

Explicit Solutions for Some Simple Decentralized Detection Problems

GEORGE POLYCHRONOPOULOS
JOHN N. TSITSIKLIS
M.I.T.

We consider a decentralized detection problem in which a number of identical sensors transmit a finite-valued function of their observations to a fusion center which makes a final decision on one of M alternative hypotheses. We consider the case where the number of sensors is large and we derive (asymptotically) optimal rules for determining the messages of the sensors, for the case where the observations are generated from a simple and symmetrical set of discrete distributions. We also consider the tradeoff between the number of sensors and the communication rate of each sensor when there is a constraint on the total communication rate from the sensors to the fusion center.

I. PROBLEM DEFINITION

The decentralized detection problem is defined as follows. There are $M \geq 2$ hypotheses H_1, H_2, \dots, H_M with known a priori probabilities $P(H_j) > 0$ and N sensors. Each sensor i obtains an observation y_i , where y_i is a random variable taking values in a set Y . We assume that y_1, \dots, y_N are conditionally independent (given the true hypothesis) and identically distributed with known conditional distributions $P_Y(\cdot | H_j)$. Each sensor i evaluates a D -valued message $u_i \in \{1, \dots, D\}$, as a function of its observation, and transmits it to a fusion center. Finally, the fusion center declares one of the alternative hypotheses to be true (Fig. 1).

Let $\gamma_i : Y \mapsto \{1, \dots, D\}$, $i = 1, 2, \dots, N$, be the function (to be called a *decision rule*) used by the i th sensor to determine its message u_i ; that is, $u_i = \gamma_i(y_i)$. Let $u_0 \in \{1, \dots, M\}$ be the decision of the fusion center. This decision is made according to a decision rule $\gamma_0 : \{1, \dots, D\}^N \mapsto \{1, \dots, M\}$; that is, $u_0 = \gamma_0(u_1, \dots, u_N)$. We say that the fusion center makes an error if $u_0 = i$ and H_i is not the true hypothesis. The probability of error is completely determined by the statistics of the observations and by the decision rules $\gamma_0, \gamma_1, \dots, \gamma_N$; it is denoted by $J_N(\gamma_0, \dots, \gamma_N)$. Our problem is to choose the decision rules $\gamma_0, \gamma_1, \dots, \gamma_N$ of the sensors and of the fusion center so as to minimize the probability of error.

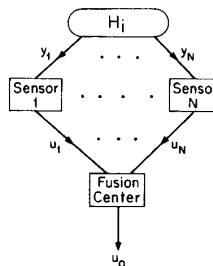


Fig. 1.

The above described problem and its variations have attracted substantial interest [1, 3, 4-6, 8-10, 12-18]. It was first introduced in [14] for the case of two hypotheses ($M = 2$), two sensors ($N = 2$), binary messages ($D = 2$), and for a fixed choice of the decision rule γ_0 of the fusion center. It was shown in [14] that under the conditional independence assumption, each sensor should evaluate its message u_i using a likelihood ratio test with an appropriate threshold. (This conclusion is not valid if the conditional independence assumption is removed in which case the problem becomes computationally intractable [17].) The optimal thresholds in the likelihood ratio tests of the different sensors can be obtained by solving a system of nonlinear equations. It is important to emphasize that the optimal decision rules for the decentralized problem are not, in general, the same as those that would be derived using the

Manuscript received November 2, 1988; revised July 14, 1989.

IEEE Log No. 33477.

This work was supported by the M.I.T.-Lincoln Laboratory and by the ONR under Contract N00014-84-K-0519 (NR 649-003).

Authors' address: Laboratory for Information and Decision Systems, Room 35-214, 77 Massachusetts Ave., Massachusetts Institute of Technology, Cambridge, MA 02139.

0018-9251/90/0300-0282 \$1.00 © 1990 IEEE

classical theory, independently for each sensor. This is because the optimal decision rules are chosen so as to optimize system-wide performance, as opposed to the performance of each individual sensor.

The performance of a decentralized detection system is generally inferior to that of a centralized system in which all raw data available are transmitted to the fusion center, due to the loss of information in the local processing. However, decentralized detection is often more practical due to the reduction of the communication requirements, as well as because the processing of the data is shared by a number of different processors. On the other hand, decentralized detection problems are qualitatively different and much more difficult than the corresponding centralized detection problems. For this reason, there are very few such problems that have been solved analytically. In fact, most of the theoretical research available is limited to the derivation of necessary conditions for optimality, and these can only be solved numerically. In contrast, we identify a special case for which an explicit solution can be obtained analytically.

We now define the particular problem to be studied. We assume that there is a one-to-one correspondence between observations and hypotheses and, more specifically, $Y = \{1, \dots, M\}$. We assume that the conditional distribution of the observation y of any sensor is given by

$$\Pr(y = i | H_j) = P_Y(i | H_j) = \begin{cases} \epsilon, & \text{if } j \neq i \\ 1 - (M - 1)\epsilon, & \text{if } j = i \end{cases}$$

where ϵ is a scalar satisfying $0 < \epsilon \leq 1/(M - 1)$. In other words, the observation of a sensor indicates the true hypothesis with probability $1 - (M - 1)\epsilon$, or it indicates a false hypothesis in which case each one of the false hypotheses is equally likely (probability ϵ). Furthermore, we assume that the number of sensors is large and we are looking for an asymptotic solution, as $N \rightarrow \infty$.

Our model is undoubtedly too structured to be an exact representation of a realistic problem, the main drawback being the assumption that there is a one-to-one correspondence between hypotheses and possible observations. This assumption becomes fairly reasonable, however, in the following situation (see Fig. 2). Each sensor i receives some observations z_i that it processes in some predetermined way, and comes up with a preliminary decision $y_i \in \{1, \dots, M\}$ on the identity of the true hypothesis. Then, each sensor i transmits to the fusion center a function $\gamma_i(y_i)$ of its preliminary decision y_i . Notice that we are restricting here the message to be a function of the processed observations instead of the raw observations. While such a restriction may result in some loss of performance, it is quite natural in certain contexts, especially if each sensor has a reason to come up with a preliminary decision in a timely manner.

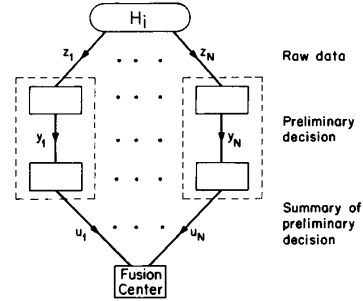


Fig. 2.

The above discussion notwithstanding, our interest in this particular problem arises mainly from the fact that an explicit solution can be obtained, as is demonstrated in the sequel. Furthermore, the solution to be derived provides insights and intuition on the nature of optimal solutions to more general problems for which explicit solutions are not possible. Such insights are very valuable because they can suggest interesting numerical experiments and heuristic guidelines for coping with more difficult problems.

The remainder of this paper is organized as follows. In Section II, we outline some results from [19] that are needed later. In Section III, we introduce some notation and terminology, and some simple preliminary facts. In Section IV, a complete solution is derived for the case where the noise parameter ϵ is small and the number of sensors is large. In Section V, we provide a partial extension of the results of Section IV to the case of a general noise parameter ϵ . Finally, in Section VI, we study the tradeoff between the number of sensors and the communication rate of each sensor when there is a constraint on the total communication rate from the sensors to the fusion center.

