Available online at www.sciencedirect.com
SCIENCE DIRECT®
Information Fusion 5 (2004) 157–167
INFORMATION FUSION
www.elsevier.com/locate/inffus

Distributed M-ary hypothesis testing with binary local decisions

Xiaoxun Zhu [a], Yingqin Yuan [b], Chris Rorres [c], Moshe Kam [b,*]

[a] Metrologic Instruments, Inc., Blackwood, NJ 08012, USA
[b] Data Fusion Laboratory, Department of Electrical and Computer Engineering, Drexel University, 3141 Chestnut Street, Philadelphia, PA 19104, USA
[c] Department of Clinical Studies, School of Veterinary Medicine, University of Pennsylvania, Kenneth Square, PA 19348, USA

Received 28 December 2002; received in revised form 20 October 2003; accepted 20 October 2003
Available online 18 November 2003

Abstract

Parallel distributed detection schemes for M-ary hypothesis testing often assume that for each observation the local detector transmits at least $\log_2 M$ bits to a data fusion center (DFC). However, it is possible for less than $\log_2 M$ bits to be available, and in this study we consider 1-bit local detectors with $M > 2$. We develop conditions for asymptotic detection of the correct hypothesis by the DFC, formulate the optimal decision rules for the DFC, and derive expressions for the performance of the system. Local detector design is demonstrated in examples, using genetic algorithm search for local decision thresholds. We also provide an intuitive geometric interpretation for the partitioning of the observations into decision regions. The interpretation is presented in terms of the joint probability of the local decisions and the hypotheses.

Keywords: Decision fusion; Distributed hypothesis testing

1. Introduction

Most studies of parallel distributed detection have been aimed at binary hypothesis testing [2,4,7,10,11,15–17]. When M-ary hypothesis testing was considered, the local detectors (LDs) were often assumed to transmit at least $\log_2 M$ bits to the Data Fusion Center (DFC) for every observation [1,12]. However, the cardinality of the local decisions need not be equal to the number of hypotheses. As Tang and his co-authors observe [14], when the total capacity of the communication channels is fixed and the information quality of each LD is identical, it is better to have a large number of short and independent messages than a smaller number of relatively long messages. Following this observation, we investigate here the effectiveness and performance of an architecture where binary messages are used even when $M > 2$. Approaches to address this problem were suggested in [3,20]. In [20], a hierarchical structure was used to break the complex M-ary decision problem into a set of much simpler binary decision fusion problems, requiring a detector at each node of the decision tree. In [3] an architecture was studied where several binary decisions are fused into a single M-ary decision and processing time constraints need to be satisfied. Our distributed detection system employs N LDs to survey a common volume for evidence of one of M hypotheses ($\{H_i\}_{i=0}^{M-1}$). These LDs are restricted to make a single binary decision per observation, i.e., they have to compress each observation into either "1" or "−1". The DFC uses the local decisions $\mathbf{u} = \{u_j\}_{j=1}^{N} \in \{-1, 1\}^N$ to make a global decision D in favor of one of the M hypotheses. In this context, it is appropriate to model the jth LD, $j \in \{1, 2, \ldots, N\}$, through a set of transition probabilities $\mathscr{R} = \{R_i^j\}_{i=0}^{M-1}$, where R_i^j is the probability that the jth detector transmits "1" to the DFC when the phenomenon H_i was present, namely

$$R_i^j = P\{u_j = 1 | H_i\}. \tag{1}$$

This model of a local detector is shown in Fig. 1.

The DFC is characterized by the set $\mathscr{E} = \{\beta_i\}_{i=0}^{M-1}$, where β_i is the probability that the DFC accepts one of the hypotheses $\{H_k\}_{k \neq i}$ given that phenomenon H_i was present. Thus,

$$\beta_i = \sum_{k=0, k \neq i}^{M-1} P\{D = H_k | H_i\}. \tag{2}$$

* Corresponding author. Tel.: +1-215-895-6920; fax: +1-215-895-1695.
E-mail address: kam@minerva.ece.drexel.edu (M. Kam).

1566-2535/$ - see front matter © 2003 Elsevier B.V. All rights reserved.
doi:10.1016/j.inffus.2003.10.004

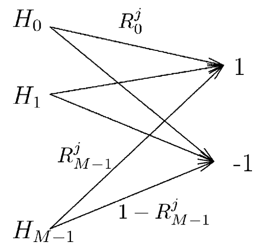

Fig. 1. Transition model (hypothesis-decision) for a local detector.

In (2) $D = H_k$ is used to indicate that the decision (D) of the DFC is to accept the kth hypothesis. The *probability of error* P_e of the DFC is

$$P_e = \sum_{i=0}^{M-1} P(H_i)\beta_i. \tag{3}$$

Our main objective is to find $\{\beta_i\}_{i=0}^{M-1}$ and to determine conditions under which $\lim_{N\to\infty} P_e = 0$. In addition, we discuss (Appendix B) the design of the local decision rules.

2. Optimal design for the DFC

The optimal DFC decision rule that minimizes P_e is

$$\begin{aligned} D &= \operatorname*{argmax}_{H_i}\{P(H_i|\mathbf{u})\} \\ &= \operatorname*{argmax}_{H_i}\left\{\frac{P(\mathbf{u}|H_i)P(H_i)}{P(\mathbf{u})}\right\} \\ &= \operatorname*{argmax}_{H_i}\{P(\mathbf{u}|H_i)P(H_i)\} \\ &= \operatorname*{argmax}_{H_i}\{\log P(\mathbf{u}|H_i) + \log P(H_i)\}. \end{aligned} \tag{4}$$

We set

$$Q_i = \log P(\mathbf{u}|H_i) + \log P(H_i). \tag{5}$$

Under the assumption that the LD observations are conditionally independent (conditioned on the hypothesis), we have

$$\begin{aligned} P(\mathbf{u}|H_i) &= \prod_{j=1}^{N} P(u_j|H_i) = \prod_{j=1}^{N} (R_i^j)^{(1+u_j)/2}(1-R_i^j)^{(1-u_j)/2} \\ &= \prod_{j=1}^{N} \left[R_i^j(1-R_i^j)\right]^{1/2}\left(\frac{R_i^j}{1-R_i^j}\right)^{u_j/2}. \end{aligned}$$

