration of performance when the equal-threshold restriction is imposed. It would be even more comforting if the optimality of equal thresholds were to be established for special cases, but no such results are available.

NOTES. The Bayesian decentralized detection problem was first studied by Tenney and Sandell [66], who established the optimality of threshold strategies. While they only considered the special case where $M = N = D = 2$ and where the cost function C does not depend on u_0, their proof contained all the essential ingredients for establishing Propositions 2.3 and 2.4 in the generality presented here. Such generalizations have been provided in [14, 15, 26, 51, 64]. The monotonicity of γ_0 has been observed in [49], among others. The case of identical sensors is discussed in [26, 48, 49, 55]. Proposition 2.5 is new. Our discussion has bypassed the question of the existence of an optimal strategy. In fact, existence results can be established, under the conditional independence assumption, for all the problems considered in this chapter [61], by appealing to a variant of Liapounoff's theorem [30].

2.4 Computational Issues

We discuss here some issues related to the computation of an optimal strategy. We first perform a rough operation count for the case when the observation spaces $\mathcal{Y}_i$ are discrete. We then discuss possible algorithms for the general case. Assumption 2.1 (conditional independence) is assumed throughout to be in effect.

2.4.1 *Binary Hypotheses and Discrete Observations*

Suppose that there are two hypotheses ($M = 2$) and that each set $\mathcal{Y}_i$ is finite, with cardinality n_0 if $i = 0$, and cardinality n if $i \neq 0$. We consider the minimum probability of error criterion.

Suppose that the peripheral sensors' decision rules $\gamma_1, \ldots, \gamma_N$ have been fixed. Assuming that the fusion center uses the corresponding MAP rule, it is easily checked that $J(\gamma)$ can be evaluated with $O(Nn + n_0 D^N)$ arithmetic operations.

Suppose now that we wish to find an optimal strategy. We only need to search over the set of monotone threshold strategies (Proposition 2.4). For each peripheral sensor S_i, its likelihood ratio $L_i(Y_i)$ can take at most n different values; thus each threshold can be restricted to a set of cardinality $n + 1$ and we only need to consider $O(n^D)$ different decision rules. (Taking into account the fact that the thresholds of a sensor form a nondecreasing sequence, the number of decision rules examined can be somewhat reduced.) We conclude that the number of choices for $(\gamma_1, \ldots, \gamma_N)$ is $O(n^{DN})$. For each such choice, we have to evaluate the associated probability of error, and the total computational cost is $O[n^{DN}(Nn + n_0 D^N)]$. Even for moderate values of n, D, and N, we see that the computational requirements can be enormous. Nevertheless, it should be noted that if the numbers D and N are held fixed, then the complexity is polynomial in n and n_0.

Suppose now that all sensors are identical and that $D = 2$. Let us impose the requirement that each sensor uses the same threshold. Even though optimality could be lost as a consequence of this assumption, the computational requirements are drastically reduced. In particular, the number of choices for $(\gamma_1, \ldots, \gamma_N)$ is reduced to $O(n)$. Notice that the number of sensors for which $U_i = 1$ is a binomial random variable and is a sufficient statistic for the fusion center. Thus, the computation needed for evaluating the expected cost of a strategy is only $O(N)$

(assuming that a table with the binomial coefficients is available). We conclude that the total computational requirements are reduced to $O(nN)$.

A corollary of the preceding discussion is that the complexity of computing an optimal strategy is, in general, exponential in N, but becomes polynomial in N when the sensors are identical, use identical decision rules, and $D = 2$. The same conclusion can also be reached for any fixed $D > 2$; however, the complexity grows exponentially with D.

2.4.2 *The General Case*

We still consider the minimum probability of error criterion. An unrealistic approach is to discretize the sets $\mathcal{Y}_i$ and use exhaustive enumeration over the set of all decision rules of the form (2.5) (or over the set of threshold rules, if $M = 2$). A more practical approach is to start with some strategy γ^0 and generate a sequence $\{\gamma^k\}$ of strategies by successively optimizing with respect to the decision rule of each sensor. Formally,

$$\gamma_i^{k+1} = \arg \min_{\gamma_i \in \Gamma_i} J(\gamma_0^{k+1}, \ldots, \gamma_{i-1}^{k+1}, \gamma_i, \gamma_{i+1}^{k}, \ldots, \gamma_N^{k}). \tag{2.16}$$

[If several elements of Γ_i attain the minimum in Eq. (2.16), one of them is selected arbitrarily.]

The algorithm of Eq. (2.16) is a general-purpose method for the minimization of multivariable functions [31], and is known as the *nonlinear Gauss–Seidel* algorithm. The provable properties of this algorithm are not particularly strong. On the positive side, the sequence $J(\gamma^k)$ of the costs corresponding to the sequence γ^k of strategies generated by this algorithm is nonincreasing. On the negative side, if the algorithm is initialized with a person-by-person optimal strategy, it will not make any progress, even if this strategy is not optimal. Furthermore, even though the sequence of costs is guaranteed to converge (by virtue of being nonincreasing and bounded below), there are no provable guarantees for the convergence of the sequence of strategies. For example, it is not known whether the sequence of strategies is guaranteed to converge or whether every limit point of the sequence of strategies is guaranteed to be person-by-person optimal. Nevertheless, the algorithm (2.16) has been found to be quite successful in practice, with convergence taking place within a reasonably small number of iterations.

The computations associated with each iteration of the algorithm (2.16) depend on the parametrization of the decision rules. Any decision rules generated by the algorithm can be parametrized by a finite set of coefficients $a_i(H_j, d)$ as in Eq. (2.5), or by a set of coefficients $b_i(H_j, d)$ and $b_0(H_j, d; u_1, \ldots, u_N)$, as in Eqs. (2.10) and (2.12) (the latter parametrization being more convenient, in general). Furthermore, in the case where $M = 2$, decision rules can be parametrized by a set of thresholds. Still, no matter what parametrization is employed, the minimization with respect to γ_i, $i \neq 0$, cannot avoid the computation of the coefficients $a_i(H_j, d)$ according to

formula (2.6). This requires a computational effort of the order of D^N, which can be quite time consuming unless D and N are small. An alternative option is to evaluate the expectation in Eq. (2.6) using a Monte Carlo method.

There is an alternative version of the algorithm (2.16), whereby each minimization with respect to some γ_i, $i \neq 0$, is immediately followed by a minimization with respect to γ_0. The rationale behind this version is that the fusion rule is much more crucial than the peripheral sensors' decision rules and therefore should be optimized more often. In a diametrically opposite variation, the fusion rule γ_0 is fixed and the decision rules of the peripheral sensors are successively minimized. However, as the optimal choice of γ_0 is unknown, this procedure has to be repeated for all $\gamma_0 \in \Gamma_0$. Assuming that the fusion center receives no observations Y_0 of its own, the set Γ_0 is finite. Furthermore, certain choices of γ_0 can be excluded a priori, due to monotonicity considerations (Proposition 2.4). Still, the number of choices for γ_0 is excessive and this approach is not practical, unless N and D are very small. At present, there are no computational results comparing the performance of the algorithm (2.16) and the alternatives discussed in this paragraph.

We now turn to the case of two hypotheses and identical sensors. If we restrict the decision rules of the peripheral sensors to be identical, the iteration (2.16) is not applicable and some other method for searching over the $(D-1)$-dimensional set of thresholds is needed. We have the following options: (a) We can use a global minimization algorithm, the simplest one being exhaustive search over a discretization of the threshold space. (b) We can perform some form of gradient search (which might however get stuck at a local, but not global, minimum). To do this, we need the gradient of the cost with respect to the thresholds. This gradient can be shown to exist, and to be easily computable, under a continuity assumption on the probability density function of the likelihood ratio $L_i(Y_i)$.

These search methods are further facilitated by the fact that the computation of $J(\gamma)$ is very easy when all peripheral sensors use the same decision rule. We are not aware of any comparison of the different search methods with the algorithm (2.16).

NOTES. The algorithm (2.16) is described in [64,65], but its use can be traced back earlier. An interesting graphical method for the case $M = D = 2$ and identical sensors is described in [48]. Most of the examples whose numerical solution has been reported in the literature use very small values of N, M, and D (typically 2). It is thus unclear what are the "largest" problems that can be efficiently solved. It is encouraging to note that numerical experiments suggest that the value of the cost function is quite insensitive to moderate variations of the values of the thresholds of the peripheral sensors. Furthermore, the optimal cost is sometimes very close to the optimal centralized cost, even if D is small (e.g., $D = 4$ or 5 [37]); thus, large values of D do not seem to be of interest.

2.5 Conditionally Dependent Observations

In Section 2.3, it was shown that under the assumption of conditional independence (Assumption 2.1), the set of all strategies can be replaced by a much

smaller set of strategies (e.g., the set of monotone threshold strategies, when $M = 2$). Unfortunately, this is no longer the case without the conditional independence assumption, as illustrated by the following example.

EXAMPLE 2.3. Consider the case of two equally likely hypotheses ($M = 2$), two peripheral sensors ($N = 2$), and let $D = 2$. Suppose that $\mathcal{Y}_1 = \mathcal{Y}_2 = \{1,2,3,4\}$, $Y_1 = Y_2$ with probability 1 (thus, Y_1 and Y_2 are not conditionally independent), and that the likelihood ratio $L_i(Y_i)$ is an increasing function of Y_i. Suppose that the fusion center also obtains some information Y_0 of its own and that the cost criterion is the fusion center's probability of error. Consider the following decision rules: $\gamma_1(Y_1) = 1$ if and only if $Y_1 \in \{1,2\}$ and $\gamma_2(Y_2) = 1$ if and only if $Y_2 \in \{1,3\}$. It is clear that the fusion center can use the messages $\gamma_1(Y_1)$ and $\gamma_2(Y_2)$ to fully reconstruct the value of Y_1 and Y_2. On the other hand, if we constrain the sensors S_1 and S_2 to use threshold rules, it is easily seen that the fusion center will not be able to reconstruct the value of Y_1 and Y_2, and this will result, in general, to some loss of performance. [For example, we can try $\gamma_1(Y_1) = 1$ if and only if $Y_1 = 1$, and $\gamma_2(Y_2) = 1$ if and only if $Y_2 \in \{1,2\}$. Then, the fusion center cannot tell the difference between $Y_1 = 3$ and $Y_1 = 4$.] We conclude that threshold strategies are not, in general, optimal.

The failure of threshold strategies to attain optimality has some dramatic computational ramifications. For example, consider the case where $N = M = D = 2$ and the sets $\mathcal{Y}_1$, $\mathcal{Y}_2$ have finite cardinality n. As was shown in Section 2.4, under the conditional independence assumption, there exists a polynomial time algorithm for finding an optimal strategy. On the other hand, without the conditional independence assumption, we may have to examine all possible strategies, and their number is exponential in n. We may wonder whether a radically different algorithm exists that could solve the problem in polynomial time. The answer seems to be negative, even for the simplest cases, because of the following result. (The reader is referred to [21] or [43] for an introduction to the theory of NP-completeness.)

Proposition 2.6. Consider the minimum probability of error decentralized detection problem for the case where $M = N = D = 2$, where the fusion center receives no information of its own (other than the messages), and where the sets $\mathcal{Y}_i$, $i = 1,2$, are finite. Let K be a rational number. Then the problem of deciding whether there exists a strategy γ such that $J(\gamma) \leq K$ is *NP*-complete. In particular, unless $P = NP$, there is no polynomial time algorithm for this problem.