II. BACKGROUND

As mentioned in the introduction, we are looking for an asymptotic solution to our problem, as the number of sensors N becomes very large. The basic theory concerning such an asymptotic solution has been developed in [19] and we review here the facts that are needed. Some experimentation [7] has shown that the asymptotically optimal decision rules perform reasonably well for moderate numbers of sensors.

We use Γ to denote the set of all possible decision rules. Due to the finiteness of the observation set Y and of the message set $\{1, \dots, D\}$, it is seen that the set Γ is also finite. We introduce the shorthand notation γ^N to denote a possible choice $(\gamma_0, \gamma_1, \dots, \gamma_N)$ of decision rules for the N -sensor problem. With a reasonable choice of γ^N , the probability of error $J_N(\gamma^N)$ converges exponentially to zero as N increases. For this reason, we focus on the exponent of the error

probability, defined by

$$r_N(\gamma^N) = \frac{\log J_N(\gamma^N)}{N}. \quad (1)$$

Let $R_N = \inf_{\gamma^N} r_N(\gamma^N)$, where the infimum is taken over all possible choices of decision rules for the N -sensor problem. Thus, R_N is the optimal exponent. As N tends to infinity, R_N has a limit [19] which is denoted by Λ^* . In the sequel, we are concerned with choosing the decision rules so that the corresponding error exponent approaches the optimal exponent Λ^* .

Consider a sensor that uses a particular decision rule $\gamma \in \Gamma$. Conditioned on H_i , the probability that the transmitted message takes a particular value $d \in \{1, \dots, D\}$ is given by $\Pr(\gamma(y) = d | H_i)$. For every $i, j \in \{1, \dots, M\}$ and every decision rule $\gamma \in \Gamma$, we define a function $\mu_{ij}(\gamma, \cdot) : [0, 1] \mapsto [-\infty, +\infty)$ by

$$\begin{aligned} & \mu_{ij}(\gamma, s) \\ &= \log \left[\sum_{d=1}^D (\Pr(\gamma(y) = d | H_i))^{1-s} (\Pr(\gamma(y) = d | H_j))^s \right]. \end{aligned} \quad (2)$$

(The convention $0^0 = 0$ is used in this formula.) It is easily verified that $\mu_{ij}(\gamma, s) \leq 0$ for every $i, j, \gamma \in \Gamma$, $s \in [0, 1]$, and it is also known that $\mu_{ij}(\gamma, s)$ is a convex function of s , for every $i, j, \gamma \in \Gamma$ [11]. Furthermore, as long as there exists some $y \in Y$ such that $P_Y(y | H_i) \cdot P_Y(y | H_j) \neq 0$, then $\mu_{ij}(\gamma, s) > -\infty$, for every $s \in [0, 1]$. This turns out to be always the case for our problem except for the uninteresting situation where $M = 2$ and $\epsilon = 1$.

The optimal exponent is given [19] by

$$\Lambda^* = \min_{\{x_\gamma | \gamma \in \Gamma\}} \max_{\{(i,j) | i \neq j\}} \min_{s \in [0,1]} \sum_{\gamma \in \Gamma} x_\gamma \mu_{ij}(\gamma, s) \quad (3)$$

where the outer minimization is carried out over all choices of vectors $\{x_\gamma | \gamma \in \Gamma\}$ satisfying $x_\gamma \geq 0$ for all $\gamma \in \Gamma$, and $\sum_{\gamma \in \Gamma} x_\gamma = 1$. In the sequel, we use x to denote a vector $\{x_\gamma | \gamma \in \Gamma\}$. Furthermore, we use X to denote the set of all such vectors which satisfy the constraints just stated.

The variable x_γ in (3) should be interpreted as the fraction of the sensors that use decision rule γ . More specifically, let us fix some $x \in X$. For each $\gamma \in \Gamma$, let $\lfloor Nx_\gamma \rfloor$ sensors use decision rule γ . (If for some γ the value of Nx_γ is not integer this determines the decision rules for fewer than N sensors. However, the remaining sensors constitute a vanishingly small fraction of the total, as $N \rightarrow \infty$, and are inconsequential.) Then, the asymptotic exponent (as $N \rightarrow \infty$) of the probability of error is given by [19]

$$\max_{\{(i,j) | i \neq j\}} \min_{s \in [0,1]} \sum_{\gamma \in \Gamma} x_\gamma \mu_{ij}(\gamma, s). \quad (4)$$

In particular, if the fractions x_γ are chosen to minimize the exponent in (4), then the optimal exponent Λ^* is obtained [compare with (3)]. Notice that the problem formulation has taken a somewhat different, but equivalent, form: instead of choosing the decision rule of each sensor, we are now trying to choose the fraction x_γ of the sensors that use a given decision rule $\gamma \in \Gamma$.

Equation (4) has a simple interpretation. The quantity $\min_{s \in [0,1]} \sum_{\gamma \in \Gamma} x_\gamma \mu_{ij}(\gamma, s)$ is the exponent in the Chernoff bound for the probability of confusing hypotheses H_i and H_j ([11, 20]), and such a bound is known to be asymptotically tight. The maximization over all i and j in (4) corresponds to the fact that the dominant term in the probability of error comes from the worst (i.e., the largest) of the exponents corresponding to the different pairs.

The outer minimization in (3) appears to be simple because it involves linear constraints and a cost function which is linear in the variables x_γ . However, the inner minimization (with respect to s) severely complicates the computation of Λ^* and of the optimal values of the variables x_γ . In Sections III and IV, we get around this difficulty by exploiting the symmetry of the problem to remove the dependence on s .

III. PRELIMINARIES

Consider a decision rule $\gamma : Y \mapsto \{1, \dots, D\}$ and let $Y_{d,\gamma} = \{y | \gamma(y) = d\}$. We notice that the sets $Y_{d,\gamma}$, $d = 1, \dots, D$, are disjoint, their union equals Y , and therefore determine a partition of Y into D disjoint sets. It is possible that two different decision rules determine the same partition in which case they can be considered equivalent, since they convey the same information to the fusion center. From now on, we do not distinguish between equivalent decision rules and we consider them to be identical. We are therefore adopting the alternative definition that a decision rule is a partition of Y into subsets $Y_{1,\gamma}, \dots, Y_{D,\gamma}$. We assume that the sets $Y_{d,\gamma}$ are arranged in order of increasing cardinality; that is, $|Y_{1,\gamma}| \leq \dots \leq |Y_{D,\gamma}|$.

DEFINITION. Two observations $i, j \in Y$ are *separated* by a decision rule γ if i and j belong to different elements $Y_{d,\gamma}$ of the partition corresponding to γ . We let Γ_{ij} be the set of all $\gamma \in \Gamma$ that separate i and j . The *number of separations* corresponding to a decision rule γ is defined as the number of (unordered) pairs of observations $i, j \in Y$ which are separated by γ .

Notice that an M -ary hypothesis testing problem can be viewed as a collection of several binary hypothesis testing problems, one for each pair of hypotheses. The number of separations corresponding to a decision rule γ can be interpreted as the number of binary problems for which a message $\gamma(y_i)$ provides useful information.

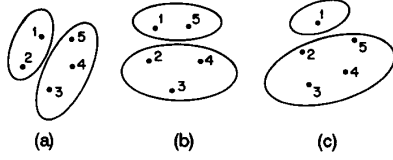


Fig. 3. Let $M = 5$, $D = 2$. The decision rules in (a) and (b) belong to the class $C^{2,3}$ and the corresponding number of separations is 6. The decision rule in (c) belongs to the class $C^{1,4}$ and the corresponding number of separations is 4.