Hence we can rewrite Q_i as

$$\begin{aligned} Q_i &= \log P(H_i) + \frac{1}{2}\sum_{j=1}^{N} \log[R_i^j(1-R_i^j)] \\ &\quad + \frac{1}{2}\sum_{j=1}^{N} u_j \log\left(\frac{R_i^j}{1-R_i^j}\right) = w_i^0 + \sum_{j=1}^{N} w_i^j u_j, \end{aligned}$$

where for $i = 0, 1, \ldots, M-1$, we have set

$$w_i^0 = \log P(H_i) + \frac{1}{2}\sum_{j=1}^{N} \log[R_i^j(1-R_i^j)]$$

and

$$w_i^j = \frac{1}{2}\log\frac{R_i^j}{1-R_i^j} \quad j = 1, 2, \ldots, N.$$

Using the optimal DFC decision rule, it is also possible to rewrite β_i in (2) as

$$\begin{aligned} \beta_i &= \sum_{k=0,k\neq i}^{M-1} P\{D = H_k|H_i\} \\ &= \sum_{k=0,k\neq i}^{M-1} P\{Q_k = \max\{Q_j\}_{j=0}^{M-1}|H_i\} \\ &= \sum_{k=0,k\neq i}^{M-1} \left[\sum_{\mathbf{u}\in\mathcal{U}} P(\mathbf{u}|H_i) P\left\{Q_k = \max\{Q_j\}_{j=0}^{M-1}|\mathbf{u}\right\}\right] \\ &= \sum_{k=0,k\neq i}^{M-1} \left\{\sum_{\mathbf{u}\in\mathcal{U}} P(\mathbf{u}|H_i)\left[\prod_{m=0,m\neq k}^{M-1} \mathrm{U}_{-1}\left\{w_k^0 - w_m^0 \right.\right.\right. \\ &\qquad \left.\left.\left. + \sum_{j=1}^{N}(w_k^j - w_m^j)u_j\right\}\right]\right\}, \end{aligned} \tag{6}$$

where $\mathcal{U} = \{-1, 1\}^N$ is the set of all possible values of $\mathbf{u}$, and $\mathrm{U}_{-1}\{\cdot\}$ is the unit-step function,

$$\mathrm{U}_{-1}\{x\} = \begin{cases} 1 & x > 0, \\ 0 & \text{otherwise.} \end{cases} \tag{7}$$

Thus we possess an optimal (min-P_e) decision rule for the DFC, and an expression ((3) and (6)) for the global performance of this architecture.

3. System performance with identical LDs

We consider a simpler system where all LDs are identical (the case for non-identical LDs is analyzed along a similar path but is somewhat more demanding in bookkeeping and notation). Under this assumption, the local transition probabilities are denoted

$$\{R_i\}_{i=0}^{M-1}, \quad \text{where } R_i = R_i^j \quad \text{for } j = 1, 2, \ldots, N.$$

The error probabilities for the DFC are ($i = 0, 1, \ldots, M-1$)

$$\begin{aligned} \beta_i &= \sum_{k=0,k\neq i}^{M-1} \left\{\sum_{\mathbf{u}\in\mathcal{U}} P(\mathbf{u}|H_i)\left[\prod_{m=0,m\neq k}^{M-1} \mathrm{U}_{-1}\left\{w_k^0 - w_m^0 \right.\right.\right. \\ &\qquad \left.\left.\left. + (w_k - w_m)\sum_{j=1}^{N} u_j\right\}\right]\right\}, \end{aligned} \tag{8}$$

where

$$w_i^0 = \log P(H_i) + \frac{N}{2}\log[R_i(1-R_i)]$$

and

$$w_i = \frac{1}{2}\log\frac{R_i}{1-R_i}.$$

3.1. Decision region for H_i

The binary decisions received by the DFC are governed by a discrete probability distribution function (pdf). Under H_i, each value of $\mathbf{u} \in \mathscr{U}$ has a probability of being realized which depends on R_i. If the local detectors are identical, then

$$P(\mathbf{u} \text{ has } k \text{ “1”s and } N-k \text{ “}-1\text{”s } |H_i) = \binom{N}{k} R_i^k (1-R_i)^{N-k}.$$

Therefore, for each hypothesis H_i, there exists a corresponding binomial distribution of order N. For the distributions to be distinguishable from each other, i.e., for H_i to be distinguished from H_m $(i \neq m)$, there must exist at least one value of $\mathbf{u}$ such that

$$P(\mathbf{u} \text{ has } k \text{ “1”s and } N-k \text{ “}-1\text{”s}|H_i) \neq P(\mathbf{u} \text{ has } k \text{ “1”s and } N-k \text{ “}-1\text{”s}|H_m). \quad (9)$$

For all hypotheses to be distinguishable, (9) must hold for all i and m with $i \neq m$. In the case of identical local detectors, this is equivalent to $R_i \neq R_m$. Furthermore, if any two hypotheses are distinguishable, we can re-index the hypotheses $\{H_i\}_{i=0}^{M-1}$ such that

$$R_0 < R_1 < \cdots < R_{M-1}. \quad (10)$$

3.2. Criterion for accepting H_i

For the DFC to make a decision $D = H_i$, the following must be true:

$$\mathrm{U}_{-1}\left\{w_i^0 - w_m^0 + (w_i - w_m)\sum_{j=1}^{N} u_j\right\} = 1, \quad \forall m \neq i. \quad (11)$$

Suppose that L out of N LDs make the decision 1, while the other $N-L$ LDs make the decision -1. Eq. (11) becomes

$$w_i^0 - w_m^0 + (w_i - w_m)(2L - N) > 0, \quad \forall m \neq i, \quad (12)$$

which can be written as

$$L\log\frac{R_m(1-R_i)}{R_i(1-R_m)} < \log\frac{P(H_i)}{P(H_m)} + N\log\frac{1-R_i}{1-R_m}, \quad \forall m \neq i. \quad (13)$$

Let

$$T_{i,m} = \left(\log\frac{P(H_i)}{P(H_m)} + N\log\frac{1-R_i}{1-R_m}\right)\bigg/\log\frac{R_m(1-R_i)}{R_i(1-R_m)}, \quad \forall m \neq i. \quad (14)$$

We note that $T_{i,m} = T_{m,i}$. If $\frac{R_m}{1-R_m} > \frac{R_i}{1-R_i}$, i.e., $R_m > R_i$, which according to (10) corresponds to $m > i$, Eq. (13) becomes $L < T_{i,m}$; else when $m < i$, Eq. (13) becomes $L > T_{i,m}$.

Let

$$L_i^{\min} = \max_{m=0}^{i-1}\{\lceil T_{i,m}\rceil\}, \quad (15)$$

$$L_i^{\max} = \min_{m=i+1}^{M-1}\{\lfloor T_{i,m}\rfloor\}, \quad (16)$$

where $\lceil x\rceil$ is the smallest integer greater than or equal to x, and $\lfloor x\rfloor$ is the largest integer less than or equal to x. Then if $L_i^{\min} < L_i^{\max}$, the DFC accepts the hypothesis H_i if and only if $L_i^{\min} \leqslant L < L_i^{\max}$. However, if $L_i^{\min} \geqslant L_i^{\max}$, the hypothesis H_i is not detectable by the DFC. In Section 5 we elaborate further on the nature of these “decision intervals”.

3.3. DFC performance and asymptotic properties

For identical LDs, the probability of error under H_i (Eq. (8)) becomes

$$\beta_i = \sum_{k=0,k\neq i}^{M-1}\left\{\sum_{j=L_k^{\min}}^{L_k^{\max}}\binom{N}{j}(R_i)^j(1-R_i)^{N-j}\right\}. \quad (17)$$

Moreover, the *probability of detection* for phenomenon $P(H_i)$, $\zeta_i = P\{D = H_i|H_i\}$, can be written as

$$\zeta_i = \sum_{j=L_i^{\min}}^{L_i^{\max}}\binom{N}{j}(R_i)^j(1-R_i)^{N-j}. \quad (18)$$

Theorem 1 now provides conditions for asymptotic detection of the correct hypothesis by the DFC, namely conditions for the probability of error to go to zero, as the number of sensors, N, goes to infinity.