Thus, the discrete version of the decentralized detection problem, without the conditional independence assumption, is an inherently intractable combinatorial problem. Furthermore, the problem's difficulty is not particular to the discrete version because, as shown in [45], complexity results for discrete team decision problems can be translated to precise negative results for the continuous versions of these problems.

From the algorithmic point of view, the iterative algorithm (2.16) can be used, and it is not unreasonable to expect that this method would often converge to a person-by-person optimal strategy. On the other hand, given the intractability of

the problem, either this person-by-person optimal strategy will be nonoptimal, or the number of computations until termination will be excessive. Nevertheless, intractability results such as Proposition 2.6 refer to the worst case. It is conceivable that in practical problems, algorithm (2.16) could have acceptable performance. However, no computational experience is available.

A difficulty related to the nonoptimality of threshold strategies is that there does not exist a finite-dimensional parametrization of the candidate strategies (unlike the conditionally independent case). For example, suppose that the set $\mathcal{Y}_i$ is the real line and that $D = 2$. Any decision rule γ_i partitions the real line into a number of intervals and the value of $\gamma_i(Y_i)$ changes each time that we cross from one interval to the next. Such a decision rule can be parametrized by specifying the end points of the intervals. However, there is no a priori bound on the number of intervals, and therefore this parametrization is infinite dimensional. (If an a priori bound existed, then, for the discrete version of the problem, the number of candidate strategies would be polynomial in the cardinality of the set $\mathcal{Y}_i$ and a polynomial time algorithm would result, which is unlikely in view of Proposition 2.6.) Thus, any attempt to express the person-by-person optimality conditions as a system of equations in a finite number of scalar parameters is condemned to fail.

A remedy to the above described infinite dimensionality of the problem is to impose an artificial constraint forcing the strategies into a finite-dimensional set. For example, we could restrict to threshold strategies. This results to loss of optimality, as demonstrated by Example 2.3, but might perform reasonably well in practice.

NOTES. Proposition 2.6 is from [57,62]. An approximate computational approach is suggested in [64] in which certain dependencies of the Y_i are ignored at certain stages of the algorithm.

3. NEYMAN–PEARSON FORMULATION

In this section, we consider the Neyman–Pearson variant of the problem of Section 2, for the case of two hypotheses. In particular, we establish the optimality of monotone threshold strategies (under certain assumptions) and we discuss some issues related to the computation of an optimal strategy. While in the theory of centralized detection, the Bayesian and the Neyman–Pearson problems are almost equivalent, it will be seen that the situation is somewhat more complex in decentralized detection.

Throughout this section, we will be assuming that $M = 2$ and that Assumption 2.1 (conditional independence) is in effect.

3.1 Problem Definition

For any strategy γ, we define the *false-alarm* and *detection* probabilities, by

$$J^F(\gamma) = \Pr(\gamma_0[Y_0, \gamma_1(Y_1), \ldots, \gamma_N(Y_N)] = 2 \mid H_1), \tag{3.1}$$

$$J^{D}(\gamma) = \Pr(\gamma_0[Y_0, \gamma_1(Y_1), \ldots, \gamma_N(Y_N)] = 2 \mid H_2), \tag{3.2}$$

respectively. Let $\alpha \in (0,1)$ be a scalar that prescribes the allowed tolerance on the false-alarm probability. The Neyman–Pearson problem is as follows:

$$\text{maximize } J^{D}(\gamma) \qquad \text{subject to} \qquad J^{F}(\gamma) = \alpha. \tag{3.3}$$

There are three different versions of problem (3.3), depending on whether we constrain γ to lie in the set Γ (deterministic strategies), $\overline{\Gamma}$ (strategies with independent randomization), or Γ^* (strategies with dependent randomization). While for the Bayesian problem, strategies could be restricted to the set Γ without any loss of performance (see Proposition 2.1), this is not the case here.

For the three sets of strategies under consideration (Γ, $\overline{\Gamma}$, and Γ^*), we define

$$Q = \{(J^{F}(\gamma), J^{D}(\gamma)) \mid \gamma \in \Gamma\}, \tag{3.4}$$

$$\overline{Q} = \{(J^{F}(\gamma), J^{D}(\gamma)) \mid \gamma \in \overline{\Gamma}\}, \tag{3.5}$$

$$Q^* = \{(J^{F}(\gamma), J^{D}(\gamma)) \mid \gamma \in \Gamma^*\}. \tag{3.6}$$

The upper boundary of Q, $\overline{Q}$, and Q^* corresponds to the optimal solutions of problem (3.3) over the sets Γ, $\overline{\Gamma}$, and Γ^*, respectively, for the different values of α. Thus, the sets Q, $\overline{Q}$, and Q^* essentially generalize the classical concept of the receiver operating characteristic (ROC) curve.

Proposition 3.1. (a) $Q \subset \overline{Q} \subset Q^*$.
(b) Q^* is the convex hull of Q and $\overline{Q}$,.

Proof. (a) This is obvious because $\Gamma \subset \overline{\Gamma} \subset \Gamma^*$.

(b) Let $\gamma \in \Gamma^*$ and suppose that γ uses the deterministic strategies $\gamma^{(1)}, \ldots, \gamma^{(K)} \in \Gamma$, with respective probabilities $p_1, \ldots, p_K$. It is easily seen that

$$J^{F}(\gamma) = \sum_{k=1}^{K} p_k J^{F}(\gamma^{(k)}), \qquad J^{D}(\gamma) = \sum_{k=1}^{K} p_k J^{D}(\gamma^{(k)}),$$

which shows that $(J^{F}(\gamma), J^{D}(\gamma))$ is a convex combination of $J^{F}(\gamma^{(k)}), J^{D}(\gamma^{(k)}))$, $k = 1, \ldots, K$. Thus, Q^* is the convex hull of Q. Using part (a), it follows that Q^* is also the convex hull of $\overline{Q}$. Q.E.D.

We now introduce a condition that can lead to some convenient simplifications, by eliminating the need for tie-breaking rules.

Definition 3.1. For any i, we say that the likelihood ratio $L_i(Y_i)$ has *no point mass* if

$$\Pr(L_i(Y_i) = x \mid H_j) = 0, \qquad \forall \quad x \in [0,\infty], j = 1, 2.$$

The property of Definition 3.1 typically holds when Y_i has a continuous distribution. An important case where this property does not hold is when $i = 0$ and the fusion center has no information of its own. [In that case, $L_0(Y_0)$ is identically equal to 1.] An easy consequence of the no-point-mass condition is the following:

Proposition 3.2. If $L_i(Y_i)$ has no point mass and $\bar{\gamma}_i \in \bar{\Gamma}_i$ is a threshold rule, then there exists a deterministic threshold rule $\gamma_i \in \Gamma_i$ such that $\gamma_i(Y_i) = \bar{\gamma}_i(Y_i)$, with probability 1.

Proof. With a randomized threshold rule, randomization can only take place when $L_i(Y_i)$ is equal to a threshold, which occurs with probability zero, by the no-point-mass assumption.
Q.E.D.

3.2 Preliminaries

Let us fix some sensor S_i, $i \neq 0$. Let $c_1, \ldots, c_D$ and $\alpha_1, \ldots, \alpha_D$ be two collections of scalars such that $\alpha_1 + \cdots + \alpha_D = 1$ and $a_d \geq 0$ for each d. Consider the problem

$$\text{maximize} \quad \sum_{d=1}^{D} c_d \Pr(\gamma_i(Y_i) = d \mid H_2) \tag{3.7}$$

$$\text{subject to} \quad \Pr(\gamma_i(Y_i) = d \mid H_1) = \alpha_d, \quad d = 1, \ldots, D. \tag{3.8}$$

Notice that if we let $D = 2$, $c_1 = 0$, $c_2 = 1$, then problem (3.7)–(3.8) coincides with the classical (centralized) Neyman–Pearson problem.

Proposition 3.3. Consider the optimization problem (3.7)–(3.8) over the set $\bar{\Gamma}_i$ of all randomized decision rules and assume that an optimal solution exists.

(a) If $c_1 \leq c_2 \leq \cdots \leq c_D$, then there exists a monotone threshold rule $\gamma_i \in \bar{\Gamma}_i$ that is optimal.
(b) In the general case, there exists a threshold rule $\gamma_i \in \bar{\Gamma}_i$ that is optimal.

Sketch of the proof. Part (b) is a consequence of part (a), because the elements of $\{1, \ldots, D\}$ can be renamed so that the relation $c_1 \leq c_2 \leq \cdots \leq c_D$ holds. As for part (a), it is proved by an argument identical to the standard proof of the Neyman–Pearson lemma. That is, if γ_i is not a monotone threshold rule, we can transform it to one by performing pairwise interchanges of sets

$A_d \subset \{Y_i \mid \gamma_i(Y_i) = d\}$, $A_{d'} \subset \{Y_i \mid \gamma_i(Y_i) = d'\}$, in a way that constraints (3.8) remain valid, and without reducing the value of the expression (3.7). Q.E.D.

Let us now consider the following variation of problem (3.7)–(3.8):

$$\text{maximize} \quad \sum_{d=1}^{D} c_d \Pr(\gamma_i(Y_i) = d \mid H_2) \tag{3.9}$$

$$\text{subject to} \sum_{d=1}^{D} h_d \Pr(\gamma_i(Y_i) = d \mid H_1) = \alpha, \tag{3.10}$$

over the set $\overline{\Gamma}_i$ of all randomized decision rules, where $\alpha, c_1, h_1, \ldots, c_D, h_D$ are given scalars. Let γ_i^* be an optimal solution to this problem and let $a_d = \Pr(\gamma_i^*(Y_i) = d \mid H_1)$, $d = 1, \ldots, D$. Consider problem (3.7)–(3.8) with this particular choice of $\alpha_1, \ldots, \alpha_D$. It is clear that γ_i^* is an optimal solution of problem (3.7)–(3.8) as well. Furthermore, any optimal solution of the problem (3.7)–(3.8) is also an optimal solution of problem (3.9)–(3.10). We can now apply Proposition 3.3 to obtain the following.

Proposition 3.4. Consider the optimization problem (3.9)–(3.10) over the set $\overline{\Gamma}_i$ of all randomized decision rules and assume that an optimal solution exists.

(a) If $c_1 \le c_2 \le \cdots \le c_D$, then there exists a monotone threshold rule $\gamma_i \in \overline{\Gamma}_i$ that is optimal.

(b) In general, there exists a threshold rule $\gamma_i \in \overline{\Gamma}_i$ that is optimal.

3.3 The Case of Independent Randomization

We return to the original problem (3.3) and consider the case where the optimization is to be carried out over the set $\overline{\Gamma}$ of strategies with independent randomization. The main result is the following.

Proposition 3.5. Let Assumption 2.1 hold. Suppose that the optimization problem (3.3) over the set $\overline{\Gamma}$ of randomized strategies has an optimal solution. Then, there exists a monotone threshold strategy that is optimal.