DEFINITION. Let $\delta_1, \dots, \delta_D$ be a collection of nonnegative integers satisfying $\delta_1 \leq \delta_2 \leq \dots \leq \delta_D$ and $\sum_{d=1}^D \delta_d = M$. The class $C^{\delta_1, \dots, \delta_D}$ is the set of all $\gamma \in \Gamma$ such that $|Y_{d,\gamma}| = \delta_d$ for every d .

These definitions are illustrated in Fig. 3.

Let L be the number of different classes. In order to facilitate notation, we assume that the different classes have been arranged according to some arbitrary order and we use the simpler notation C_l to denote the l th class, $l = 1, \dots, L$. Thus, the set Γ of all decision rules is equal to $\bigcup_{l=1}^L C_l$.

It is seen that the number of separations is the same for all decision rules belonging to the same class C_l [see Fig. 3], and is denoted by S_l . In particular,

$$S_l = \frac{1}{2} \sum_{d=1}^D \delta_d (M - \delta_d) \quad (5)$$

where $\delta_1, \dots, \delta_D$ are such that $C_l = C^{\delta_1, \dots, \delta_D}$. [The factor 1/2 in (5) is present in order to avoid counting the same unordered pair twice.]

Let Q_l be the cardinality of the set of all triples (i, j, γ) such that $\gamma \in C_l$ and γ separates i and j . [The two triples (i, j, γ) and (j, i, γ) are only counted once.] Since the number of separations corresponding to any $\gamma \in C_l$ is S_l , we see that $Q_l = |C_l| \cdot S_l$. On the other hand, every pair (i, j) is separated by exactly $|C_l \cap \Gamma_{ij}|$ elements of C_l . By symmetry, the cardinality of $C_l \cap \Gamma_{ij}$ is the same for every i and j . Furthermore, since there exist $M(M-1)/2$ different (unordered) pairs (i, j) , we conclude that $Q_l = |C_l \cap \Gamma_{ij}| \cdot M(M-1)/2$. By equating the two alternative expressions for Q_l , we obtain

$$\frac{|C_l \cap \Gamma_{ij}|}{|C_l|} = \frac{2S_l}{M(M-1)} \quad (6)$$

a fact that will be useful later.

We now derive the form of the functions $\mu_{ij}(\gamma, s)$. Suppose that $i \in Y_{\eta,\gamma}$ and $j \in Y_{\zeta,\gamma}$. Using the notation $\delta_\eta = |Y_{\eta,\gamma}|$ and $\delta_\zeta = |Y_{\zeta,\gamma}|$, it is seen (cf. (2)) that

$$\begin{aligned} \mu_{ij}(\gamma, s) = & \log[(1 - (M - \delta_\eta)\epsilon)^{1-s} (\delta_\eta\epsilon)^s \\ & + (\delta_\zeta\epsilon)^{1-s} (1 - (M - \delta_\zeta)\epsilon)^s \\ & + (M - \delta_\eta - \delta_\zeta)\epsilon], \quad \text{if } \eta \neq \zeta \quad (7) \end{aligned}$$

and

$$\mu_{ij}(\gamma, s) = 0, \quad \text{if } \eta = \zeta. \quad (8)$$

Notice that the case $\eta = \zeta$ [cf. (8)] corresponds to the case $\gamma \notin \Gamma_{ij}$. Finally, from either (2) or (7), it is seen that

$$\mu_{ij}(\gamma, s) = \mu_{ji}(\gamma, 1 - s) \quad (9)$$

which is useful later.

IV. SMALL NOISE CASE

In this section, we derive the solution of our problem for the case where the noise parameter ϵ is small. This is accomplished by showing that the minimum with respect to s in (3) is approximately attained for $s = \frac{1}{2}$, which allows us to eliminate s .

LEMMA 1. Fix some ϵ_0 such that $0 < \epsilon_0 < 1/(M-1)$. Then, there exist constants G_1 and G_2 such that, for every $\epsilon \in (0, \epsilon_0)$, every $i, j \in \{1, \dots, M\}$ such that $i \neq j$, and every $x \in X$, we have

$$\begin{aligned} G_1 + \frac{1}{2} \log \epsilon \sum_{\gamma \in \Gamma_{ij}} x_\gamma & \leq \min_{s \in [0, 1]} \sum_{\gamma \in \Gamma} x_\gamma \mu_{ij}(\gamma, s) \\ & \leq G_2 + \frac{1}{2} \log \epsilon \sum_{\gamma \in \Gamma_{ij}} x_\gamma. \end{aligned}$$

PROOF. We first prove the right-hand side inequality. Consider some $\gamma \in \Gamma_{ij}$ and suppose that $i \in Y_{\eta,\gamma}$, $j \in Y_{\zeta,\gamma}$. We have (cf. (7))

$$\begin{aligned} e^{\mu_{ij}(\gamma, 1/2)} & = (1 - (M - \delta_\eta)\epsilon)^{1/2} (\delta_\eta\epsilon)^{1/2} \\ & \quad + (\delta_\zeta\epsilon)^{1/2} (1 - (M - \delta_\zeta)\epsilon)^{1/2} \\ & \quad + (M - \delta_\eta - \delta_\zeta)\epsilon \\ & \leq (\delta_\eta\epsilon)^{1/2} + (\delta_\zeta\epsilon)^{1/2} + (M - \delta_\eta - \delta_\zeta)\epsilon^{1/2} \\ & \leq H_2 \epsilon^{1/2} \end{aligned}$$

where $H_2 = \delta_\eta^{1/2} + \delta_\zeta^{1/2} + M - \delta_\eta - \delta_\zeta > 1$. Taking logarithms, we obtain $\mu_{ij}(\gamma, \frac{1}{2}) \leq G_2 + (\log \epsilon)/2$, where $G_2 = \log H_2 > 0$. Furthermore, if $\gamma \notin \Gamma_{ij}$, we have $\mu_{ij}(\gamma, 1/2) = 0$ (cf. (8)). It follows that

$$\sum_{\gamma \in \Gamma} x_\gamma \mu_{ij} \left(\gamma, \frac{1}{2} \right) \leq G_2 + \frac{1}{2} \log \epsilon \sum_{\gamma \in \Gamma_{ij}} x_\gamma.$$

If a minimization over s is carried out, the resulting value is no larger than the one corresponding to $s = \frac{1}{2}$, and this proves the right-hand side inequality.