Theorem 1. *If condition* (10) *holds, then* $\lim_{N\to\infty}\zeta_i = 1$ *and* $\lim_{N\to\infty}\beta_i = 0$ *for* $i = 0, 1, \ldots, M-1$. *The probability of error of the DFC,* $P_e = \sum_{i=0}^{M-1}P(H_i)\beta_i$, *converges to zero at least exponentially as* $N \to \infty$.

The proof is in Appendix A.

4. Examples

4.1. Gaussian populations—same variance, different means

Let the local observations be drawn from one of five Gaussian populations with the same variance ($\sigma^2 = 1$) but different means ($H_0: -2m$, $H_1: -m$, $H_2: 0$, $H_3: m$, and $H_4: 2m$). The observations are statistically independent, and all local detectors are identical. The a priori probabilities are equal ($P(H_0) = P(H_1) = \cdots = P(H_4) = 1/5$).

The local detectors employ the following decision rule based on the observation z:

$$u = \begin{cases} 1 & z > 0, \\ -1 & \text{otherwise.} \end{cases} \tag{19}$$

The DFC employs the optimal decision rule in Eq. (4). We calculate the probability of error (P_e) of the DFC with respect to N, the number of local detectors, for different values of m. Fig. 2 shows P_e.

It is not surprising that when m is small (e.g., $m = 0.5$), the DFC performance is poor. As m grows (e.g., $m = 1.5$), performance improves. It may appear somewhat counter-intuitive that as m increases further (e.g., $m = 2$), the performance of the DFC does not continue to improve. However, as m becomes very large, the LD transition probabilities for certain hypotheses, e.g., R_2 and R_3 become closer in value. As a result, the discrete distributions of the local decisions under hypotheses H_2 and H_3 becomes less and less distinguishable, thus increasing the probability of error at the DFC for large values of m.

4.2. Gaussian populations—different variance, same means

We assume that the observations are drawn from one of four zero-mean Gaussian populations with different variances ($H_0 : \sigma^2 = 1$, $H_1 : 4$, $H_2 : 9$, and $H_3 : 16$). The observations are statistically independent, and all local detectors are identical. The a priori probabilities are equal ($P(H_0) = P(H_1) = \cdots = P(H_3) = 1/4$).

The local detector employs the following decision rule based on the observation z:

$$u = \begin{cases} 1 & z^2 > t, \\ -1 & \text{otherwise.} \end{cases} \tag{20}$$

The DFC employs the optimal decision rule in Eq. (4). We calculate the probability of error (P_e) of the DFC with respect to N, the number of local detectors, for different values of t. Fig. 3 shows that P_e decreases exponentially for different values of t (the graphs tend to a straight line).

When t is small, the transition probabilities for all hypothesis are very close in value (most of the area under the probability density functions is outside the interval $[-\sqrt{t}, \sqrt{t}]$). As t increases, these probabilities become more distinct from each other. However, when t is large enough the area under the probability density functions is mostly confined within $[-\sqrt{t}, \sqrt{t}]$, and the transition probabilities are close in value once again. The resulting degradation in performance is demonstrated (for $t = 6$) in Fig. 3.

4.3. Poisson populations

We assume that the observations are drawn from one of four Poisson populations with different means ($H_0 : m$, $H_1 : 2m$, $H_2 : 3m$ and $H_3 : 4m$). The observations are statistically independent, and all local detectors are identical. The a priori probabilities are equal ($P(H_0) = P(H_1) = \cdots = P(H_3) = 1/4$).

The local detector employs the following decision rule based on the observation z:

$$u = \begin{cases} 1 & z > m/\ln(3/2), \\ -1 & \text{otherwise.} \end{cases} \tag{21}$$

The DFC employs the optimal decision rule in Eq. (4). We calculate the probability of error (P_e) of the DFC with respect to N, the number of local detectors, for different values of m. Fig. 4 shows that P_e decreases

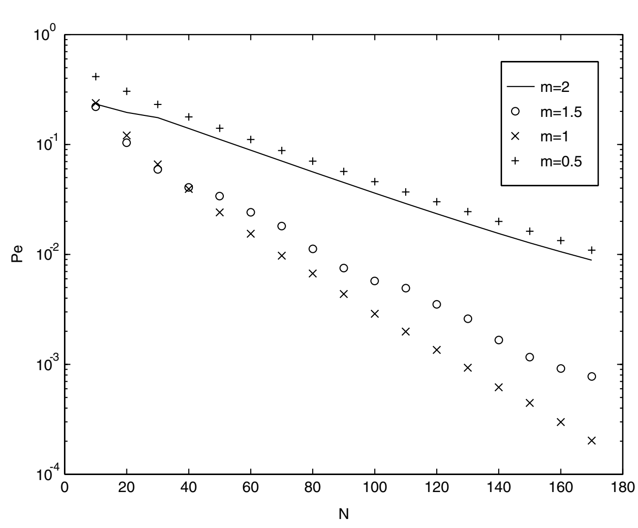

Fig. 2. Probability of error (P_e) vs. N (example 4.1).

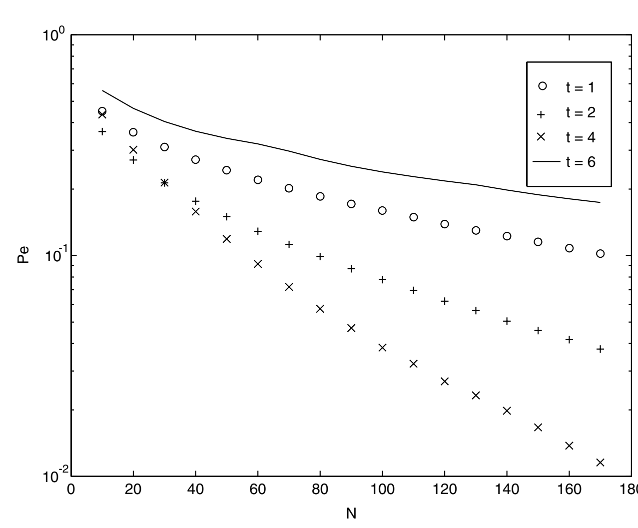

Fig. 3. Probability of error (P_e) vs. N (example 4.2).

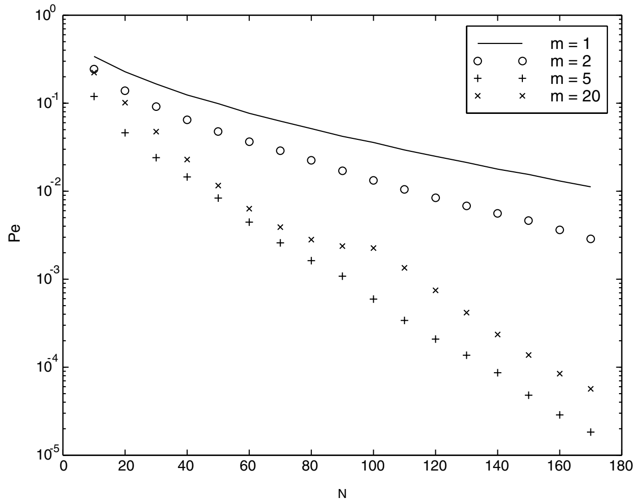

Fig. 4. Probability of error (P_e) vs. N (example 4.3).