Proof. Let $\gamma^* = (\gamma_0^*, \ldots, \gamma_N^*) \in \overline{\Gamma}$ be an optimal strategy. Let us consider sensor S_1. Then γ^* must maximize $J^D(\gamma)$ over all $\gamma \in \overline{\Gamma}$ such that $J^F(\gamma) = \alpha$ and $\gamma_j = \gamma_j^*$ for $j \neq 1$. We will express $J^D(\gamma)$ and $J^F(\gamma)$ in terms of γ_1. Let U be the random vector defined by $U = (\gamma_2^*(Y_2), \ldots, \gamma_N^*(Y_N))$. Then

$$J^D(\gamma) = \Pr(\gamma_0^*(Y_0, \gamma_1(Y_1), U) = 2 \mid H_2)$$

$$= \sum_{d=1}^{D} \Pr(\gamma_0^*(Y_0, d, U) = 2 \mid H_2) \Pr(\gamma_1(Y_1) = d \mid H_2) \tag{3.11}$$

$$= \sum_{d=1}^{D} c_d \Pr(\gamma_1(Y_1) = d \mid H_2),$$

where $c_1, \ldots, c_D$ are defined in the obvious manner. A similar calculation shows that

$$J^F(\gamma) = \sum_{d=1}^{D} h_d \Pr(\gamma_1(Y_1) = d \mid H_1), \tag{3.12}$$

for some suitable scalars $h_1, \ldots, h_D$. Notice that the problem of maximizing (3.11) over all $\gamma_i \in \overline{\Gamma}_1$ and subject to (3.12), has the form (3.9)–(3.10). By Proposition 3.4, this problem has an optimal solution $\overline{\gamma}_1$ that is a threshold rule. We now replace γ_i^* by $\overline{\gamma}_i$, and we still have an optimal strategy for the original problem (3.3) . By repeating this argument for sensors $S_2, \ldots, S_N$, we end up with an optimal solution in which the decision rule of each peripheral sensor is a threshold rule. Furthermore, threshold rules can be replaced by monotone threshold rules without changing the information conveyed to the fusion center, and thus without losing optimality.

Assuming now that $\gamma_1, \ldots, \gamma_N$ are monotone threshold rules, the fusion center is faced with a classical (centralized) Neyman–Pearson problem and should use a threshold rule. The fact that γ_0 can be chosen monotone follows exactly as in the proof of Proposition 2.4. Q.E.D.

Notice that (by Proposition 3.2), if $L_i(Y_i)$ has no point mass for each i, then independent randomization offers no advantage over deterministic strategies. The following result elaborates further on the necessity of randomization.

Proposition 3.6. Let Assumption 2.1 hold and suppose that $L_i(Y_i)$ has no point mass for every $i \neq 0$. If there exists a deterministic strategy $\gamma \in \Gamma$ that is optimal for problem (3.3) over the set $\overline{\Gamma}$, then such a deterministic strategy can be found within the set of deterministic monotone threshold strategies.

Proof. The argument parallels the proof of Proposition 3.5. We start with a deterministic strategy γ^*, assumed to be optimal over the set $\overline{\Gamma}$ (the existence of such a strategy is assumed in the proposition's statement). The same argument as in Proposition 3.5, together with Proposition 3.2, shows that the decision rules of all peripheral sensors can be replaced by deterministic monotone threshold rules. Now, the fusion center is faced with a classical Neyman–Pearson problem

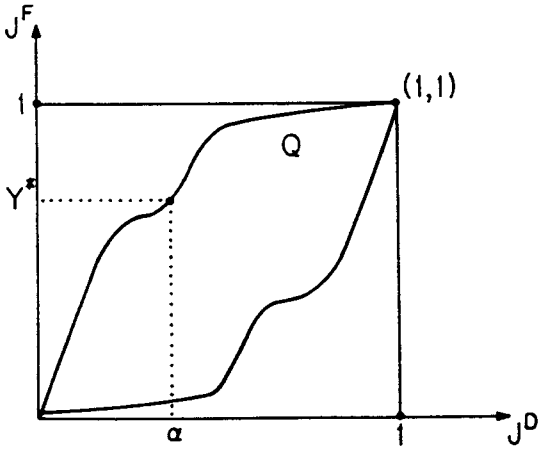

Figure 3.1. Here (α, y^*) corresponds to an optimal solution of the Neyman–Pearson problem. However, due to the nonconvexity of Q, (α, y^*) does not minimize $\lambda x - y$ over Q, no matter how λ is chosen.

and γ_0^* is an optimal solution that happens to be deterministic. It follows from centralized theory that γ_0^* can be replaced by a deterministic monotone threshold rule. Q.E.D.

The no-point-mass assumption is indispensable in Proposition 3.6. Without it, one can construct examples in which (i) there exists a deterministic strategy that is optimal, (ii) there exists a monotone threshold strategy that is optimal, and (iii) no deterministic threshold strategy is optimal!

Proposition 3.5 provides a characterization of optimal strategies that is strikingly similar to the one provided by Proposition 2.4 for the Bayesian case. Despite that, finding an optimal solution of the Neyman–Pearson problem (over either the set Γ or $\overline{\Gamma}$) seems to be much more difficult. We discuss the drawbacks of two particular approaches that are implicit in the literature:

(i) We can associate a Lagrange multiplier λ to the constraint $J^F(\gamma) = \alpha$ and replace the constrained minimization problem (3.3) with the unconstrained problem of minimizing $\lambda J^F(\gamma) - J^D(\gamma)$ over the set Γ or $\overline{\Gamma}$. This is equivalent to minimizing $\lambda x - y$ over all $(x, y) \in Q$ or $\overline{Q}$, respectively. However, there is no guarantee that the sets Q and $\overline{Q}$, are convex[8] and an optimal solution to the Neyman–Pearson problem could fail to minimize the Lagrangian (see Figure 3.1).

(ii) As in the case of the Bayesian problem, we can try an algorithm based on the person-by-person optimality conditions, as in Eq. (2.16). For example, we have seen that γ_1^* must maximize the right-hand side of (3.11) subject to the right-hand side of (3.12) being equal to α. Unfortunately, this approach is inappropriate. To see this, suppose that γ is a nonoptimal threshold strategy such that $J^F(\gamma) = \alpha$. Furthermore, suppose that $D = 2$, so that each decision rule γ_i, $i \neq 0$, is determined by a single threshold t_i. We then see that optimizing t_i while keeping the thresholds of the other sensors fixed cannot lead to a new value for t_i because the constraint $J^F(\gamma) = \alpha$ would be violated.

3.4 The Case of Dependent Randomization

We now consider the Neyman–Pearson problem (3.3) when dependent randomization ($\gamma \in \Gamma^*$) is allowed. The following result shows that this problem is closely related to a Bayesian problem, for a suitable choice of the prior probabilities.

Proposition 3.7. Consider problem (3.3) under the constraint $\gamma \in \Gamma^*$ and suppose that an optimal solution exists. Then:

(a) A strategy γ^* is an optimal solution of the Neyman–Pearson problem (3.3) if and only if $J^F(\gamma^*) = \alpha$ and there exists some $\lambda \geq 0$ such that γ^* minimizes $\lambda J^F(\gamma) - J^D(\gamma)$ over all $\gamma \in \Gamma^*$.

(b) There exists some $p \in [0,1]$ and two monotone threshold strategies $\gamma^{(1)}, \gamma^{(2)} \in \overline{\Gamma}$ such that: (i) the randomized strategy $\gamma^* \in \Gamma^*$ that uses strategies $\gamma^{(1)}$ and $\gamma^{(2)}$ with probabilities p and $1 - p$, respectively, is optimal; (ii) the strategies $\gamma^{(1)}$ and $\gamma^{(2)}$ minimize $\lambda J^F(\gamma) - J^D(\gamma)$ over all $\gamma \in \Gamma^*$ for some $\lambda \geq 0$.

(c) If, in addition, $L_i(Y_i)$ has no point mass for every $i \neq 0$, then the strategies $\gamma^{(1)}, \gamma^{(2)}$ of part (b) can be chosen to be deterministic monotone threshold strategies.

Proof. (a) Notice that problem (3.3) is equivalent to maximizing q_2 subject to $q_1 = \alpha$ and $(q_1, q_2) \in Q^*$. Let $q_2^* \in [0,1]$ be the maximum of q_2 subject to these constraints. It is clear that (α, q_2^*) lies at the upper boundary of Q^*. [Otherwise, $(\alpha, q_2^* + \varepsilon)$ would also belong to Q^*, for sufficiently small $\varepsilon > 0$, which would contradict the definition of q_2^*.] By the supporting hyperplane theorem [31], and because Q^* is convex, there exist scalars λ_1, λ_2 such that (α, q_2^*) minimizes $\lambda_1 q_1 + \lambda_2 q_2$ over the set Q^*. Notice that the set Q^* is contained in the unit rectangle $[0,1]^2$ and that the points (0,0) and (1,1) belong to it. It follows (see Figure 3.2) that $\lambda_1 \geq 0$ and $\lambda_2 < 0$. By defining $\lambda = -\lambda_1/\lambda_2$ we see that (α, q_2^*) minimizes $\lambda q_1 - q_2$ over the set Q^*.

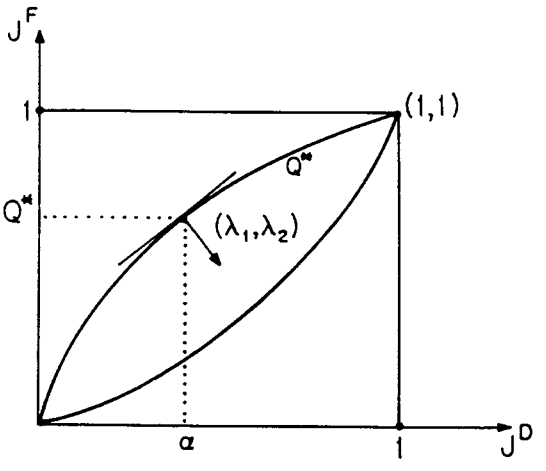

Figure 3.2.

Now, if $\gamma \in \Gamma^*$ is an optimal strategy, then $(J^F(\gamma^*),\ J^D(\gamma^*)) = (\alpha, q_2^*)$. Thus, γ^* minimizes $\lambda J^F(\gamma) - J^D(\gamma)$ over all $\gamma \in \Gamma^*$.

Conversely, if γ^* satisfies $J^F(\gamma^*) = \alpha$ and if there exists some $\lambda \geq 0$ such that γ^* minimizes $\lambda J^F(\gamma) - J^D(\gamma)$ over all $\gamma \in \Gamma^*$, then it is easily seen that $(J^F(\gamma),\ J^D(\gamma) = (\alpha, q_2^*)$, and therefore γ^* is optimal.