We now prove the left-hand side inequality. We fix some i, j , some $\gamma \in \Gamma_{ij}$, and some $s \in [0, \frac{1}{2}]$. We assume again that $i \in Y_{\eta,\gamma}$ and $j \in Y_{\zeta,\gamma}$. We have

$$\begin{aligned} e^{\mu_{ij}(\gamma, s)} & = (1 - (M - \delta_\eta)\epsilon)^{1-s} (\delta_\eta\epsilon)^s \\ & \quad + (\delta_\zeta\epsilon)^{1-s} (1 - (M - \delta_\zeta)\epsilon)^s + (M - \delta_\eta - \delta_\zeta)\epsilon \\ & \geq (1 - (M - \delta_\eta)\epsilon)^{1-s} (\delta_\eta\epsilon)^s \geq (1 - (M - \delta_\eta)\epsilon)\epsilon^s \\ & \geq (1 - (M - 1)\epsilon_0)\epsilon^s = H_1 \epsilon^s \geq H_1 \epsilon^{1/2} \end{aligned}$$

where $H_1 = 1 - (M - 1)\epsilon_0 > 0$. Taking logarithms, we obtain $\mu_{ij}(\gamma, s) \geq G_1 + (\log \epsilon)/2$, where $G_1 = \log H_1 < 0$. The same conclusion is obtained by a symmetrical argument for the case $s \in [\frac{1}{2}, 1]$. Using again the fact that $\mu_{ij}(\gamma, s) = 0$ if $\gamma \notin \Gamma_{ij}$, we obtain

$$\begin{aligned} \min_{s \in [0,1]} \sum_{\gamma \in \Gamma} x_\gamma \mu_{ij}(\gamma, s) &\geq \sum_{\gamma \in \Gamma} x_\gamma \min_{s \in [0,1]} \mu_{ij}(\gamma, s) \\ &\geq \sum_{\gamma \in \Gamma_{ij}} x_\gamma \left(\frac{1}{2} \log \epsilon + G_1 \right) \\ &\geq G_1 + \frac{1}{2} \log \epsilon \sum_{\gamma \in \Gamma_{ij}} x_\gamma \end{aligned}$$

which completes the Proof. Q.E.D.

We notice that as ϵ approaches zero, $\log \epsilon$ tends to $-\infty$, while the constants G_1, G_2 of Lemma 1 remain unchanged. Therefore, by retaining the dominant term, Λ^* can be approximated, for small ϵ , by

$$\Lambda^* = \frac{1}{2} \min_{x \in X} \max_{\{(i,j)|i \neq j\}} \log \epsilon \sum_{\gamma \in \Gamma_{ij}} x_\gamma. \quad (10)$$

Since $\log \epsilon$ is negative, an equivalent optimization problem is

$$\max_{x \in X} \min_{\{(i,j)|i \neq j\}} \sum_{\gamma \in \Gamma_{ij}} x_\gamma. \quad (11)$$

We now derive the solution of (11).

PROPOSITION 1. *Let $S^* = \max_i S_i$. Then, a vector $x \in X$ is an optimal solution of the problem (11) if and only if the following two conditions hold: 1) the value of $\sum_{\gamma \in \Gamma_{ij}} x_\gamma$ is the same for every pair (i, j) such that $i \neq j$, and 2) if $\gamma \in C_l$ and $S_l < S^*$, then $x_\gamma = 0$. Furthermore, the optimal value of (11) is equal to $2S^*/(M(M - 1))$.*

PROOF. Suppose that a vector $x^* \in X$ satisfies conditions 1) and 2), and let c be such that $c = \sum_{\gamma \in \Gamma_{ij}} x_\gamma^*$, for $i \neq j$. Summing over all unordered pairs (i, j) , we obtain

$$\begin{aligned} c \frac{M(M-1)}{2} &= \sum_{\{(i,j)|i \neq j\}} \sum_{\gamma \in \Gamma_{ij}} x_\gamma^* = \sum_{\gamma \in \Gamma} \sum_{\{(i,j)|\gamma \in \Gamma_{ij}\}} x_\gamma^* \\ &= \sum_{\gamma \in \Gamma} S^* x_\gamma^* = S^*. \end{aligned}$$

(Here we used the fact that if $\gamma \in C_l$, then the cardinality of the set $\{(i, j) | \gamma \in \Gamma_{ij}\}$ is S_l , by definition; we then used property 2) to replace S_l by S^* .) We conclude that if conditions 1) and 2) hold, then $c = 2S^*/(M(M - 1))$.

In order to show that the vector x^* is actually optimal, it is sufficient to show that

$$\min_{\{(i,j)|i \neq j\}} \sum_{\gamma \in \Gamma_{ij}} x_\gamma \leq \frac{2S^*}{M(M-1)}$$

for every vector $x \in X$. We use the elementary fact that the minimum of a set of numbers is no larger than their average, to obtain

$$\begin{aligned} &\frac{M(M-1)}{2} \min_{\{(i,j)|i \neq j\}} \sum_{\gamma \in \Gamma_{ij}} x_\gamma \\ &\leq \sum_{\{(i,j)|i \neq j\}} \sum_{\gamma \in \Gamma_{ij}} x_\gamma = \sum_l \sum_{\gamma \in C_l} \sum_{\{(i,j)|\gamma \in \Gamma_{ij}\}} x_\gamma \\ &= \sum_l \sum_{\gamma \in C_l} S_l x_\gamma \leq S^* \sum_l \sum_{\gamma \in C_l} x_\gamma = S^* \end{aligned} \quad (12)$$

as desired. We conclude that x^* is optimal.

For the converse, let us suppose that a vector $x \in X$ is optimal. We have already established that the optimal value of the objective function under consideration is equal to $2S^*/(M(M - 1))$. Therefore, all inequalities in (12) must be equalities. Since the first inequality in (12) is not strict, condition 1) follows. Furthermore, since the second inequality in (12) is not strict, condition 2) follows. Q.E.D.

Using Proposition 1, one optimal solution for the problem (11) is the following. Choose a class C_{l^*} such that $S_{l^*} = S^* = \max_l S_l$ and let

$$x_\gamma = \begin{cases} 0, & \text{if } \gamma \notin C_{l^*} \\ 1/|C_{l^*}|, & \text{if } \gamma \in C_{l^*}. \end{cases} \quad (13)$$

It is seen that this vector x is feasible ($x \in X$) and satisfies the optimality conditions of Proposition 1. Let us point out that an optimal solution of the problem (11) is in general not unique. The solution provided by (13) can be singled out because of its special symmetry properties.

The class C_{l^*} , which is a class of decision rules with a maximal number of separations, should be viewed as a "best" class. According to Proposition 1 only decision rules in such a class should be used. This is very intuitive because each $\gamma \in C_l$ provides information to the fusion center which is useful in discriminating S_l pairs of hypotheses (by the definition of S_l). The larger the value of S_l , the larger the contribution of a decision rule $\gamma \in S_l$ in discriminating between the different hypotheses.

We now proceed to determine the best class C_{l^*} . Suppose that $C_{l^*} = C^{\delta_1, \dots, \delta_D}$, for some integer coefficients $\delta_1, \dots, \delta_D$ whose sum is equal to M . Suppose that there exist some η and ζ such that $\delta_\eta - \delta_\zeta > 1$. Consider a new class $C_{l'} = C^{\delta'_1, \dots, \delta'_D}$, where $\delta'_\eta = \delta_\eta - 1$, $\delta'_\zeta = \delta_\zeta + 1$, and $\delta'_d = \delta_d$ if $d \neq \eta$ and

$d \neq \zeta$. Using (5), we obtain

$$\begin{aligned} 2(S_{I'} - S_{I^*}) &= \delta'_\zeta(M - \delta'_\zeta) + \delta'_\eta(M - \delta'_\eta) \\ &\quad - \delta_\zeta(M - \delta_\zeta) - \delta_\eta(M - \delta_\eta) \\ &= (\delta_\zeta + 1)(M - \delta_\zeta - 1) + (\delta_\eta - 1)(M - \delta_\eta + 1) \\ &\quad - \delta_\zeta(M - \delta_\zeta) - \delta_\eta(M - \delta_\eta) \\ &= 2(\delta_\eta - \delta_\zeta - 1) > 0 \end{aligned}$$

which contradicts the optimality of S_{I^*} . This shows that $|\delta_n - \delta_\zeta| \leq 1$ for all η, ζ . Given that the average of the coefficients δ_d must be equal to M/D , it follows that for every d we must have either $\delta_d = \lfloor M/D \rfloor$ or $\delta_d = \lceil M/D \rceil$. In particular, if M is divisible by D , then $\delta_d = M/D$ for every d . If M is not divisible by D , the number of δ_d s for which $\delta_d = \lfloor M/d \rfloor$ is uniquely determined by the requirement $\sum_{d=1}^D \delta_d = M$.