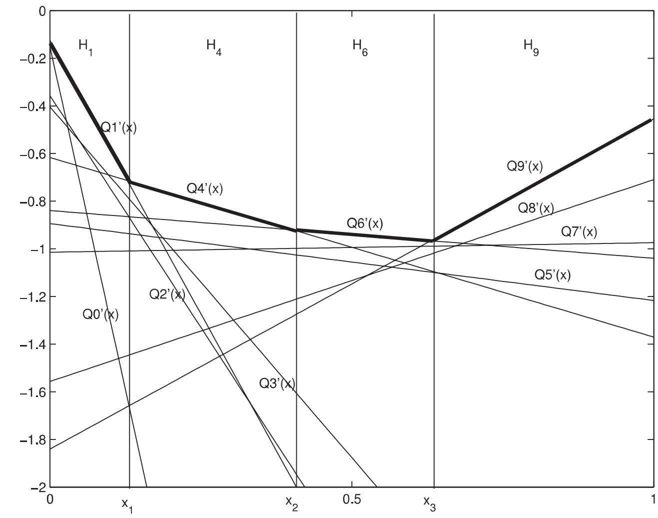

Fig. 5. A typical decision region diagram (here $M = N = 10$).

exponentially for different values of m (the graphs tend to a straight line).

Similar to example 4.1, when m is small, the DFC performance is poor. It improves as m increases to 2 and 5. However, as m becomes larger (e.g., $m = 20$), some of the LD transition probabilities (e.g., those of H_2 and H_3) become so close that the DFC performance start to degrade again.

5. Decision region for H_i—geometric interpretation

In this section we discuss an intuitive geometric interpretation of the partition of decision regions for each hypothesis in $\{H_i\}_{i=0}^{M-1}$. From (5), we have

$$Q'_i = \frac{Q_i}{N} = \frac{1}{N}\log P(H_i) + \frac{L}{N}\log R_i + \left(1 - \frac{L}{N}\right)\log(1 - R_i), \quad i = 0, \ldots, M-1. \tag{22}$$

Given the number L of LDs that make the decision 1, our optimal decision rule is to select H_i that corresponds to the largest Q'_i.

Let us denote by x the fraction of LDs that make the decision 1; that is, $x = L/N$, $0 \leqslant x \leqslant 1$. Then (22) can be written as

$$Q'_i(x) = x\log\left(\frac{R_i}{1 - R_i}\right) + \frac{1}{N}\log P(H_i) + \log(1 - R_i), \quad i = 0, \ldots, M-1, \tag{23}$$

where we now write $Q'_i(x)$ to emphasize the dependence of Q'_i on x. For each i, $Q'_i(x)$ is a linear function of x with slope $\log[R_i/(1 - R_i)]$. When the probabilities R_i are ordered as in (10) these slopes are correspondingly ordered by

$$\log\left(\frac{R_1}{1 - R_1}\right) < \log\left(\frac{R_2}{1 - R_2}\right) < \cdots < \log\left(\frac{R_{M-1}}{1 - R_{M-1}}\right). \tag{24}$$

Fig. 5 is a typical diagram for ten LDs showing the ten linear functions $\{Q'_0(x), Q'_1(x), \ldots, Q'_9(x)\}$ plotted as functions of x for the case $M = N = 10$. Let us denote by $Q'_{\max}(x)$ the function

$$Q'_{\max}(x) = \max_x\{Q'_0(x), Q'_1(x), \ldots, Q'_{M-1}(x)\}. \tag{25}$$

As can be seen from Fig. 5, this function is a continuous, piecewise linear, convex-up function of x over the interval $[0, 1]$. In the particular example illustrated, only four of the ten functions, (namely, $Q'_1(x), Q'_4(x), Q'_6(x)$ and $Q'_9(x)$), enter into the formation of $Q'_{\max}(x)$. Likewise, only the four corresponding hypothesis (H_1, H_4, H_6 and H_9), can be distinguished by the DFC depending on which of four intervals $[0, x_1), [x_1, x_2), [x_2, x_3)$ and $[x_3, 1]$ the fraction x lies in. We remark, however, that because x can only assume the discrete values in the set $\{0, 1/N, 2/N, \ldots, (N-1)/N, 1\}$, it is possible that x may not be able to fall in the one of the non-empty intervals determined by the function $Q'_{\max}(x)$. This would further restrict the set of hypotheses that could be distinguished by the DFC.

There are two situations in which all M hypotheses can be potentially distinguished by the DFC. The first situation is when the probabilities $P(H_i)$ of the M hypotheses are all equal to each other. Letting the common probability be P^*, (23) reduces to

$$Q'_i(x) = x\log\left(\frac{R_i}{1 - R_i}\right) + \frac{1}{N}\log P^* + \log(1 - R_i), \quad i = 0, \ldots, M-1. \tag{26}$$

As can be verified, each of the straight lines represented by (26) is tangent to the function

$$E(x) = x\log\left(\frac{x}{1-x}\right) + \frac{1}{N}\log P^* + \log(1-x) \tag{27}$$

at $x = R_i$. In other words, $E(x)$ is the *envelope* of any one of the linear functions $Q_i'(x)$ as the parameter R_i varies over all values between 0 and 1. As indicated in Fig. 6, $E(x)$ is a smooth convex-up function over the interval $[0,1]$, and so each of the straight lines determined by each $Q_i'(x)$ lies below the function $E(x)$ for all $x \in [0,1]$. This means that for any particular index i, the function $Q_{\max}'$ must equal $Q_i'(x)$ for all x in some neighborhood of the point of tangency $x = R_i$.

More specifically, because of the ordering $R_0 < R_1 < \cdots < R_{M-1}$, the interval $[0,1]$ can be partitioned into M non-empty subintervals, $[0,x_1),[x_1,x_2),\ldots,[x_{M-2},x_{M-1})$ and $[x_{M-1},1]$, such that

(1) $x_i < R_i < x_{i+1}, \quad i = 1,\ldots M-1,$
where $x_0 = 0$ and $x_M = 1$.

(2) $$Q_{\max}'(x) = \begin{cases} Q_i'(x), & x \in [x_i, x_{i+1}), \quad 0 \leqslant i < M-1, \\ Q_{M-1}'(x), & x \in [x_{M-1}, 1]. \end{cases} \tag{28}$$

The optimal decision rule is then to choose hypothesis H_i if $x \in [x_i, x_{i+1})$, for some $i = 0,\ldots,M-2$, or hypothesis H_{M-1} if $x \in [x_{M-1},1]$. In this way, hypothesis H_i is selected by the DFC if the fraction of the LDs that register 1 is in the subinterval containing R_i. As before, however, it is possible that none of the $N+1$ possible values of the discrete variables x lies in some of these subintervals, and so the corresponding hypothesis cannot be selected.