(b) Let q_2^* be defined as in part (a) of the proof and let $q^* = (\alpha, q_2^*)$. We have $q^* \in Q^*$, and Q^* is the convex hull of $\overline{Q}$ (Proposition 3.1). Thus, there exist some integer K and some nonnegative coefficients $p_1, \ldots, p_K$ that sum to 1, and some elements $q^1, \ldots, q^K$ of $\overline{Q}$ such that $\Sigma_{k=1}^K p_k q^k = q^*$. Let, as in the proof of part (a), $\lambda \geq 0$ be such that (α, q_2^*) minimizes $\lambda q_1 - q_2$ over the set Q^*. Let $\overline{\lambda} = (\lambda, -1)$, and let $\overline{\lambda}'$ be its transpose. We have

$$\overline{\lambda}' \sum_{k=1}^{K} p_k q^k = \overline{\lambda}' q^*. \tag{3.13}$$

On the other hand, the definition of λ implies that $\overline{\lambda}' q \geq \overline{\lambda}' q^*$ for all $q \in Q^*$. This, together with Eq. (3.13), implies that $\overline{\lambda}' q^k = \overline{\lambda}' q^*$, for all k. It follows that the vectors $q^*, q^1, \ldots q^K$ lie on a straight line. Consequently, q^* can be expressed as a convex combination of only two of the vectors q^k. We conclude that $q^* = pq^1 + (1-p)q^2$, for some $p \in [0,1]$ and some $q^1, q^2 \in \overline{Q}$. Furthermore, q^1 and q^2 minimize $\overline{\lambda}' q$ over all $q \in Q^*$.

For $i = 1,2$, let $\gamma^{(i)}$ be an element of $\overline{\Gamma}$ such that $q^i = (J^F(\gamma^{(i)}), (J^D(\gamma^{(i)}))$. (Such a $\gamma^{(i)}$ exists because $q^i \in \overline{Q}$. Since q^i minimizes $\lambda q_1 - q_2$ over all $(q_1, q_2) \in Q^*$, it follows that $\gamma^{(i)}$ minimizes $\lambda J^F(\gamma) - J^D(\gamma)$ over all $\gamma \in \Gamma^*$, thus establishing

property (ii) in the statement of the proposition.

Using part (a) of the proposition, $\gamma^{(i)}$ is an optimal solution of the Neyman–Pearson problem

$$\text{maximize} \quad J^D(\gamma) \qquad \text{subject} \quad J^F(\gamma) = J^F(\gamma^{(i)}) \text{ and } \gamma \in \Gamma^*. \tag{3.14}$$

A fortiori, the same is true if Γ^* is replaced by $\overline{\Gamma}$. Then, Proposition 3.5 applies and shows that $\gamma^{(i)}$ can be chosen to be a monotone threshold strategy.

We now let γ^* be the randomized strategy that uses the strategies $\gamma^{(1)}$ and $\gamma^{(2)}$ with probabilities p and $1 - p$, respectively. We then have $J^F(\gamma^*)$, $(J^D(\gamma^*)) = pq^1 + (1-p)q^2 = q^*$, and therefore γ^* is optimal.

(c) Suppose now that Y_i has no point mass for every $i \neq 0$. We argue exactly as in the proof of part (b), except that $\overline{Q}$ is replaced throughout by Q. Then, $\gamma^{(i)}$ is a deterministic strategy that is an optimal solution of problem (3.14). A fortiori, it is also an optimal strategy when the optimization problem (3.14) is considered over the set $\overline{\Gamma}$. Proposition 3.6 applies that and shows that there exists a deterministic monotone threshold strategy that attains the minimum in Eq. (3.14). Using this strategy in the place of $\gamma^{(i)}$, the desired result is established. Q.E.D.

It is easily seen that the problem of minimizing the Lagrangian $\lambda J^F(\gamma) - J^D(\gamma)$ is equivalent to the Bayesian problem of minimizing the error probability when $\lambda = \Pr(H_1)/\Pr(H_2)$. We conclude that the Neyman–Pearson problem is closely related to a Bayesian problem, for a suitable choice of the prior probabilities. In particular, if an optimal strategy $\gamma \in \overline{\Gamma}$ for the Bayesian problem happens to satisfy $J^F(\gamma) = \alpha$, then it is also an optimal strategy for the Neyman–Pearson problem. If on the other hand, no Bayesian optimal strategy $\gamma \in \overline{\Gamma}$ satisfies $J^F(\gamma) = \alpha$, then two strategies in $\overline{\Gamma}$ have to be combined to form an optimal strategy for the Neyman–Pearson problem. This latter case can only arise when the set Q is nonconvex.

Computing an optimal strategy for the Neyman–Pearson problem with dependent randomization seems particularly difficult if $J^F(\gamma) \neq \alpha$ for all optimal strategies $\gamma \in \overline{\Gamma}$ of the Bayesian problem. It seems that a search over the possible values of λ is needed until a λ with the property stated in Proposition 3.7 is found. Then, two optimal solutions of the corresponding Bayesian problem have to be found and combined in order to satisfy the constraint $J^F(\gamma) = \alpha$. In fact, scanning the possible values of λ is the most natural method for generating the upper boundary of the set Q^*.

NOTES. Decentralized Neyman–Pearson detection was first considered, somewhat informally, in [11]. Subsequently, it was studied in [8,25,48,52,53,68]. In several of these papers, it is established that an optimal strategy over the set $\overline{\Gamma}$ is a monotone threshold strategy, by starting from the assumption that the Neyman–Pearson problem is related to the minimization of $\lambda J^F(\gamma) - J^D(\gamma)$. However, this argument may run into the difficulties illustrated in Figure 3.1, and is therefore unjustified. A different (and correct) proof of Proposition 3.5 is given in [69]

for the case $D = 2$ and in [71] for the case of general D. Other proofs are provided in [72] and in [61]. Propositions 3.6 and 3.7 are new. Joint randomization was discussed in [48] where it was observed that it leads to the convexification of the team-ROC. The role of randomization is discussed further in [73].

4. MORE COMPLEX CONFIGURATIONS

In this section, more complex sensor configurations are considered. Under the assumption that the sensors form a tree, it will be seen (Sections 4.2–4.3) that the qualitative conclusions of Sections 2 and 3 remain valid. This is not the case, however, when more general configurations are considered (Section 4.4). Finally, Section 4.5 compares some alternative configurations. Throughout, the conditional independence assumption (Assumption 2.1) will be in effect.

4.1 Tree Configurations

In a tree configuration, a set of sensors $S_0, S_1, \ldots, S_N$ are connected so as to form a tree (that is, a connected graph with no cycles). Sensor S_0 will be called again the fusion center. Since the graph of sensors forms a tree, there exists a unique path from every sensor S_i, $i \neq 0$, to sensor S_0. We orient the arcs of the tree so that they all point toward the fusion center (see Figure 4.1). Thus, the sensors can be viewed as a directed graph with nodes $S_0, S_1, \ldots, S_N$ and a set A of directed arcs as just described.

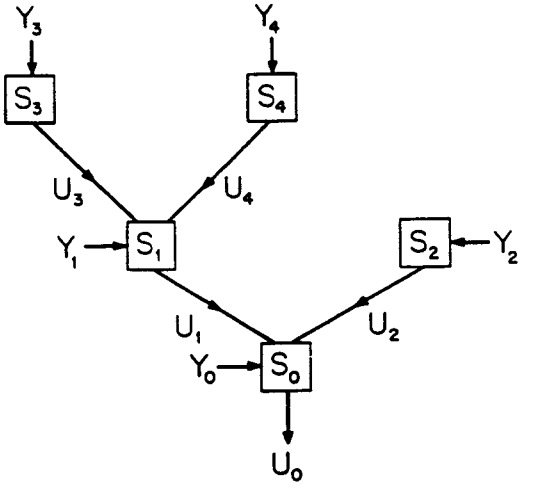

Figure 4.1. A tree configuration. Here $P(1) = \{3,4\}$ and $P(2)$ is empty. Also, $I(1) = \{0,2\}$ and $I(2) = \{0,1,3,4\}$.

We define some terminology. We say that S_i is an *immediate predecessor* (respectively, *immediate successor*) of S_j if $(S_i, S_j) \in A$ [respectively, if $(S_i, S_j) \in A$]. We say that S_i is a *predecessor* (respectively, *successor*) of S_j if there exist $S_{i(1)}, \ldots, S_{i(k)}$ such that $(S_{i(l)}, S_{i(l+1)}) \in A$ for $l = 1, \ldots k-1$, and $S_{i(1)} = i$, $S_{i(k)} = j$ (respectively, $S_{i(1)} = j$, $S_{i(k)} = i$). Let $P(i)$ be the set of all immediate predecessors of i. Finally, let $I(i)$ be the set of all sensors S_j such that $j \neq i$ and S_j is not a predecessor of S_i. (See Figure 4.1 for an illustration.)

Each sensor receives an observation Y_i, which is a random variable taking values in a set $\mathcal{Y}_i$, and with known probability distribution, conditioned on each one of a set $H_1, \ldots, H_M$ of hypotheses. The organization works as follows: Each sensor S_i with no predecessors (that is, every "leaf") sends a message to its immediate successor. Then any sensor S_i, upon receiving a message from all of its immediate predecessors, computes its own message, as a function of the messages received and of its own observation Y_i. Finally, the fusion center S_0, upon receiving a message from all of its immediate predecessors, makes a final decision, as a function of its own observation Y_0 and of the received messages. We assume that the message of each sensor S_i, $i \neq 0$, takes values in the set $\{1, \ldots, D\}$, and that the final decision of the fusion center takes values in the set $\{1, \ldots, M\}$.

More formally, let $|P(i)|$ be the cardinality of the set of immediate predecessors of sensor S_i. A decision rule for sensor S_i, $i \neq 0$, is a function $\gamma_i: \mathcal{Y}_i \times \{1, \ldots, D\}^{|P(i)|} \to \{1, \ldots, D\}$. Similarly, a decision rule for sensor S_0 is a function $\gamma_0: \mathcal{Y}_0 \times \{1, \ldots, D\}^{|P(0)|} \to \{1, \ldots, M\}$. Let Γ_i be the set of all (deterministic) decision rules for sensor S_i and let $\Gamma = \prod_{i=0}^{N} \Gamma_i$ be the set of all (deterministic) strategies.

Once a strategy $\gamma \in \Gamma$ is fixed, the message (or decision) of sensor S_i becomes a well-defined random variable to be denoted by U_i. (Clearly, the distribution of U_i depends on γ.) Formally, we have

$$U_i = \gamma_i(Y_i, U^i), \tag{4.1}$$

where U^i is defined as the vector whose components are the random variables $U_j, j \in P(i)$.

4.2 Bayesian Formulation

We consider a tree configuration and we assume that each hypothesis H_j has a prior probability $\Pr(H_j) > 0$. For simplicity, we also assume that the cost criterion is the fusion center's probability of error. That is, for any strategy $\gamma \in \Gamma$, we let

$$J(\gamma) = \sum_{j=1}^{M} \Pr(U_0 \neq j \mid H_j) \Pr(H_j),$$

where U_0 has the probability distribution determined by γ. We are interested in finding a strategy that minimizes $J(\gamma)$ over the set Γ. (Randomization is unnecessary, for the same reasons as in Section 2; see Proposition 2.1).