We conclude that with decision rules belonging to the best class C_{I^*} , the corresponding partitions of the observation set Y are as even as possible. For example, if $D = 2$ and M is even, the set Y is to be partitioned into two subsets with equal cardinalities. Also, for the example of Fig. 3 in which $M = 5$ and $D = 2$, the best class is the class $C^{2,3}$. Notice that $C^{2,3}$ has 10 different elements; thus, an optimal solution is to divide the sensors in ten groups of equal cardinality and letting all the sensors in each group use a particular decision rule belonging to the class $C^{2,3}$.

Example

We discuss here, in qualitative terms, a simple example, in order to provide some feeling for the nature of our solution. Each sensor obtains an image of an animal and does some preprocessing to decide between the four hypotheses Black-Cat, White-Cat, Black-Dog, White-Dog (abbreviated as BC, WC, BD, WD, respectively). Each sensor transmits a binary message to a fusion center who declares one of the four hypotheses to be true. We restrict the messages to depend only on the preliminary decisions of the respective sensors (as opposed to arbitrary functions of their raw data).

With decision rules in the class $C^{1,3}$ each sensor concentrates on a particular hypothesis and its message reveals whether the sensor believes that hypothesis to be true. A typical such rule would correspond to a partition like $(\{BC\}, \{WC, BD, WD\})$. With such decision rules, each message helps the fusion center in resolving 3 different pairs of hypotheses.

Our solution suggests that decision rules in the class $C^{2,2}$ should be preferable. For example, half of the sensors could use the partition $(\{BC, BD\}, \{WC, WD\})$ (that is, provide color information to the center) and the other half could use the partition $(\{BC, WC\}, \{BD, WD\})$, thus providing species

information. Each message is more informative now because it helps in resolving 4 pairs of hypotheses. However, this solution is not optimal. In particular, all messages help in deciding between BC and WD, but only half of the messages are helpful in deciding between BC and WC. For this reason, we should also introduce a third partition $(\{BC, WD\}, \{WC, BD\})$ and let one third of the sensors use each one of the introduced partitions.

We have thus arrived at the interesting conclusions that a) it is not optimal to have each sensor focus on a particular hypothesis, and b) it is not optimal to decompose the four hypotheses into two binary hypotheses. Of course, these conclusions are only valid under our symmetry assumption, namely, that all errors are equally likely at the sensors. This makes sense if an error is caused by a burst of noise that destroys most useful information. On the other hand, one can also think of situations where, under the true hypothesis of a black cat, it is more likely that a sensor sees a black dog than a white dog. In the latter case, our conclusions do not apply.

We close this section by conjecturing that, for the small noise case, the same results hold even when N is small (and not only asymptotically). However, this would require a very different mathematical approach.

V. GENERAL CASE

We now consider the case where ϵ does not tend to zero but is fixed instead at some nonzero value in the range $0 < \epsilon < 1/(M - 1)$. Unfortunately, despite the symmetry of the optimization problem defining Λ^* , symmetry considerations alone are not sufficient to ascertain that the optimal value of the vector x possesses symmetry properties similar to the ones obtained in the previous section. We demonstrate this by means of a simple example.¹

Example

Let there be three hypotheses ($M = 3$) and let the messages be binary ($D = 2$). In this case there are exactly three decision rules, the following: the i th decision rule γ_i , $i = 1, 2, 3$, is defined by $\gamma_i(i) = 1$ and $\gamma_i(j) = 2$ if $j \neq i$. Notice that $\mu_{12}(\gamma_3, s) = \mu_{13}(\gamma_2, s) = \mu_{23}(\gamma_1, s) = 0$, for every s . Let

$$\nu(s) = \log[(1 - 2\epsilon)^{1-s}\epsilon^s + (2\epsilon)^{1-s}(1 - \epsilon)^s].$$

It is seen (cf. (2)) that $\mu_{ij}(\gamma_i, s) = \nu(s)$ and $\mu_{ij}(\gamma_j, s) = \nu(1 - s)$, for every $i \neq j$. Substituting in (3), and using

¹This example also corrects an error in a corresponding example in [19].

the notation $x_i = x_{\gamma_i}$, we obtain

$$\begin{aligned} \Lambda^* &= \min_{x \in X} \max \left\{ \begin{aligned} &\min_{s \in [0,1]} [x_1 \mu_{12}(\gamma_1, s) + x_2 \mu_{12}(\gamma_2, s) + x_3 \mu_{12}(\gamma_3, s)] \\ &\min_{s \in [0,1]} [x_1 \mu_{13}(\gamma_1, s) + x_2 \mu_{13}(\gamma_2, s) + x_3 \mu_{13}(\gamma_3, s)] \\ &\min_{s \in [0,1]} [x_1 \mu_{23}(\gamma_1, s) + x_2 \mu_{23}(\gamma_2, s) + x_3 \mu_{23}(\gamma_3, s)] \end{aligned} \right\} \\ &= \min_{x \in X} \max \left\{ \begin{aligned} &\min_{s \in [0,1]} [x_1 \nu(s) + x_2 \nu(1-s)] \\ &\min_{s \in [0,1]} [x_1 \nu(s) + x_3 \nu(1-s)] \\ &\min_{s \in [0,1]} [x_2 \nu(s) + x_3 \nu(1-s)] \end{aligned} \right\}. \end{aligned}$$

Consider the symmetric solution where the numbers of sensors using each decision rule are equal ($x_i = \frac{1}{3}$ for each i). The corresponding exponent is seen to be $\frac{1}{3} \min_{s \in [0,1]} [\nu(s) + \nu(1-s)] = \frac{2}{3} \nu(\frac{1}{2})$. (The last equality follows because we are minimizing a convex function which is symmetric around the point $1/2$.) Let us now consider the nonsymmetric solution $x_1 = x_2 = \frac{1}{2}$, $x_3 = 0$. The corresponding exponent is equal to

$$\max \left\{ \nu \left(\frac{1}{2} \right), \frac{1}{2} \min_{s \in [0,1]} \nu(s) \right\}.$$

In particular, if $\frac{1}{2} \min_{s \in [0,1]} \nu(s) < \frac{2}{3} \nu(\frac{1}{2})$, then the symmetric solution is not optimal. We have investigated this issue numerically by computing the value of the exponent corresponding to different vectors $x \in X$ (over a fairly dense grid of points in X and for a few different values of ϵ) and we have reached the conclusion that the symmetric solution is always the optimal one. However, an analytical method for establishing that this is the case is not apparent, even though it can be proved that the symmetric solution is a strict local minimum. (The proof of the latter fact is outlined in the Appendix.)