The second situation in which all M hypotheses can be distinguished is when the number of detectors N is sufficiently large so that the term $\frac{1}{N}\log P(H_i)$ in (23) can be neglected. We then have

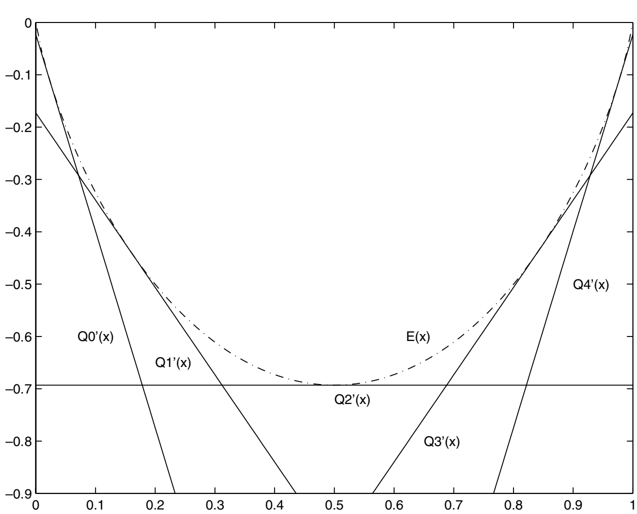

Fig. 6. Envelope $E(x)$ and the family of functions Q_i' $(N \to \infty)$.

$$Q_i' \approx x\log\left(\frac{R_i}{1-R_i}\right) + \log(1-R_i), \quad i = 0,\ldots,M-1. \tag{29}$$

Each of the straight lines in (29) is tangent to the curve

$$E(x) = x\log\left(\frac{x}{1-x}\right) + \log(1-x), \tag{30}$$

which is simply a vertical translation of the curve (27). Fig. 6 thus serves to also illustrate the case when the number of the local detectors is large. Consequently, all M hypotheses can be distinguished if N is sufficiently large and, furthermore, a sufficiently large N guarantees that the $N+1$ possible values of the discrete variable $x = L/N$ will be distributed among all of the M intervals that distinguish the hypothesis.

5.1. Examples—Gaussian populations, same variance, different means

We provide a geometric interpretation of our decision rule, using data from example 4.1, where the local observations were drawn from one of five Gaussian populations with the same variance ($\sigma^2 = 1$) but different means ($H_0 : -2m$, $H_1 : -m$, $H_2 : 0$, $H_3 : m$, and $H_4 : 2m$). The observations were statistically independent, and all local detectors were identical. The a priori probabilities were equal ($P(H_0) = P(H_1) = \cdots = P(H_4) = 1/5$).

Fig. 7 shows the decision regions for hypotheses H_0, H_2 and H_4, with $m = 1$, using graphs of $E(x)$ and $Q_i'(x)$.

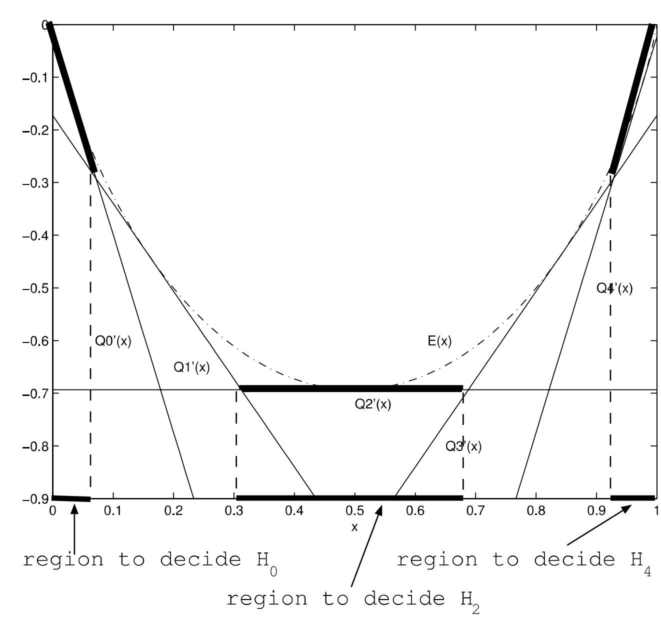

Fig. 7. Decision region illustration ($m = 1$).

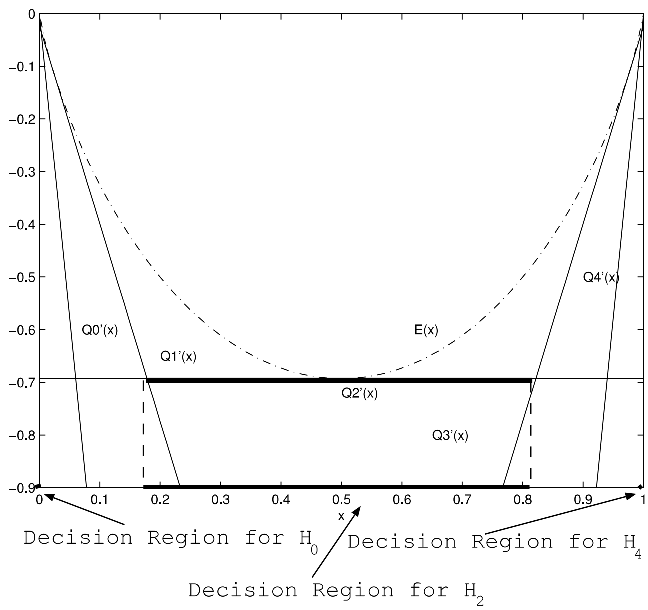

Fig. 8. Decision region illustration ($m = 2$).

We observed before that when m increases, the performance of the DFC does not necessarily always improve (e.g., $m = 2$). Fig. 8 demonstrates that as R_3 and R_4 become closer (and very small) in value, the decision region for H_4 becomes very small, causing the optimal decision rule to choose H_3 over H_4 most of the time. Similarly, H_1 is usually chosen rather than H_0. The observed increase in the probability of error at the DFC (with the increase of m) corresponds to the diminishing decision regions for certain hypotheses.

6. Conclusion

We studied a distributed detection system that performs M-ary hypothesis testing with identical LDs making binary decisions. We showed that when the local detectors compress the observation into a single bit, the data fusion center (DFC) is able to distinguish the hypotheses provided that the distribution functions of the local decisions are not identical. We also showed that the probability of error decreases exponentially as the number of local detectors increases. Practical examples were given for discovering the local decision rules using genetic algorithm searches for local thresholds.

Additional investigation is needed to determine the exact trade-off between compression level (i.e., number of bits per observation) and hardware complexity, namely, the number of sensors required to achieve a certain error rate. This is tied to the design of the local decision rules, since the transition probabilities determine the rate of convergence of the probability of error.