As it turns out, the development and the results are exactly parallel to the results of Section 2. We will therefore just sketch an outline of the arguments involved. Let us focus on a particular sensor S_i, and suppose that a decision rule γ_j^* has been fixed for every other sensor S_j, $j \neq i$. We are interested in a decision rule γ_i^* that minimizes $J(\gamma_0^*, \ldots, \gamma_{i-1}^*, \gamma_i, \gamma_{i+1}^*, \ldots, \gamma_N^*)$ over all $\gamma_i \in \Gamma_i$. Let Z^i be the vector whose components are the random variables Y_j, with $j \in I(i)$. [For example, in Figure 4.1, we have $Z^2 = (Y_0, Y_1, Y_3, Y_4)$]. It is easily seen that the decision of the fusion center can be expressed as a function of the message of sensor S_i and the vector Z^i. In other words, for every strategy γ, there exists a function δ_i such that

$$U_0 = \delta_i(U_i, Z^i) \tag{4.2}$$

Furthermore, the functional form of the function δ_i does not depend on γ_i. Using Eq. (4.1), we have

$$U_0 = \delta_i(\gamma_i(Y_i, U^i), Z^i).$$

Let us define $C(d, H_j)$ to be 1 if $d \neq j$ and zero otherwise. Then

$$J(\gamma) = E[C(U_0, H)] = E[C(\delta_i(\gamma_i(Y_i, U^i), Z^i), H)].$$

This expression is to be minimized with respect to γ_i. Since δ_i is independent of γ_i, and since the distribution of U^i is also independent of γ_i, it is easily seen that this minimization is of the form considered in Proposition 2.2, provided that we identify X with (Y_i, U^i), Z with Z^i, and $F(d, Z, H_j)$ with $C(\delta(d, Z^i), H_j)$. Proposition 2.2 uses the assumption that X and U are conditionally independent. In our context we need (Y_i, U^i) to be conditionally independent from Z^i. This condition is satisfied because (Y_i, U^i) and Z^i are functions of conditionally independent sets of random variables Y_j. Therefore, Proposition 2.2 applies and shows the following:

Proposition 4.1. Let Assumption 2.1 hold. Let $i \neq 0$. Suppose that γ_j has been fixed for every $j \neq i$. Then γ_i minimizes $J(\gamma)$ over all $\gamma_i \in \Gamma_i$ if and only if

$$\gamma_i(Y_i, U^i) = \arg \min_{d=1,\ldots,D} \sum_{j=1}^{M} \Pr(H_j \mid Y_i, U^i) a_i(H_j, d), \quad \text{w.p.1,} \tag{4.3}$$

where

$$a_i(H_j, d) = E[C(\delta_i(d, Z^i), H_j) \mid H_j],$$

and δ_i is defined by Eq. (4.2).

Any optimal strategy must be person-by-person optimal and the decision rule of each sensor S_i, $i \neq 0$, must satisfy Eq. (4.3). Furthermore, for the fusion center, it is obvious that the MAP rule should be used.

Notice that the structure of decision rules prescribed by Eq. (4.3) is identical to the one obtained for the simpler configuration studied in Section 2. Thus, most of the discussion in Section 2 still applies. For example, we obtain the following result, whose proof is identical to the proof of Proposition 2.4.

Proposition 4.2. Suppose that there are two hypotheses ($M = 2$), that Assumption 2.1 holds, and that there exists an optimal strategy. Then there exists an optimal strategy in which $\gamma_i(Y_i, U^i)$ is nondecreasing in $L_i(Y_i)$ and U^i, for each i.

Regarding the computation of an optimal strategy, the discussion in Section 2.4 still applies. In particular, a most natural algorithm is a successive person-by-person optimization as in Eq. (2.16).

4.3 Neyman–Pearson Formulation

In this subsection, we assume that $M = 2$ (binary hypotheses) and that we are still dealing with a tree configuration and conditionally independent observations. We keep the same notation as in Section 4.2. Similarly to Section 3, we also consider randomized strategies (with dependent or independent randomization).

Let $J^D(\gamma) = \Pr(U_0 = 2 \mid H_2)$ and $J^F(\gamma) = \Pr(U_0 = 2 \mid H_1)$. We consider the problem of maximizing $J^D(\gamma)$ subject to $J^F(\gamma) = \alpha$, where α is a given scalar in (0,1).

We consider the case of independent randomization, and we argue as in Proposition 3.5. We fix the decision rules of all sensors other than sensor S_i, $i \neq 0$, and consider the optimization with respect to γ_i. Letting δ_i be the function introduced in Eq. (4.2), and using the conditional independence assumption, we have

$$J^D(\gamma) = \sum_{d=1}^{D} \Pr(\delta_i(d, Z^i) = 2 \mid H_2) \Pr(\gamma_i(Y_i, U^i) = d \mid H_2)$$

$$= \sum_{d=1}^{D} c_d \Pr(\gamma_i(Y_i, U^i) = d \mid H_2).$$

Similarly

$$J^F(\gamma) = \sum_{d=1}^{D} h_d \Pr(\gamma_i(Y_i, U^i) = d \mid H_1)$$

for some suitable coefficients h_d. Thus, the problem facing sensor S_i, has the same form as problem (3.9)–(3.10), provided that we identify Y_i with (Y_i, U^i). We can therefore repeat the argument in the proof of Proposition 3.5 and obtain the following:

Proposition 4.3. Let Assumption 2.1 hold. Suppose that the Neyman–Pearson problem over the set of strategies with independent randomization has an optimal solution. Then there exists an optimal strategy in which $\gamma_i(Y_i, U^i)$ is nondecreasing in $L_i(Y_i)$ and U^i, for each i.

The Neyman–Pearson problem can be also considered when dependent randomization is allowed. Proposition 3.7 remains valid (with a verbatim repetition of the proof). The computation of Neyman–Pearson optimal strategies runs into all of the difficulties discussed in Section 3, and can be expected to be demanding.

4.4 Arbitrary Configurations

There is very little that can be said when one considers configurations more general than trees. In particular, threshold rules are no longer optimal. We illustrate this by a simple example.

Consider the configuration of Figure 4.2. Suppose that sensors S_0 and S_3 are the only ones that receive some information from the environment. Furthermore, suppose that $\mathcal{Y}_3 = \{1,2,3,4\}$ and that $D = M = 2$. Suppose that sensor S_3 sends messages as follows:

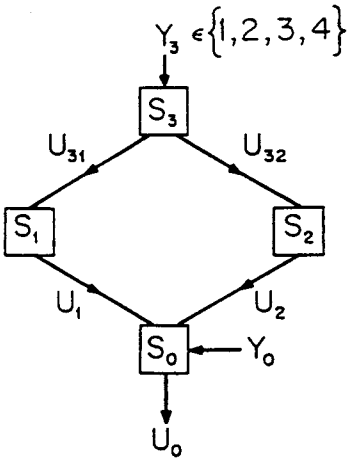

Figure 4.2.

$$U_{31}(Y_3) = 1 \qquad \text{if and only if } Y_3 \in \{1, 2\},$$

$$U_{32}(Y_3) = 1 \qquad \text{if and only if } Y_3 \in \{1, 3\}.$$

The sensors S_1 and S_2 can forward the messages they have received to the fusion center. Then, the fusion center is able to uniquely identify the value of Y_3. Suppose now that the likelihood ratio $L_3(Y_3)$ is an increasing function of Y_3. It is easily seen that if the messages U_{31} and U_{32} are chosen by comparing $L_3(Y_3)$ to some thresholds, the fusion center will not be always able to reconstruct the value of Y_3. Thus, the restriction to threshold rules leads to a reduction in the information available to the fusion center. Then, the unspecified parts of the example (the distribution of Y_0) can be easily chosen so that the loss of information leads to loss of performance.

4.5 Comparison of Alternative Configurations

Suppose that we have a set of sensors each one receiving its own information Y_i, but that we have not yet decided on the configuration of the sensors. We can then pose the problem of choosing a configuration under which the overall performance (under an optimal strategy) is best. Unfortunately, very little progress can be made on this problem because the overall performance (under an optimal strategy) of different configurations cannot be calculated in closed form. We discuss below the few available results on this problem.

The first result is a simple observation and states that the "tandem" configuration of Figure 4.3(a) is always at least as good as the "parallel" configuration of Figure 4.3(b), assuming that the fusion center in the parallel configuration has no information of its own. The reason is that in the parallel configuration, the final decision is given by

$$U_0 = \gamma_0(\gamma_1(Y_1), \gamma_2(Y_2)). \tag{4.4}$$

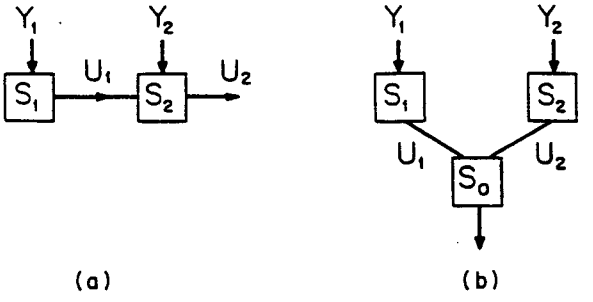

Figure 4.3.

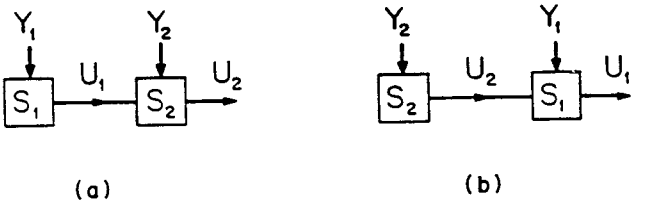

Figure 4.4.

Such a decision can also be implemented in the tandem configuration, by having sensor S_1 transmit the value of $\gamma_1(Y_1)$ to sensor S_2. Then, sensor S_2 could compute U_0 according to Eq. (4.4), thereby attaining the performance level of the parallel configuration.

The above observation motivates the next question: Is it better to have a tandem configuration of N sensors or a parallel configuration with N peripheral sensors? We saw that the tandem configuration is better when $N = 2$. For moderate values of N, there seems to be no unequivocal answer. In Section 5.2, it will be seen that the parallel configuration is typically better when N is large.

The last case we consider compares the two configurations shown in Figure 4.4. Here, a tandem configuration is assumed, but the two sensors are not identical, and the problem is to decide which one of the two should play the role of the fusion center. We put some more structure into the problem by assuming that one of the two sensors, say S_1, is better than the other in the following sense. The *ROC-curve* of sensor S_i is a function of a scalar parameter $\alpha \in [0,1]$ defined by

$$F_i(\alpha) = \max_{D \subset \mathcal{Y}_i} \Pr(D \mid H_2) \qquad \text{subject to } \Pr(D \mid H_1) = \alpha.$$

Let us say that S_1 is *better* than S_2 if $F_1(\alpha) \geq F_2(\alpha)$, for all $\alpha \in [0,1]$. It was conjectured in [16] that when S_1 is better, then S_1 should be the fusion center [as in Figure 4.4(b)] for all choices of the prior probabilities. This conjecture is false. In particular, [40] provides an example where S_1 is better than S_2 but each one of the two candidate configurations is preferable for a particular range of the prior probabilities.

NOTES. The study of tree configurations was initiated in [16,17] where the optimality of monotone threshold rules was established for the Bayesian formulation, and an iterative algorithm resembling Eq. (2.16) was suggested. The techniques in these references readily generalize to yield Propositions 4.1 and 4.2. The optimality conditions of Proposition 4.1 are elaborated further in [48,50], for the case $M = D = 2$, and several special cases are studied in detail. A tandem of two sensors for general values of D is studied in [41]. The Neyman–Pearson problem for a tandem configuration was studied in [70], where the optimality of threshold strategies is established for $D = 2$. This reference also contains the observation regarding the comparison of the configurations shown in Figure 4.3. Proposition 4.3 is new. The comparison of the configurations of Figure 4.4 is studied in [38,40], where the conjecture of [16] is disproved (in general) and is established for certain Gaussian problems.