Without any guaranteed symmetry properties, little progress can be made analytically towards the computation of Λ^* . For this reason, we shall impose a symmetry requirement and proceed to solve the problem of (3) subject to this additional constraint. Motivated by the structure of an optimal solution for the low noise case (cf. Eq. (13)), we require that the value of x_γ be the same for every γ belonging to the same class. Given any vector $x \in X$ satisfying this requirement, let $y_l = \sum_{\gamma \in C_l} x_\gamma$. We then have $x_\gamma = y_l / |C_l|$ for every $\gamma \in C_l$. Using this expression for x_γ , the minimization problem of (3) becomes

$$\Lambda^* = \min_{y_1, \dots, y_L} \max_{\{(i,j)|i \neq j\}} \min_{s \in [0,1]} \sum_{l=1}^L \frac{y_l}{|C_l|} \sum_{\gamma \in C_l} \mu_{ij}(\gamma, s) \quad (14)$$

where the variables y_1, \dots, y_L are subject to the constraints $y_l \geq 0$, for each l , and $\sum_{l=1}^L y_l = 1$.

PROPOSITION 2.

a) Fix some class C_l . Then, the value of

$$\frac{1}{|C_l \cap \Gamma_{ij}|} \sum_{\gamma \in C_l \cap \Gamma_{ij}} \mu_{ij} \left(\gamma, \frac{1}{2} \right)$$

is the same for all i, j such that $i \neq j$, and is denoted by α_l .

b) Let l^* be such that $S_{l^*} |\alpha_{l^*}| = \max_l S_l |\alpha_l|$. Then, the choice $y_{l^*} = 1$, and $y_l = 0$ if $l \neq l^*$, is an optimal solution of the problem (14).

PROOF.

a) This is evident from the definition of $\mu_{ij}(\gamma, \frac{1}{2})$ and symmetry considerations.

b) Fix some pair (i, j) , with $i \neq j$. For any $\gamma \in \Gamma$, define a new decision rule $\sigma(\gamma)$ in which the positions of i and j in the partition corresponding to γ are interchanged (see Fig. 4). It is seen that σ is a one-to-one and onto mapping of any given class C_l into itself. Furthermore, it follows easily from the definition of μ_{ij} that $\mu_{ij}(\sigma(\gamma), s) = \mu_{ji}(\gamma, s) = \mu_{ij}(\gamma, 1-s)$. Therefore,

$$\begin{aligned} \sum_{\gamma \in C_l} \mu_{ij}(\gamma, s) &= \frac{1}{2} \sum_{\gamma \in C_l} [\mu_{ij}(\gamma, s) + \mu_{ij}(\sigma(\gamma), s)] \\ &= \frac{1}{2} \sum_{\gamma \in C_l} [\mu_{ij}(\gamma, s) + \mu_{ij}(\gamma, 1-s)]. \end{aligned} \quad (15)$$

Thus, the expression in the left-hand side of (15) is symmetric, as a function of s , around the value $s = \frac{1}{2}$. It follows that the minimization with respect to s in (14) involves a function which is convex and symmetric around the point $s = \frac{1}{2}$. Hence, the minimum is attained at $s = \frac{1}{2}$ and (14) simplifies to

$$\Lambda^* = \min_{y_1, \dots, y_L} \max_{\{(i,j)|i \neq j\}} \sum_{l=1}^L \frac{y_l}{|C_l|} \sum_{\gamma \in C_l} \mu_{ij} \left(\gamma, \frac{1}{2} \right). \quad (16)$$

Now, using part a) of the proposition,

$$\begin{aligned} \frac{1}{|C_l|} \sum_{\gamma \in C_l} \mu_{ij} \left(\gamma, \frac{1}{2} \right) &= \frac{1}{|C_l|} \sum_{\gamma \in C_l \cap \Gamma_{ij}} \mu_{ij} \left(\gamma, \frac{1}{2} \right) \\ &= \frac{|C_l \cap \Gamma_{ij}|}{|C_l|} \alpha_l = \frac{2S_l \alpha_l}{M(M-1)} \end{aligned} \quad (17)$$

where we have made use of (6) in the last step. We now use (17) to further simplify (16) and obtain

$$\Lambda^* = \min_{y_1, \dots, y_L} \sum_{l=1}^L \frac{2y_l}{M(M-1)} S_l \alpha_l. \quad (18)$$

Notice that the inequality $\alpha_l \leq 0$ holds for each l . Therefore, an optimal solution to the optimization problem of (18) is obtained by choosing a class C_{l^*} .

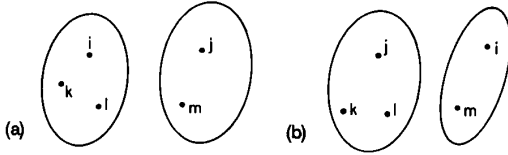


Fig. 4. A decision rule γ is shown in (a) and the corresponding decision rule $\sigma(\gamma)$ (in which the positions of i and j are interchanged) is shown in (b).

for which the value of $S_l|\alpha_l|$ is maximized and letting $y_{l^*} = 1$, and $y_l = 0$ if $l \neq l^*$. Q.E.D.

Our conclusions are therefore similar to the small noise case. In particular, there exists a best class and all decision rules to be used should belong to a best class. The nature of the best class is interesting. The constant α_l can be interpreted as a measure of the contribution of an “average” element of C_l to a pair of hypotheses which are separated by that decision rule (see Proposition 2a)). The product $S_l|\alpha_l|$ weighs the number of separations of a decision rule in C_l by the “quality measure” α_l and the value of this product is used to determine a best class.

The identity of the best class cannot be determined analytically because the formulas for the coefficients α_l are somewhat cumbersome. On the other hand, for any given value of ϵ , the value of α_l is easy to compute numerically. We have done so for the case where $D = 2$ and for $M = 5, 10, 20, 30$ [7]. We summarize the results. When ϵ is very small, then the optimal class is the one which partitions evenly the observation set, in agreement with the results of Section IV. Interestingly enough, this same class remains optimal for larger values of ϵ as well, up to approximately $1/M$. At about that point, the identity of the optimal class changes, and the optimal class is a most uneven one, namely the class $C^{1, M-1}$. This latter class remains the best one for all ϵ up to $1/(M-1)$ (which is the largest allowed value for ϵ).

The case $\epsilon = 1/(M-1)$ has an interesting interpretation. Here, the probability $\Pr(y = i | H_j)$ is equal to ϵ if $i \neq j$, and is zero if $i = j$. Thus, an observation $y = i$ provides absolute proof that H_i is not the true hypotheses. If the sensors use decision rules $\gamma \in C^{1, M-1}$ of the form $\gamma(i) = 1$ and $\gamma(j) = 2$, for $j \neq i$, then a message with the value 1 allows the fusion center to eliminate one of the hypotheses. On the other hand, if decision rules in classes other than $C^{1, M-1}$ are used, then the fusion center is not able to make unequivocal inferences. This argument suggests that $C^{1, M-1}$ is the optimal class, as confirmed by our numerical experiments.

VI. DESIGN OF OPTIMAL COMMUNICATION RATE FOR SMALL NOISE CASE

A fundamental design problem is decentralized decision making concerns the choice of the

communication rate (or available bandwidth) between the different decision making units. Such design problems are usually very hard and very little analysis is possible, except for simple situations. For this reason, the solution of even idealized problems can provide valuable intuition. We consider such a design problem, in the context of our decentralized detection problem, under the small noise assumption.