Appendix A. Proof of Theorem 1

Lemma 1. *If p and r are such that $p \neq r$ and $0 < p, r < 1$, then*

$$\left(\frac{r}{p}\right)^p \left(\frac{1-r}{1-p}\right)^{1-p} < 1. \tag{A.1}$$

Proof. Let

$$F(p,r) = p\log\frac{r}{p} + (1-p)\log\frac{1-r}{1-p}. \tag{A.2}$$

Thus,

$$\frac{\partial}{\partial p}F(p,r) = \log\frac{r}{p} - \log\frac{1-r}{1-p}, \tag{A.3}$$

$$\frac{\partial^2}{\partial p^2}F(p,r) = -\frac{1}{p} - \frac{1}{1-p}. \tag{A.4}$$

This indicates that $F(p,r)$ has maxima where

$$\log\frac{r}{p} - \log\frac{1-r}{1-p} = 0. \tag{A.5}$$

It is easy to show that the maxima are at $p = r$, $0 < p, r < 1$.

Also, for any given r, $0 < r < 1$, we note from (A.3) that $\frac{\partial}{\partial p}F(p,r) > 0$ for $p < r$, and $\frac{\partial}{\partial p}F(p,r) < 0$ for $p > r$. Hence the maxima at $p = r$ is also a global maxima for $0 < p, r < 1$. Therefore, $\forall p \neq r$, $0 < p, r < 1$, $F(p,r) < F(p,p) = 0$. Eq. (A.1) follows. □

Lemma 2. *Given any $r \in (0,1)$ and p_1, p_2 such that*

1. $1 > p_1 > p_2 > r > 0$; *or*
2. $1 > r > p_1 > p_2 > 0$,

it follows that

$$\frac{\log\frac{p_1}{r}}{\log\frac{1-r}{1-p_1}} > \frac{\log\frac{p_2}{r}}{\log\frac{1-r}{1-p_2}}. \tag{A.6}$$

Proof. Let $1 > p > r > 0$, and

$$f(p,r) = \log\frac{p}{r} \Big/ \log\frac{1-r}{1-p} \tag{A.7}$$

and

$$g(p,r) = \frac{\partial f(p,r)}{\partial p} = \left[(1-p)\log\frac{1-r}{1-p} - p\log\frac{p}{r}\right] \Big/ \left[p(1-p)\left(\log\frac{1-r}{1-p}\right)^2\right].$$

From Lemma 1, $g(p,r) < 0$ for $1 > p > r > 0$. This implies that $\forall p_1, p_2$, $1 > p_1 > p_2 > r > 0$, Eq. (A.6) is true. Similarly, for $1 > r > p > 0$, it is easy to show that $g(p,r) < 0$. Therefore Eq. (A.6) is also true, $\forall p_1, p_2$, $1 > r > p_1 > p_2 > 0$. □

Lemma 3. *If condition* (10) *holds, then* $\forall k$, $m,n \in \{0,1,\ldots,M-1\}$, $\exists N_{m,n}^k \in \Re$, *such that* $\forall N > N_{m,n}^k$, $T_{k,m} > T_{k,n}$, *provided that either* $m > n > k$ *or* $k > m > n$.

Proof. Let

$$N_{m,n}^k = \left(\frac{\log \dfrac{P(H_k)}{P(H_m)}}{\log \dfrac{R_m(1-R_k)}{R_k(1-R_m)}} - \frac{\log \dfrac{P(H_k)}{P(H_n)}}{\log \dfrac{R_n(1-R_k)}{R_k(1-R_n)}} \right) \Bigg/ \left(\frac{1}{1+\log \dfrac{R_m}{R_k} \Big/ \log \dfrac{1-R_k}{1-R_m}} - \frac{1}{1+\log \dfrac{R_n}{R_k} \Big/ \log \dfrac{1-R_k}{1-R_n}} \right). \tag{A.8}$$

Since either $m > n > k$ or $k > m > n$ is true, from Eq. (10), we have either $R_m > R_n > R_k$ or $R_k > R_m > R_n$. From Lemma 2,

$$\frac{1}{1+\log \frac{R_m}{R_k} / \log \frac{1-R_k}{1-R_m}} > \frac{1}{1+\log \frac{R_n}{R_k} / \log \frac{1-R_k}{1-R_n}}. \tag{A.9}$$

Provided that $N > N_{m,n}^k$, from Eqs. (A.8) and (14) we have $T_{k,m} > T_{k,n}$. □

Lemma 4. *If condition* (10) *holds, then* $\forall k \in \{1,\ldots,M-1\}$, $\exists N_{\min}^k$, *such that* $\forall N > N_{\min}^k$, $L_k^{\min} = \lceil T_{k,k-1} \rceil$; $\forall k \in \{0,\ldots,M-2\}$, $\exists N_{\max}^k$, *such that* $\forall N > N_{\max}^k$, $L_k^{\max} = \lfloor T_{k,k+1} \rfloor$.

Proof. Let $N_{\min}^k = \max_{m,n=0}^{k-1} N_{m,n}^k$ and $N_{\max}^k = \min_{m,n=k+1}^{M-1} N_{m,n}^k$, where $N_{m,n}^k$ is as defined in Eq. (A.8). This lemma follows from Lemma 3 and Eqs. (15) and (16). □

Proof of Theorem 1. Using the DeMoivre–Laplace Theorem [9, pp. 49–50],

$$\lim_{N\to\infty} \zeta_i = \lim_{N\to\infty} G\left(\frac{L_i^{\max} - NR_i}{\sqrt{NR_i(1-R_i)}} \right) - \lim_{N\to\infty} G\left(\frac{L_i^{\min} - NR_i}{\sqrt{NR_i(1-R_i)}} \right), \tag{A.10}$$

where

$$G(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \mathrm{e}^{-y^2/2}\, \mathrm{d}y. \tag{A.11}$$

When N is sufficiently large (Lemma 4),

$$L_i^{\max} = \lfloor T_{i,i+1} \rfloor > T_{i,i+1} - 1, \tag{A.12}$$

$$L_i^{\min} = \lceil T_{i,i-1} \rceil < T_{i,i-1} + 1. \tag{A.13}$$

Therefore

$$\frac{L_i^{\max} - NR_i}{\sqrt{NR_i(1-R_i)}} > \frac{1}{\sqrt{R_i(1-R_i)}} [(A_{i,i+1} - 1)/\sqrt{N} + B_{i,i+1}\sqrt{N}], \tag{A.14}$$

$$\frac{L_i^{\min} - NR_i}{\sqrt{NR_i}(1-R_i)} < \frac{1}{\sqrt{R_i(1-R_i)}} [(A_{i,i-1} + 1)/\sqrt{N} + B_{i,i-1}\sqrt{N}], \tag{A.15}$$

where

$$A_{k,m} = \frac{\log \dfrac{P(H_k)}{P(H_m)}}{\log \dfrac{R_m(1-R_k)}{R_k(1-R_m)}} \tag{A.16}$$

and

$$B_{k,m} = \frac{1}{1+\log \frac{R_m}{R_k} \Big/ \log \frac{1-R_k}{1-R_m}} - R_k. \tag{A.17}$$

From Lemma 1 and $R_{i+1} > R_i$,

$$\frac{1}{1+\log \frac{R_{i+1}}{R_i} \Big/ \log \frac{1-R_i}{1-R_{i+1}}} > R_i. \tag{A.18}$$