5. ASYMPTOTIC CONSIDERATIONS

In previous sections, we derived several optimality conditions that can be useful for computing optimal solutions to decentralized detection problems. Unfortunately, these conditions do not lead to closed-form solutions and virtually no analytical results (other than general optimality conditions) are available. In this section, we show that some further progress is possible when one considers asymptotic solutions, as the number of sensors becomes very large. In Section 5.1, we consider a parallel configuration of N identical sensors. A fairly complete solution is available here, as N tends to infinity. In particular, if the number of hypotheses is equal to two, we show that we can restrict the sensors to use the same decision rule, with negligible loss of performance. In Section 5.2, we consider a tandem configuration of N sensors and we examine the probability of error as N tends to infinity. It will be seen that the tandem configuration is generally inferior to the parallel configuration when N is large.

5.1 Parallel Configurations

We consider here the Bayesian problem of Section 2, under the probability of error criterion. We assume that there are two hypotheses ($M = 2$) and that the observations $Y_1, \ldots, Y_N$ of the peripheral sensors are conditionally independent and identically distributed. Let Y be a random variable taking values in a set $\mathcal{Y}$ that has the same conditional distributions as the random variables Y_i $i \neq 0$. We let Δ be the set of all functions (deterministic decision rules) δ: $\mathcal{Y} \rightarrow \{1, \ldots, D\}$. We assume that the fusion center receives no information of its own.

We start with the observation that the probability of error drops quickly as N increases.

Proposition 5.1. Suppose that Assumption 2.1 holds and that the two conditional distributions of Y are not identical. Then there exists a decision rule $\delta \in \Delta$ and constants $\alpha, \beta > 0$ (independent of N) such that, if all sensors use δ and the fusion center uses the MAP rule, then the probability of error is less than $\alpha e^{-\beta N}$.

Proof. Choose a set $A \subset \mathcal{Y}$ such that $\Pr(A \mid H_1) \neq \Pr(A \mid H_2)$, and let $\delta(Y) = 1$ if $Y \in A$, and $\delta(Y) = 2$ otherwise. Then, the distributions of $\delta(Y)$ under the two hypotheses are different. The fusion center is faced with a hypothesis testing problem, and the data received by the fusion center are a sequence of conditionally independent, identically distributed, Bernoulli random variables. Then, the exponential decrease of the error probability follows from standard results [9]. Q.E.D.

Because of Proposition 5.1, we focus on the error exponent $r(\gamma)$ defined by $r(\gamma) = \log(J(\gamma)/N)$, where $J(\gamma)$ is the error probability associated to strategy γ. We define

the optimal exponent r_N as the infimum of $r(\gamma)$ over all strategies γ for the N-sensor problem. Let us now consider those strategies in which all peripheral sensors use the same decision rule. We define r_N^S as the infimum of $r(\gamma)$ over all such strategies. Clearly, $r_N \leq r_N^S$ for every N. Our next result states that the reverse inequality is also valid, asymptotically.

For any decision rule $\delta \in \Delta$ and any scalar $s \in [0,1]$, we define (cf. [9])

$$\mu(\delta, s) = \log \left[\sum_{d=1}^{D} \Pr(\delta(Y) = d \mid H_1)^{1-s} \Pr(\delta(Y) = d \mid H_2)^{s} \right]. \tag{5.1}$$

(The convention $0^0 = 0$ is used here.)

Assumption 5.1. (a) $|\mu(\delta, s)| < \infty$ for all $\delta \in \Delta$ and $s \in [0,1]$.

(b) For $j = 1,2$, there holds $E[\log^2 L(Y) \mid H_j] < \infty$, where $L(Y)$ is the likelihood ratio of the two conditional distributions of Y.

Proposition 5.2. Under Assumptions 2.1 and 5.1, we have

$$\lim_{N\to\infty} r_N = \lim_{N\to\infty} r_N^S = \inf_{\delta\in\Delta} \inf_{s\in[0,1]} \mu(\delta, s). \tag{5.2}$$

Proof. (Outline). Using certain error bounds from [54] (see [59] for details), we have

$$r(\gamma) = \frac{1}{N} \inf_{s\in[0,1]} \sum_{i=1}^{N} \mu(\gamma_i, s) + o(N), \tag{5.3}$$

for every strategy $\gamma = (\gamma_0, \gamma_1, \ldots, \gamma_N)$ in which γ_0 is the MAP rule. Here, $o(N)$ is a term that converges to zero, as $N \to \infty$, uniformly over all strategies γ. It follows that

$$\lim_{N\to\infty} r_N \geq \inf_{\delta\in\Delta} \inf_{s\in[0,1]} \mu(\delta, s).$$

On the other hand, by constraining each γ_i in Eq. (5.3) to be the same, we obtain

$$\lim_{N\to\infty} r_N^S = \inf_{\delta\in\Delta} \inf_{s\in[0,1]} \mu(\delta, s).$$

These two inequalities, together with $r_N \leq r_N^S$, establish the desired result.

Q.E.D.

Proposition 5.2 states that we can constrain all peripheral sensors to use the same decision rule, without worsening the error exponent. Furthermore, an asymptotically optimal decision rule is found by performing the minimization in Eq. (5.2). This may seem difficult, but a moment's thought shows that only monotone threshold rules need to be considered, as is shown below.

Corollary 5.1. Under Assumption 5.1, there holds

$$\inf_{\delta\in\Delta}\ \inf_{s\in[0,1]}\ \mu(\delta, s). = \inf_{\delta\in\Delta_T}\ \inf_{s\in[0,1]}\ \mu(\delta\, s), \tag{5.4}$$

where Δ_T is the subset of Δ consisting of deterministic monotone threshold rules.

Proof. The left-hand side in Eq. (5.4) is the optimal error exponent, by Proposition 5.1. The right-hand side is the optimal error exponent, when we restrict to monotone threshold rules. By Proposition 2.4, the restriction to deterministic monotone threshold does not increase the optimal probability of error and therefore does not increase the value of the optimal error exponent.

Q.E.D.

In the special case where $D = 2$, monotone threshold rules can be parametrized by a single threshold. In that case, the minimization in Eq. (5.2) reduces to a one-dimensional search over the set of all thresholds. Computational experience [36] suggests that this one-dimensional search is not computationally demanding and that the resulting optimal threshold leads to very good performance (compared to the truly optimal) even when the number N is moderately small (e.g., $N \approx 10$).

5.1.1 *The Neyman–Pearson Variant*

Similar results are available for the Neyman–Pearson variant of the problem whereby we wish to maximize $J^D(\gamma)$ subject to $J^F(\gamma) = \alpha$. Here, we keep α fixed, but we let N converge to infinity. If all peripheral sensors are using the same decision rule $\delta \in \Delta$, then, as $N \to \infty$, $J^D(\gamma)$ converges to one exponentially fast and

$$\lim_{N\to\infty} \frac{1}{N} \log[1 - J^D(\gamma)] = -K(\delta),$$

where $K(\delta)$ is the Kullback–Leibler information distance between the distributions of the random variable $\delta(Y)$ under the two hypotheses. Thus, finding an asymptotically optimal strategy, subject to every sensor using the same decision rule, is equivalent to choosing δ so as to maximize $K(\delta)$. As in Proposition 5.2, it can be shown [59] (under a minor technical assumption) that the restriction to identical decision rules does not affect the exponent that governs the convergence rate of

$J^D(\gamma)$ to 1. Finally, Corollary 5.1 has the following natural counterpart stating that (under certain technical conditions)

$$\inf_{\delta \in \Delta} K(\delta) = \inf_{\delta \in \Delta_T} K(\delta), \tag{5.5}$$

where Δ_T is the subset of Δ consisting of deterministic monotone threshold rules.

5.1.2 *The M-ary Case*

We have just seen that for the case of binary hypotheses, we can constrain all sensors to use the same decision rule without worsening the asymptotic error exponent. When $M > 2$, this is no more true. Nevertheless, it can be proved [59] that we may restrict the sensors to use at most $M(M-1)/2$ distinct decision rules. (Notice that for $M = 2$, this recovers Proposition 5.2.) Determining which particular $M(M-1)/2$ decision rules should be employed, and how many sensors should use each one of these decision rules, can be formulated as an optimization problem related to (but much more difficult than) the one in Eq. (5.2) [59].

5.2 Tandem Configurations

We consider here a tandem of N sensors, and we are interested in the error probability of the last sensor, for the case where N is large. We assume again that the observations of each sensor are conditionally independent and have the same distribution as a common random variable Y taking values in a set $\mathcal{Y}$. Let us say that the asymptotic probability of error is *bounded away from zero* if for any given positive prior probabilities, there exists a scalar $\alpha > 0$ such that for every number N of sensors in the tandem and for every strategy, the probability of error by the last sensor is at least α. In the contrary case, we say that the asymptotic probability of error is zero.

Assuming that the sensors are numbered consecutively, with S_1 being the first and S_N the last, and using our standard notation, the system is described by $U_1 = \gamma_1(Y_1)$ and

$$U_{i+1} = \gamma_i(Y_i, U_i), \qquad i \geq 1. \tag{5.6}$$

Furthermore, we have the constraint that $U_i \in \{1, \ldots, D\}$, for $i < N$, and $U_N \in \{1, \ldots, M\}$. The probability of error of decisions generated according to Eq. (5.6) has been extensively studied in the context of "finite memory hypothesis testing." The following is a representative result.

Proposition 5.3. Let Assumption 2.1 hold and assume that $M = D = 2$. Then, the asymptotic probability of error is bounded away from zero if and only if there exists some $B > 0$ such that

$$\frac{1}{B} \leq L(Y) \leq B, \qquad \text{w.p.1,} \tag{5.7}$$

under either hypothesis.

The fact that the error probability is bounded away from zero, when $M = D = 2$ and Eq. (5.7) holds, is in sharp contrast to the exponential decrease of the error probability for the parallel configuration. In this case, the superiority of the parallel configuration is clear for large N. We actually suspect that the value of N at which the parallel configuration becomes better than the tandem configuration is typically very small. For many interesting cases (e.g., when the two conditional distributions of Y are Gaussian or exponential), Eq. (5.7) fails to hold, and therefore the probability of error can be driven to zero, as $N \to \infty$. However, the convergence of the error probability to zero is, in general, much slower than exponential, at least for the strategies suggested in [10,27]. Thus, the parallel configuration seems to be generically preferable.

NOTES. The material of Section 5.1, with the exception of Corollary 5.1, is from [59]. An alternative derivation of some similar results (but in different form) for the case of M hypotheses has been obtained later in [1]. The fact that an optimal decision rule is a minimizer in Eq. (5.2) is essentially stated in [9], provided that the constraint $\gamma_i = \gamma$, for all i, is in effect. Thus, the novelty in Proposition 5.1 is the verification that this constraint is harmless. Corollary 5.1 is proved in [28] using a direct but more tedious argument. We suspect that the validity of Eq. (5.5) must have been known but we are not aware of any earlier reference. Reference [6] provides an asymptotic solution to the problem of minimizing $\mu(\delta, s)$, as $D \to \infty$. Concerning the M-ary case, an explicit asymptotic solution for the parallel configuration is provided in [46] for a simple special case.