We express the communication rate of each sensor as a function of the variable D . In particular, we view the number $\lceil \log_2 D \rceil$ as the number of binary messages that each sensor must transmit to the fusion center.² Clearly, a higher value of D leads to better performance (smaller probability of error at the fusion center) since a decision is made with more information. On the other hand, communication resources may be scarce, in which case an upper bound can be imposed on the total communication rate in the system. Accordingly, we assume that

$$N \lceil \log_2 D \rceil \leq K \quad (19)$$

where K is a given positive integer. Given such a constraint, we pose the question: “Is it better to have few sensors communicating at high rate, or more sensors communicating at low rate”? We formulate the above described problem in mathematical terms. We view the optimal error exponent Λ^* as a function of D and we use the more suggestive notation $\Lambda^*(D)$. Furthermore, we consider the small noise case for which we can use the approximation (cf. (10) and Proposition 1)

$$\Lambda^*(D) = \log \epsilon \frac{S^*(D)}{M(M-1)} \quad (20)$$

where (cf. (5))

$$S^*(D) = \max_{\delta_D \in \Delta_D} \frac{1}{2} \sum_{d=1}^D \delta_d (M - \delta_d) \quad (21)$$

and Δ_D is the set of all vectors $\delta = (\delta_1, \dots, \delta_D)$ such that each δ_d is a nonnegative integer and $\sum_{d=1}^D \delta_d = M$. Recall that the error probability behaves, asymptotically as $N \rightarrow \infty$, like $e^{N\Lambda^*(D)}$. We are then led to the problem

$$\min_{N, D} N \Lambda^*(D) \quad (22)$$

subject to the constraint (19). (Of course N and D are also constrained to be integers satisfying $N \geq 1$ and $D \geq 2$.)

PROPOSITION 3. *An optimal solution of the problem defined by (19) and (22) is given by $D = 2$, $N = K$.*

²In an alternative formulation we could use $\log_2 D$ instead of $\lceil \log_2 D \rceil$. Which one of these choices is more appropriate could depend on the particular coding method used for transmission. In any case, our subsequent results can be shown to remain valid under this alternative formulation as well.

PROOF. We use (20) and (21) and the fact that $\log \epsilon$ is negative to formulate the problem (22) in the form

$$\max_{N,D} NF(D) \quad (23)$$

where

$$F(D) = \max_{\delta \in \Delta_D} \sum_{d=1}^D \delta_d (M - \delta_d). \quad (24)$$

Let us recall that the optimization problem in the definition of $F(D)$ was solved in the end of Section IV. In particular, it is seen that

$$F(2) = \begin{cases} \frac{M^2}{2}, & \text{if } M \text{ is even} \\ \frac{M^2 - 1}{2}, & \text{if } M \text{ is odd} \end{cases}$$

and

$$F(D) \leq \sum_{d=1}^D \frac{M}{D} \left(M - \frac{M}{D} \right) = M^2 \left(1 - \frac{1}{D} \right), \quad \forall D.$$

(The above inequality is obtained because $\delta_1 = \dots = \delta_D = M/D$ is the optimal solution in (24) when the integrality constraints are relaxed.)

We compare the solution $N = K, D = 2$, with the solution $N = \lfloor K/2 \rfloor, D = 3$. It is easily verified that

$$K \frac{M^2 - 1}{2} \geq \frac{K}{2} M^2 \left(1 - \frac{1}{3} \right), \quad \forall M \geq 2$$

which shows that the solution with $D = 2$ is preferable. Similarly,

$$K \frac{M^2 - 1}{2} \geq \frac{K}{2} M^2 \left(1 - \frac{1}{4} \right), \quad \forall M \geq 2$$

and $D = 2$ is also preferable to $D = 4$. Finally, if $D > 4$, then $\lceil \log_2 D \rceil \geq 3$ and $N \leq K/3$. We have

$$K \frac{M^2 - 1}{2} \geq \frac{K}{3} M^2 \geq \frac{K}{3} M^2 \left(1 - \frac{1}{D} \right), \quad \forall M \geq 2, \quad \forall D > 4,$$

and $D = 2$ is again preferable. Q.E.D.

Generally speaking, intuition suggests that it is better to have several sensors transmitting low rate but independent information, rather than few sensors transmitting detailed information. (One could argue that, other things being equal, independent messages carry more information in the information-theoretic sense, thus leading to smaller error probability; however, it does not seem that this argument can be made rigorous.) Our result (Proposition 3) corroborates this intuition, at least for the particular problem under study. An alternative statement of this result, which is pertinent to organizations involving human decision makers, is the following: if a decision maker is to receive a set of reports of a given total length, it is preferable to receive many partial but

independently drafted reports, rather than a few lengthy ones.

VII. CONCLUSIONS

We have considered the asymptotic (as the number of sensors goes to infinity) solution of a particularly simple symmetric problem in decentralized detection. While the problem is very idealized, the conclusions obtained agree with intuition and could be useful in guiding principles for more general problems. Roughly stated, the following guidelines suggest themselves.

1) It is preferable to have several independent sensors transmitting low rate (coarse) information instead of few sensors transmitting high rate (very detailed) information. (Of course, this guideline is meaningful if it is assumed that the addition of more sensors does not lead to increased "setup" costs; in other words, it is assumed that many sensors are readily available and the only question is whether they can be usefully employed.)

2) An M -ary hypothesis testing problem can be viewed as a collection of $M(M-1)/2$ binary hypothesis testing problem. Under this point of view, the most useful messages by the sensors (decision rules) are those which provide information to the fusion center that is relevant to the largest possible number of these binary hypothesis testing problems. This conclusion has been reached for the small case, because each decision rule is of the "same quality". More generally, one should multiply the number of hypotheses pairs for which a decision rule is helpful by a certain "quality measure" and employ decision rules for which this product is highest.

To what extent the above two guidelines can be verified analytically or experimentally in more realistic problems is an interesting question which is left for further research. Of particular interest is the solution of the same problem when N is finite (as opposed to its asymptotic limit), as well the generalization to problems having less symmetry or to problems with different observation models. Unfortunately, our results are very sensitive to minor changes in problem formulation and different mathematical approaches would be needed to tackle more general problems.

APPENDIX

We outline here a proof that the symmetric solution ($x_i = \frac{1}{3}$, for $i = 1, 2, 3$) is a strict local minimum for the problem considered in the example of Section V. The problem under consideration can be stated as:

$$\Lambda^* = \min_{x \in X} F(x)$$

where

$$F(x) = \max_{i < j} F_{ij}(x) \quad (A1)$$

and

$$F_{ij}(x) = \min_{s \in [0,1]} [x_i \nu(s) + x_j \nu(1-s)]$$

where $i, j \in \{1,2,3\}$. Let $x^* = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$. The function $\nu(\cdot)$ is strictly convex and continuously differentiable, and the minimum in the definition of $F_{ij}(x^*)$ is uniquely attained at $s = \frac{1}{2}$. We can then use Danskin's Theorem [2] to obtain

$$\begin{aligned} \frac{\partial F_{ij}}{\partial x_k}(x^*) &= \frac{\partial}{\partial x_k} \left[x_i \nu\left(\frac{1}{2}\right) + x_j \nu\left(\frac{1}{2}\right) \right] \Big|_{x=x^*} \\ &= \begin{cases} 0, & \text{if } k \neq i \text{ and } k \neq j \\ \nu\left(\frac{1}{2}\right), & \text{if } k = i \text{ or } k = j. \end{cases} \end{aligned}$$