Therefore,

$$B_{i,i+1} > 0. \tag{A.19}$$

Since $A_{i,i+1}$ is bounded,

$$\lim_{N\to\infty} G\left(\frac{1}{\sqrt{R_i(1-R_i)}} \left[(A_{i,i+1} - 1)/\sqrt{N} + B_{i,i+1}\sqrt{N} \right] \right) = 1. \tag{A.20}$$

From Eq. (A.14),

$$\lim_{N\to\infty} G\left(\frac{L_i^{\max} - NR_i}{\sqrt{NR_i(1-R_i)}} \right) = 1. \tag{A.21}$$

Similarly, since $R_i > R_{i-1}$, one can show from Lemma 1 that

$$\frac{1}{1+\log \frac{R_{i-1}}{R_i} \Big/ \log \frac{1-R_i}{1-R_{i-1}}} < R_i. \tag{A.22}$$

Hence

$$B_{i,i-1} < 0. \tag{A.23}$$

Since $A_{i,i-1}$ is bounded,

$$\lim_{N\to\infty} G\left(\frac{1}{\sqrt{R_i(1-R_i)}} \left[(A_{i,i-1} + 1)/\sqrt{N} + B_{i,i-1}\sqrt{N} \right] \right) = 0. \tag{A.24}$$

From Eq. (A.15),

$$\lim_{N\to\infty} G\left(\frac{L_i^{\min} - NR_i}{\sqrt{NR_i(1-R_i)}}\right) = 0. \tag{A.25}$$

From Eqs. (A.21) and (A.25),

$$\lim_{N\to\infty} \zeta_i = 1, \quad i = 0, 1, \ldots, M-1. \tag{A.26}$$

From Eq. (A.10),

$$\lim_{N\to\infty} \beta_i = 1 - \lim_{N\to\infty} G\left(\frac{L_i^{\max} - NR_i}{\sqrt{NR_i(1-R_i)}}\right) + \lim_{N\to\infty} G\left(\frac{L_i^{\min} - NR_i}{\sqrt{NR_i(1-R_i)}}\right). \tag{A.27}$$

Use the inequalities in Eqs. (A.14) and (A.15),

$$\lim_{N\to\infty} \beta_i \leqslant \lim_{N\to\infty} \epsilon_1(N) + \lim_{N\to\infty} \epsilon_2(N), \tag{A.28}$$

where

$$\epsilon_1(N) = 1 - G\left(\frac{1}{\sqrt{R_i(1-R_i)}}\left[(A_{i,i+1} - 1)/\sqrt{N} + B_{i,i+1}\sqrt{N}\right]\right) \tag{A.29}$$

and

$$\epsilon_2(N) = G\left(\frac{1}{\sqrt{R_i(1-R_i)}}\left[(A_{i,i-1} + 1)/\sqrt{N} + B_{i,i-1}\sqrt{N}\right]\right). \tag{A.30}$$

Using the property $G(X) = 1 - G(-X)$, we can rewrite $\epsilon_2(N)$

$$\epsilon_2(N) = 1 - G\left(\frac{1}{\sqrt{R_i(1-R_i)}}\left[(-A_{i,i-1} - 1)/\sqrt{N} - B_{i,i-1}\sqrt{N}\right]\right). \tag{A.31}$$

When N is large enough such that

$$N > -\frac{A_{i,i+1} + A_{i,i-1}}{B_{i,i+1} + B_{i,i-1}}, \tag{A.32}$$

if $B_{i,i+1} + B_{i,i-1} < 0$, Eq. (A.32) yields

$$(A_{i,i+1} - 1)/\sqrt{N} + B_{i,i+1}\sqrt{N} < (-A_{i,i-1} - 1)/\sqrt{N} - B_{i,i-1}\sqrt{N}, \tag{A.33}$$

otherwise,

$$(A_{i,i+1} - 1)/\sqrt{N} + B_{i,i+1}\sqrt{N} > (-A_{i,i-1} - 1)/\sqrt{N} - B_{i,i-1}\sqrt{N}. \tag{A.34}$$

Therefore, since $1 - G(x)$ is monotonically decreasing for $x > 0$,

$$\lim_{N\to\infty} \beta_i < \begin{cases} 2\lim_{N\to\infty} \epsilon_1(N) & B_{i,i+1} + B_{i,i-1} < 0, \\ 2\lim_{N\to\infty} \epsilon_2(N) & \text{otherwise.} \end{cases} \tag{A.35}$$

From [19, p. 39],

$$1 - G(X) < \frac{1}{\sqrt{2\pi}X} e^{-X^2/2}, \quad X > 0. \tag{A.36}$$

Therefore

$$\begin{aligned} \epsilon_1(N) &< \frac{\sqrt{R_i(1-R_i)}}{\sqrt{2\pi}\left[(A_{i,i+1} - 1)/\sqrt{N} + B_{i,i+1}\sqrt{N}\right]} \exp\left\{-\frac{1}{2R_i(1-R_i)}\left[(A_{i,i+1} - 1)/\sqrt{N} + B_{i,i+1}\sqrt{N}\right]^2\right\} \\ &= \frac{\sqrt{R_i(1-R_i)}}{\sqrt{2\pi}\left[(A_{i,i+1} - 1)/\sqrt{N} + B_{i,i+1}\sqrt{N}\right]} \exp\left\{-\frac{1}{2R_i(1-R_i)}\left[(A_{i,i+1} - 1)^2/N + B_{i,i+1}^2 N + 2B_{i,i+1}(A_{i,i+1} - 1)\right]\right\} \\ &< \frac{\sqrt{R_i(1-R_i)}}{\sqrt{2\pi}\left[(A_{i,i+1} - 1)/\sqrt{N} + B_{i,i+1}\sqrt{N}\right]} \exp\left\{-\frac{1}{2R_i(1-R_i)}\left[B_{i,i+1}^2 N + 2B_{i,i+1}(A_{i,i+1} - 1)\right]\right\}. \end{aligned} \tag{A.37}$$

Similarly,

$$\begin{aligned} \epsilon_2(N) &< \frac{\sqrt{R_i(1-R_i)}}{\sqrt{2\pi}\left[(-A_{i,i-1} - 1)/\sqrt{N} - B_{i,i-1}\sqrt{N}\right]} \exp\left\{-\frac{1}{2R_i(1-R_i)}\left[(-A_{i,i-1} - 1)/\sqrt{N} - B_{i,i-1}\sqrt{N}\right]^2\right\} \\ &= \frac{\sqrt{R_i(1-R_i)}}{\sqrt{2\pi}\left[(-A_{i,i-1} - 1)/\sqrt{N} - B_{i,i-1}\sqrt{N}\right]} \exp\left\{-\frac{1}{2R_i(1-R_i)}\left[(A_{i,i-1} + 1)^2/N + B_{i,i-1}^2 N + 2B_{i,i-1}(A_{i,i-1} + 1)\right]\right\} \\ &< \frac{\sqrt{R_i(1-R_i)}}{\sqrt{2\pi}\left[(-A_{i,i-1} - 1)/\sqrt{N} - B_{i,i-1}\sqrt{N}\right]} \exp\left\{-\frac{1}{2R_i(1-R_i)}\left[B_{i,i-1}^2 N + 2B_{i,i-1}(A_{i,i-1} + 1)\right]\right\}. \end{aligned} \tag{A.38}$$