The asymptotic results of Section 5.1 have been generalized and unified in [72] and [61], where it is shown that threshold rules optimize a large class of so-called Ali–Silvey distance measures. Earlier and less general results of this type can be found in [18, 29, 42]. Reference [61] also contains some generalizations to the case of M hypotheses.

In related work, reference [56] studies the effects of feedback on the error exponent of Proposition 5.2. Also, references [12,13,19,20] provide asymptotic analyses for a variety of problems, including robust detection and the case of weak signals.

Concerning Proposition 5.3, the "only if" direction was established in [10] and [23]. The "if" direction was established in [23], under the additional constraint that γ_i is the same for $i = 2, \ldots, N-1$. Without this additional constraint, a proof is contained in [38] and [41]. It also seems that a proof in [27] can be adapted to yield the same result. The reference [27] also shows that for M hypotheses having a particular structure, the asymptotic error probability is equal to zero if and only if $D \geq M + 1$. See also [33] for a related early reference. Finally, [38] contains further analysis of some suboptimal schemes.

The parallel and tandem configurations have been compared in [48] and [70]. The first reference provides extensive numerical results that corroborate the subexponential decrease of the error probability in a tandem configuration. The second provides a condition under which the tandem configuration is superior for a finite value of N, but this condition is hard to check without first finding an optimal strategy for the parallel configuration.

References [2] and [22] study decentralized hypothesis testing from another asymptotic viewpoint. Namely, they consider the case where each sensor has access to a very long block of data; in this case block-coding techniques are applicable. A complete solution is given in [22] for the case $D = 2$ and a parallel configuration with two peripheral sensors or a tandem of two sensors. Interestingly enough, conditional independence is not assumed in the model of [2] and [22].

6. SEQUENTIAL PROBLEMS

In this section, we present a brief review of decentralized sequential detection problems. In order to keep this section short, we only provide a rough description of the problems and the corresponding results. Detailed expositions can be found in the references provided.

The key aspect of sequential detection problems is that the sensors have the options to stop and decide in favor of some hypothesis, or to receive more information. Decentralized variants can be classified into two main categories, depending on whether stopping decisions are the responsibility of the fusion center or of the peripheral sensors. We start by considering the case where the fusion center makes the stopping decisions.

Let there be two hypotheses, H_1 and H_2, with known prior probabilities, and N peripheral sensors $S_1, \ldots, S_N$. At each time $t \in \{1, 2, \ldots\}$ each sensor S_i observes a random variable $Y_{i,t}$ and sends a message $U_{i,t} \in \{1, \ldots, D\}$ to a fusion center. We assume that the random variables $Y_{i,t}$ are conditionally independent. At each time t, the fusion center receives the messages $U_{i,t}$, $i = 1, \ldots, N$, together with some additional information of its own, and has three options: decide in favor of H_1, in favor of H_2, or defer the decision to a later stage. The objective is to choose the decision rules of the peripheral sensors and of the fusion center so as to minimize the probability of an incorrect decision by the fusion center plus the total cost of deferring the decision (the latter being equal to a constant multiple of the time at which the decision is made).

Once the decision rules of the peripheral sensors are fixed, the fusion center is faced with a classical (centralized) sequential detection problem and must therefore use a *sequential probability ratio test*. Namely [5], at each time t, the fusion center forms a likelihood ratio L_t (as a function of all information it has accumulated) and compares it to two thresholds T_t and T_t'. If $L_t \leq T_t$, then H_1 is chosen; if $L_t \geq T_t'$, then H_2 is chosen; if $T_t < L_t < T_t'$, then the decision is deferred.

Let us now consider the peripheral sensors. There are two different cases to be considered depending on whether $U_{i,t}$ is allowed to depend only on $Y_{i,t}$ or on past $Y_{i,\tau}$, $\tau < t$, as well.

CASE A. Suppose that $U_{i,t}$ is constrained to be of the form

$$U_{i,t} = \gamma_{i,t}(Y_{i,t}).$$

Then, the problem faced by the peripheral sensors is identical in structure to the problem studied in Section 2. (Just imagine that each $Y_{i,t}$ is observed by a different sensor $S_{i,t}$.) Thus, the optimality of threshold rules follows, as in Proposition 2.4. (An interesting special case, with a single peripheral sensor and a time horizon of two time units, has been studied in detail in [39].)

CASE B. Suppose that $U_{i,t}$ is allowed to be of the more general form

$$U_{i,t} = \gamma_{i,t}(Y_{i,1}, Y_{i,2}, \ldots, Y_{i,t}).$$

Clearly, this is a more natural formulation than the one in Case A. Unfortunately, there exist simple counterexamples (see [58]) in which the only optimal solutions are not threshold rules (in disagreement with the results claimed in [24]). Furthermore, we are not aware of any positive results on this problem.

We now turn to the case where a fusion center is absent and the stopping decisions are the responsibility of the peripheral sensors. In particular, suppose that each sensor chooses a stopping time of its own and, upon stopping, decides in favor of one of the two hypotheses. Furthermore, let us allow the decision of a sensor to depend on all accumulated observations. The cost function is a positive linear combination of the stopping times of the sensors plus a term of the form $C(u_1, \ldots, u_N, H)$, where u_i is the final decision of sensor S_i. This problem has been studied in [63] where is was established that optimal decision rules for each sensor can be found within the class of sequential probability ratio tests. The reason underlying this result is that once the decision rules of all sensors S_j, $j \neq i$, are fixed, then sensor S_i is faced with a classical sequential detection problem. This point is clarified in [32], where the continuous time variant of the problem is solved.

References [3,4] deal with related problems. An interesting variation is studied in [67]. Here, two detectors wish to detect the time when a Markov chain changes state. The cost function consists of a penalty for the time between the state change and its detection, together with a penalty for detecting a state change before it actually occurs. As it turns out, optimal strategies have a certain threshold property [67]. A threshold property is also established in a signaling problem studied in [60].

ACKNOWLEDGMENTS

The author is grateful to Professors Michael Athans and Jason Papastavrou for several discussions on this subject over a number of years.

NOTES

1. Research supported by the ONR under Contract N00014-84-K-0519 (NR 649-003) and by the ARO under contract DAAL03-86-K-0171.

2. Throughout, we use uppercase letters to denote random variables and lowercase letters to indicate particular values in the range of a random variable. The only exception to this rule is that we use H_j (instead of h_j) to indicate the jth hypothesis, in order to conform to customary usage.

3. We will occasionally consider the case where the fusion center has no observation of its own. This amounts to omitting Y_0 from the model or, alternatively, letting Y_0 be a constant random variable.

4. In a slightly more general formulation, we could assume that each u_i takes values in a set $\{1, \ldots, D_i\}$, where the D_i are generally different positive integers. This added generality is avoided in order to simplify notation.

5. In order to simplify the exposition, we will not address the measure-theoretic aspects of the problem, even though some attention to them is needed for a fully rigorous treatment. The reader familiar with measure theory should have no difficulty in supplying the missing details. To mention one example, it should have been assumed that each set $\mathcal{Y}_i$ is endowed with a σ-field $\mathcal{F}_i$, and we should have defined Γ_i, $i \neq 0$, as the set of all $\mathcal{F}_i$-measurable functions from $\mathcal{Y}_i$ into $\{1, \ldots, D\}$. A similar measurability condition is needed for Γ_0. Let us simply state here that the results to be derived remain unchanged when measurability issues are properly taken into account.

6. All the formulas to be derived are also true under the alternative assumption that each set $\mathcal{Y}_i$ is finite, provided that the probability density functions are interpreted as probability mass functions. These formulas also generalize to the case of arbitrary probability distributions, provided that the probability density functions are replaced by Radon–Nikodym derivatives with respect to a suitable reference measure.

7. Formally, let $A \subset \mathcal{Y}_i$ be such that $\Pr(A \mid H_1) = 1$ and such that $\mathcal{P}_i|_{H_2}$ is absolutely continuous with respect to $\mathcal{P}_i|_{H_1}$, on the set A. We then let $L_i(y_i) = \infty$ for $y_i \notin A$. Also, for $y_i \in A$, $L_i(y_i)$ is the Radon–Nikodym derivative of $\mathcal{P}_i|_{H_2}$, with respect to $\mathcal{P}_i|_{H_1}$.

8. An example in which Q is nonconvex can be found in [48, p. 55]. It is also possible to construct examples for which $\overline{Q}$ is nonconvex.

REFERENCES

[1] Anantharam, V., "Distributed Detection with a Large Number of Sensors," preprint, 1988.

[2] Ahlswede, R., and I. Csiszar, "Hypothesis Testing with Communication Constraints," *IEEE Trans. Inform. Theory*, **IT-32**, 1986, pp. 533–542.

[3] Al-Ibrahim, M. M., and P. K. Varshney, "A Simple Multi-Sensor Sequential Detection Procedure," *Proc. 27th IEEE Conf. Decision Control, Austin, Texas, 1989.*

[4] Al-Ibrahim, M. M., and P. K. Varshney, "A Decentralized Sequential Test with Data Fusion," *Proc. 1989 Amer. Control Conf., Pittsburgh, 1989.*

[5] Bertsekas, D. P., *Dynamic Programming: Deterministic and Stochastic Models*, Prentice-Hall, Englewood Cliffs, NJ, 1987.

[6] Benitz, G. R., and J. A. Bucklew, "Asymptotically Optimal Quantizers for Detection of I.I.D. Data," *IEEE Trans. Inform. Theory*, **IT-35**, 1989, pp. 316–325.

[7] Boettcher, K. L., and R. R. Tenney, "Distributed Decisionmaking with Constrained Decisionmakers," *IEEE Trans. Systems, Man, Cybernetics*, **SMC-16**, No. 6, 1986.

[8] Barkat, M., and P. K. Varshney, "Decentralized CFAR Signal Detection," *IEEE Trans. Aerospace Electron. Syst.*, **AES-25**, 1989, pp. 141–149.

[9] Chernoff, H., "A Measure of Asymptotic Efficiency for Tests of a Hypothesis Based on a Sum of Observations," *Ann. Math. Statist.*, **23**, 1952, pp. 493–507.

[10] Cover, T. M., "Hypothesis Testing with Finite Statistics," *Ann. Math. Statist.*, **40**, 1969, pp. 828–835.

[11] Conte, E., E. D'Addio, A. Farina, and M. Longo, "Multistatic Radar Detection: Synthesis and Comparison of Optimum and Suboptimum Receivers," *IEEE Proc.*, **130**, Part F, 1983, pp. 484–494.

[12] Chau, Y. A., and Geraniotis, E., "Memoryless Quantization and Fusion for Multisensor Systems for the Detection of Weak Signals in Dependent Noise," preprint, 1990.

[13] Chau, Y. A., and Geraniotis, E., "Quantization and Fusion for Multisensor Discrimination from Dependent Observations," preprint, 1990.

[14] Chair, Z., and P. K. Varshney, "Optimum Data Fusion in Multiple Sensor Detection Systems," *IEEE Trans. Aero. Electron. Syst.*, **AES-22**, 1986, pp. 98–101.

[15] Chair, Z., and P. K. Varshney, "Distributed Bayesian Hypothesis Testing with Distributed Data Fusion," *IEEE Trans. Syst., Man, Cybernetics*, **SMC-18**, 1988, pp. 695–699.

[16] Ekchian, L. K., "Optimal Design of Distributed Detection Networks," Ph.D. dissertation, Dept. of Electrical Engineering and Computer Science, MIT, Cambridge, MA, 1982.