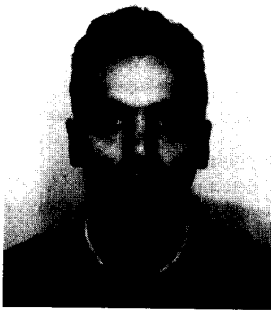
Consider any direction $d \in R^3$, $d \neq 0$, in which x^* can be perturbed without leaving the set X . [That is, $d = (d_1, d_2, d_3)$ with $d_1 + d_2 + d_3 = 0$.] The chain rule yields

$$\begin{aligned} \frac{\partial F_{ij}(x^* + \alpha d)}{\partial \alpha} \Big|_{\alpha=0} &= \sum_{k=1}^3 d_k \frac{\partial F_{ij}(x^*)}{\partial x_k} \\ &= (d_i + d_j) \nu\left(\frac{1}{2}\right). \quad (\text{A2}) \end{aligned}$$

Notice that the assumptions $d \neq 0$, $d_1 + d_2 + d_3 = 0$ imply that there exist some i, j such that $d_i + d_j < 0$. Since $\nu(1/2) < 0$, it follows that for every choice of d , the left-hand side of (A2) is positive for some pair (i, j) . Thus, for each direction d , some function $F_{ij}(x)$ has to increase. Taking (A1) into account, $F(x)$ must also increase. From this point on, it is only a small step to show that $F(x)$ is larger than $F^*(x)$ in a neighborhood of x^* , i.e., that x^* is a local minimum. (The details of this last step are omitted.)

REFERENCES

- [1] Chair, Z., and Varshney, P. K. (1986)
Optimum data fusion in multiple sensor detection systems.
IEEE Transactions on Aerospace and Electronic Systems,
AES-22 (1986), 98-101.
- [2] Danskin, J. M. (1967)
*Theory of Max-Min and its Application in Weapons
Allocation Problems*.
New York: Springer, 1967.
- [3] Ekchian, L. K., and Tenney, R. R. (1982)
Detection networks.
In *Proceedings of the 21st IEEE Conference on Decision
and Control*, (1982), 686-691.
- [4] Hoballah, I. Y., and Varshney, P. K. (1986)
Neyman-Pearson detection with distributed sensors.
In *Proceedings of the 25th IEEE Conference on Decision
and Control*, Athens, Greece, Dec. 1986, 237-241.
- [5] Kushner, H. J., and Pacut, A. (1982)
A simulation study of a decentralized detection problem.
IEEE Transactions on Automatic Control, 27 (1982),
1116-1119.
- [6] Papastavrou, J. D., and Athans, M. (1986)
A distributed hypothesis-testing team decision problem
with communications cost.
In *Proceedings of the 25th IEEE Conference on Decision
and Control*, Athens, Greece, Dec. 1986, 219-225.
- [7] Polychronopoulos, G. H. (1988)
Solution of some problems in decentralized detection by a
large number of sensors.
Technical report LIDS-TH-1781, Laboratory for
Information and Decision Systems, M.I.T., Cambridge,
MA, June 1988.
- [8] Reibman, A. R., and Nolte, L. W. (1987)
Optimal detection and performance of distributed sensor
systems.
IEEE Transactions on Aerospace and Electronic Systems,
AES-23 (1987), 24-30.
- [9] Reibman, A. R., and Nolte, L. W. (1987)
Design and performance comparison of distributed
detection networks.
IEEE Transactions on Aerospace and Electronic Systems,
AES-23 (1987), 789-797.
- [10] Sadjadi, F. (1986)
Hypothesis testing in a distributed environment.
IEEE Transactions on Aerospace and Electronic Systems,
AES-22 (1986), 134-137.
- [11] Shannon, C. E., Gallager, R. G., and Berlekamp, E. R. (1967)
Lower bounds to error probability for coding on discrete
memoryless channels.
Information and Control, 10 (1967), 65-103.
- [12] Srinivasan, R. (1986)
Distributed radar detection theory.
IEE Proceedings, 133 (1986), 55-60.
- [13] Srinivasan, R. (1986)
A theory of distributed detection.
Signal Processing, 11 (1986), 319-327.
- [14] Tenney, R. R., Sandell, N. R., Jr. (1981)
Detection with distributed sensors.
IEEE Transactions on Aerospace and Electronic Systems,
AES-17, 4 (1981), 501-510.
- [15] Teneketzis, D., and Varaiya, P. (1984)
The decentralized quickest detection problem.
IEEE Transactions on Automatic Control, AC-29 (1984),
641-644.
- [16] Thomopoulos, S. C. A., Viswanathan, R., and Bougoulas,
D. P. (1987)
Optimal decision fusion in multiple sensor systems.
IEEE Transactions on Aerospace and Electronic Systems,
AES-23 (1987), 644-653.
- [17] Tsitsiklis, J. N., and Athans, M. (1985)
On the complexity of decentralized decision making and
detection problems.
IEEE Transactions on Automatic Control, AC-30, 5 (1985),
440-446.
- [18] Tsitsiklis, J. N. (1984)
Problems in decentralized decision making and
computations.
Ph.D. dissertation, Dept. of Electrical Engineering and
Computer Science, M.I.T., Cambridge, MA, 1984.
- [19] Tsitsiklis, J. N. (1988)
Decentralized detection by a large number of sensors.
Mathematics of Control, Signals, and Systems, 1, 2 (1988),
167-182.
- [20] Van Trees, H. L. (1968)
Detection Estimation and Modulation Theory, Vol. I.
New York: Wiley, 1968.



George Polychronopoulos received the B.S. Electrical Engineering and Computer Science magna cum laude, from Princeton University in 1985. He then worked for 2 years as a Staff Associate of the Systems and Analysis Group, at the M.I.T.—Lincoln Laboratory, and enrolled in the Operations Research Program at M.I.T. from which he received the M.S. degree, in 1988. He is currently pursuing his Ph.D. at M.I.T. in Operations Research. His interests are in probabilistic and dynamic network optimization, transportation planning, and decentralized decision making.



John N. Tsitsiklis was born in Thessaloniki, Greece, on Feb. 18, 1958. He received the B.S. degree in mathematics (1980), and the B.S. (1980), M.S. (1981) and Ph.D. (1984) degrees in electrical engineering, all from the Massachusetts Institute of Technology, Cambridge, MA.

During the academic year 1983–1984 he was an Acting Assistant Professor of Electrical Engineering at Stanford University, Stanford, CA. Since 1984, he has been with the Electrical Engineering and Computer Science Department at the Massachusetts Institute of Technology, where he is currently Associate Professor. His research interests are in the areas of parallel and distributed computation, systems and control theory, and applied probability.

Dr. Tsitsiklis is the coauthor, with D. Bertsekas, of *Parallel and Distributed Computation: Numerical Methods* (1989). He has been a recipient of an IBM Faculty Development Award (1983), an NSF Presidential Young Investigator Award (1986), an Outstanding Paper Award by the IEEE Control Systems Society (for a paper coauthored with M. Athans, 1986), and of the Edgerton Faculty Achievement Award by M.I.T. (1989). He is also an Associate Editor for *Applied Mathematics Letters*.