[17] Ekchian, L. K., and R. R. Tenney, "Detection Networks," *Proc. 21st IEEE Conf. Decision Control, 1982*, pp. 686–691.

[18] Flynn, T. J., and R. M. Gray, "Encoding of Correlated Observations," *IEEE Trans. Inform. Theory*, **IT-33**, 1987, pp. 773–787.

[19] Geraniotis, E., "Robust Distributed Block and Sequential Detection in Uncertain Environments," preprint, 1990.

[20] Geraniotis, E., and Chau, Y. A., "Robust Data Fusion for Multisensor Detection Systems," *IEEE Trans. Inform. Theory*, **IT-36**, 1990, pp. 1265–1279.

[21] Garey, M. R., and D. S. Johnson, *Computers and Intractability: A Guide to the Theory of NP-Completeness*, W.H. Freeman, San Francisco, 1979.

[22] Han, T.S., "Hypothesis Testing with Multiterminal Data Compression," *IEEE Trans. Inform. Theory*, **IT-33**, 1987, pp. 759–772.

[23] Hellman, M. E., and T. M. Cover, "Learning with Finite Memory," *Ann. Math. Statist.*, **41**, 1970, pp. 765–782.

[24] Hashemi, H. R., and I. B. Rhodes, "Decentralized Sequential Detection," *IEEE Trans. Inform. Theory*, **IT-35**, 1989, pp. 509–520.

[25] Hoballah, I. Y., and P. K. Varshney, "Neyman-Pearson Detection with Distributed Sensors," *Proc. 25th IEEE Conf. Decision Control, Athens, December 1986*, pp. 237–241.

[26] Hoballah, I. Y., and P. K. Varshney, "Distributed Bayesian Signal Detection," *IEEE Trans. Inform. Theory*, **IT-35**(5), 1989, pp. 995–1000.

[27] Koplowitz, J., "Necessary and Sufficient Memory Size for m-Hypothesis Testing," *IEEE Trans. Inform. Theory*, **IT-21**, 1975, pp. 44–46.

[28] Kazakos, D., "New Error Bounds and Optimum Quantization for Multisensor Distributed Signal Detection," preprint, 1988.

[29] Kazakos, D., V. Vannicola, and M. C. Wicks, "Optimum Quantization for Multisensor Detection," unpublished manuscript, 1989.

[30] Liapounoff, A., "Sur les Fonctions–Vecteurs Completement Additives," *Bull. Acad. Sci. USSR, Ser. Math.*, **4**, 1940, pp. 465–478.

[31] Luenberger, D. G. *Linear and Nonlinear Programming*, Addison-Wesley, Reading, MA, 1984.

[32] LaVigna, A., A. M. Makowski, and J. S. Baras, "A Continuous-Time Distributed Version of Wald's Sequential Hypothesis Testing Problem," in *Proceedings of the 7th International Conference on Analysis and Optimization of Systems* (A. Bensoussan and J. L. Lions, eds.), pp. 533–543, Antibes, France, 1986.

[33] Mullis, C. T., and R. A. Roberts, "Memory Limitation and Multistage Decision Processes," *IEEE Trans. Syst. Sci. Cybernetics*, **SSC-3**, 1968, pp. 307–316.

[34] Marschak, J. R., and R. Radner, *The Economic Theory of Teams*, Yale University Press, New Haven, CT, 1972.

[35] Poor, H.V., "Fine Quantization in Signal Detection and Estimation," *IEEE Trans. Inform. Theory*, **IT-34**, 1988, pp. 960–972.

[36] Polychronopoulos, G. H., "Solution of Some Problems in Decentralized Detection by a Large Number of Sensors," technical report LIDS-TH-1781, Laboratory for Information and Decision Systems, MIT, Cambridge, MA, June 1988.

[37] Pothiawala, J., "Analysis of a Two-Sensor Tandem Distributed Detection Network," technical report LIDS-TH-1848, Laboratory for Information and Decision Systems, MIT, Cambridge, MA, 1989.

[38] Papastavrou, J. D., "Decentralized Decision Making in a Hypothesis Testing Environment," Ph.D. thesis, MIT, Cambridge, MA, May 1990.

[39] Papastavrou, J. D., and M. Athans, "A Distributed Hypothesis-Testing Team Decision Problem with Communications Cost," *Proc. 25th IEEE Conf. Decision Control, Athens, December 1986*, pp. 219–225.

[40] Papastavrou, J. D., and M. Athans, "Optimum Configuration for Distributed Teams of Two Decision-Makers," technical report LIDS-P-1813, Laboratory for Information and Decision Systems, MIT, Cambridge, MA, 1988.

[41] Papastavrou, J. D., and M. Athans, "Distributed Detection by a Large Team of Sensors in Tandem," technical report LIDS-P-1929, Laboratory for Information and Decision Systems, MIT, Cambridge, MA, 1989.

[42] Picinbono, B., and P. Duvaut, "Optimum Quantization for Detection," IEEE Trans. Commun., **COM-36**, 1988, pp. 1254–1258.

[43] Papadimitriou, C. H., and K. Steiglitz, *Combinatorial Optimization: Algorithms and Complexity*, Prentice-Hall, Englewood Cliffs, NJ, 1982.

[44] Poor, H. V., and J. B. Thomas, "Applications of Ali-Silvey Distance Measures in the Design of Generalized Quantizers for Binary Decision Systems," *IEEE Trans. Commun.*, **COM-25**, 1977, pp. 893–900.

[45] Papadimitriou, C. H., and J. N. Tsitsiklis, "Intractable Problems in Control Theory," *SIAM J. Control Optimiz.*, **24**, 1986, pp. 639–654.

[46] Polychronopoulos, G., and J. N. Tsitsiklis, "Explicit Solutions for Some Simple Decentralized Detection Problems," *IEEE Trans. Aerospace Electron. Syst.*, **AES-26**, 1990, pp. 282–291.

[47] Radner, R., "Team Decision Problems," *Ann. Math. Statist.*, **33**, 1962, pp. 857–881 .

[48] Reibman, A. R., "Performance and Fault-Tolerance of Distributed Detection Networks," Ph.D. dissertation, Dept. of Electrical Engineering, Duke University, Durham, North Carolina, 1987.

[49] Reibman, A. R., and L. W. Nolte, "Optimal Detection and Performance of Distributed Sensor Systems," *IEEE Trans. Aero. Electron. Syst.*, **AES-23**, 1987, pp. 24–30.

[50] Reibman, A. R., and L. W. Nolte, "Design and Performance Comparison of Distributed Detection Networks," *IEEE Trans. Aero. Electron. Syst.*, **AES-23**, 1987, pp. 789–797.

[51] Sadjadi, F., "Hypothesis Testing in a Distributed Environment," *IEEE Trans. Aero. Electron. Syst.*, **AES-22**, 1986, pp. 134–137.

[52] Srinivasan, R., "Distributed Radar Detection Theory," *IEEE Proc.*, **133**, 1986, pp. 55–60.

[53] Srinivasan, R., "A Theory of Distributed Detection," *Signal Proc.*, **11**, 1986, pp. 319–327.

[54] Shannon, C. E., R. G. Gallager, and E. R. Berlekamp, "Lower Bounds to Error Probability for Coding on Discrete Memoryless Channels," *Inform. Control*, **10**, 1967, pp. 65–103.

[55] Sarma, V. V. S., and K. A. Gopala Rao, "Decentralized Detection and Estimation in Distributed Sensor Systems," *Proc. 1983 IEEE Cybernetics Soc. Conf., 1983*, pp. 438–441.

[56] Shalaby, H. M. H., and A. Papamarcou, "Distributed Detection with Feedback," technical report 90-64, Systems Research Center, University of Maryland, 1990.

[57] Tsitsiklis, J. N., "Problems in Decentralized Decision Making and Computation," Ph.D. dissertation, Dept. of Electrical Engineering and Computer Science, MIT, Cambridge, MA, 1984.

[58] Tsitsiklis, J. N., "On Threshold Rules in Decentralized Detection," *Proc. 25th IEEE Conf. Decision Control, Athens, 1986*, pp. 232–236.

[59] Tsitsiklis, J. N., "Decentralized Detection by a Large Number of Sensors," *Math. Control, Signals, Syst.*, **1**, 1988, pp. 167–182.

[60] Teneketzis, D., "A Signalling Problem Arising in Decentralized Detection," technical report 256, Dept. of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, 1989.

[61] Tsitsiklis, J.N., "Extremal Properties of Likelihood-Ratio Quantizers," *IEEE Trans. Commun.*, **COM-41**(4), 1993, pp. 550–558.

[62] Tsitsiklis, J. N., and M. Athans, "On the Complexity of Decentralized Decision Making and Detection Problems," *IEEE Trans. Automatic Control*, **AC-30**(5), 1985, pp. 440–446.

[63] Teneketzis, D., and Y.-C. Ho, "The Decentralized Wald Problem," *Inform. Control*, **73**, 1987, pp. 23–44.

[64] Tang, Z.-B., K. R. Pattipati, and D. L. Kleinman, "A Distributed M-ary Hypothesis Testing Problem with Correlated Observations," *Proc. 28th Conf. Decision Control, Tampa, Florida, 1989*, pp. 562–568.

[65] Tang, Z.-B., K. R. Pattipati, and D. L. Kleinman, "An Algorithm for Determining the Decision Thresholds in a Distributed Detection Environment," preprint, 1989.

[66] Tenney, R. R., and Sandell, N. R. Jr., "Detection with Distributed Sensors," *IEEE Trans. Aero. Electron. Syst.*, **AES-17**(4), 1981, pp. 501–510.

[67] Teneketzis, D., and P. Varaiya, "The Decentralized Quickest Detection Problem," *IEEE Trans. Automatic Control*, **AC-29**, 1984, pp. 641–644.

[68] Thomopoulos, S. C. A., R. Viswanathan, and D. P. Bougoulias, "Optimal Decision Fusion in Multiple Sensor Systems," *IEEE Trans. Aero. Electron. Syst.*, **AES-23**, 1987.

[69] Thomopoulos, S. C. A., R. Viswanathan, and D. K. Bougoulias, "Optimal Distributed Decision Fusionn, *IEEE Trans. Aero. Electron. Syst.*, **AES-25**, 1989, pp. 761–764.

[70] Viswanathan, R., S. C. A. Thomopoulos, and R. Tumuluri, "Optimal Serial Distributed Decision Fusion," *IEEE Trans. Aero. Electron. Syst.*, **AES-24**, 1988, pp. 366–375.

[71] Viswanathan, R., A. Ansari, and S. C. A. Thomipoulos, "Optimal Partitioning of Observations in Distributed Detection," technical report TR-SIU-EE-3-87, Dept. of Electrical Engineering, Southern Illinois University, 1987; also, *Abstracts of Papers, Int. Symp. Inform. Theory, Kobe, Japan, June 1988*, p. 197.

[72] Warren, D. J., and P. K. Willett, "Optimal Decentralized Detection for Conditionally Independent Sensors," *Proc. 1989 Am. Control Conf.*, pp. 1326–1329.

[73] Warren, D. J., and P. K. Willett, "Randomization in Distributed Detection," *Proc. 29th Conf. Decision Control, Honolulu, December 1990*, pp. 1151–1152.