Hence,

$$\beta_i = \mathrm{O}(\mathrm{e}^{-N}/\sqrt{N}) \tag{A.39}$$

and

$$\lim_{N\to\infty} \beta_i = 0, \quad i = 0, 1, \ldots, M-1. \tag{A.40}$$

From Eq. (A.40),

$$\lim_{N\to\infty} P_\mathrm{e} = \sum_{i=0}^{M-1} P(H_i) \lim_{N\to\infty} \beta_i = 0. \tag{A.41}$$

Since β_i decreases at least exponentially for $i = 0, 1, \ldots, M-1$, P_e, a linear combination of these β_i's, also decreases at least exponentially. □

Appendix B. Design of local decision rules

For many decentralized detection problems, including the one studied here, determination of the optimal local decision rules was shown to be NP-complete [18]. The necessary conditions of the optimum are often described as a set of coupled non-linear equations that are extremely difficult to solve [5,15,20]. Several numerical methods were proposed to approximate the optimal local decision rules for such systems, including variants of Gauss–Seidel algorithm [14,21], and the use of genetic algorithms (GA) [8]. Though "most of the problems analyzed in the literature have been found to have globally optimal solutions in which each sensor uses the same threshold" [6], this is not the general case. In this Appendix B we demonstrate how the local decision rules for our architecture can be approximated, using genetic algorithm search. The best solutions that we found are for non-identical LDs.

We assume that the local detector compares each scalar local observation $z_j, j = 1, 2, \ldots, N$, to a threshold (or thresholds) in order to determine the local decision u_j (see [6], Section III). We search numerically for a minimum probability of error P_e (as defined in Eq. (3)) in terms of the threshold(s) of each LD, assuming that we know the probability density function of the local observation z_j and the a priori probabilities of the hypotheses. In general, P_e is a non-continuous non-differentiable function of the local thresholds which makes gradient-based optimizing algorithms ineffective. We therefore used genetic algorithms (GAs) for the optimization task [8,13]. The GA sought a local minimum of P_e in N thresholds, where N is the number of LDs. In the following, we provide two examples (corresponding to the examples in Sections 4.1 and 4.2) for design of the local decision rules. The calculations were made for different number of sensors (2–7), using numerical search for the decision thresholds. Our GA [13] used a varying crossover rate, and a constant mutation rate of 0.12. The search was terminated either when the number of iterations reached 20,000 or the improvement in the probability of error over the last 100 iterations was less than 10^{-4}.

B.1. Example B-1: Gaussian populations—same variance, different means

The hypothesis testing problem has four equi-probable hypotheses. Under hypothesis i ($i = 1, \ldots, 4$), the observations are Gaussian with mean m_i and standard deviation σ_i. Specifically, $m_1 = -1$, $m_2 = 0$, $m_3 = 5$ and $m_4 = 7$, and $\sigma_i = 1$ for all i. Following [6] we assume that the ith LD uses the rule described in Eq. (19). Fig. 9 shows the global probability of error for the case of two (2) LDs. Not surprisingly the optima occur when the LDs are not identical, and we find two distinct global minima (we can permute the values of the two thresholds between LDs). Table 1 presents the calculated

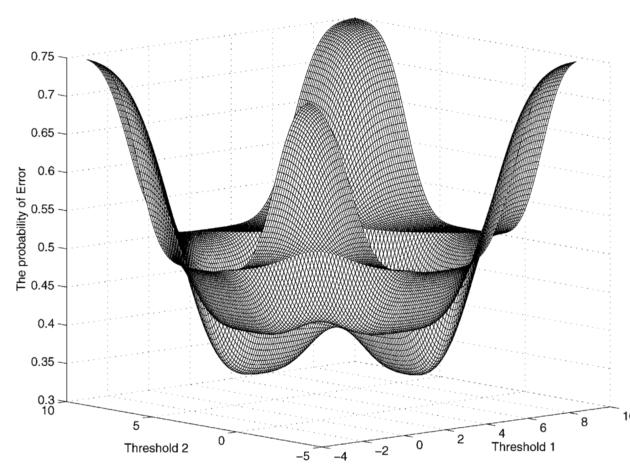

Fig. 9. Probability of error (P_e) vs. thresholds in a two-sensor case (example B-1).

Table 1
Optimization results using identical and non-identical thresholds

No. of sensors (N)	Identical thresholds		Non-identical thresholds	
	Threshold	Error probability	Thresholds	Error probability
2	4.880	0.3846	1.9750, 6.0281	0.3300
3	4.880	0.3267	−0.5065, 2.4871, 5.9247	0.2356
4	4.850	0.2993	−0.5348, 3.1490, 5.1615, 5.6864	0.2149
5	4.800	0.2865	−0.0628, 0.3386, 2.0482, 5.3941, 5.5808	0.1935
6	5.200	0.2767	−0.2424, 0.2868, 0.4751, 2.5893, 5.7135, 6.1509	0.1748
7	5.150	0.2671	−0.7008, −0.6992, −0.0608, 0.9873, 5.5744, 6.1316, 6.2272	0.1572

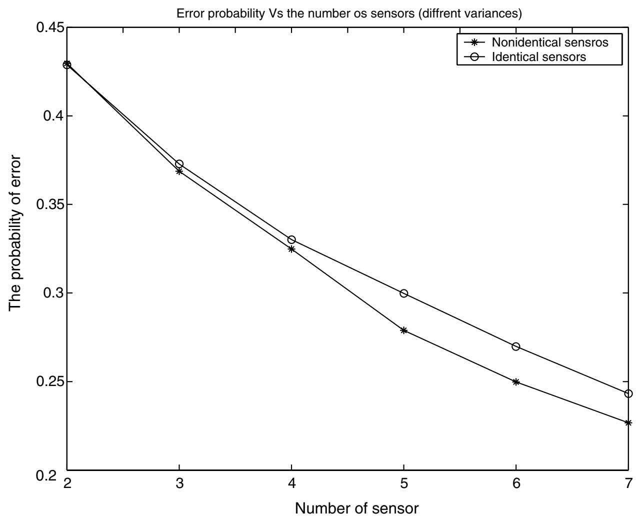

Fig. 10. Probability of error vs. number of sensors (example B-2).

thresholds for 2–7 LDs. The table also shows the solution for identical LDs. Clearly, identical LDs perform much worse than non-identical LDs.

B.2. Example B-2: Gaussian populations—different variance, same means

Following the notation in example B-1, we tested four Gaussian hypotheses. Here, $m_1 = m_2 = m_3 = m_4 = 0$ and $\sigma_1 = 1, \sigma_2 = 4, \sigma_3 = 10, \sigma_4 = 20$. The local decision rule is defined by Eq. (20). Fig. 10 shows the probability of error vs. the number of sensors, using the local thresholds found by the GA. Again, non-identical sensors provide better performance than identical sensors. However, threshold search for a large number of sensors can become computationally expensive. In this case, the results of Theorem 1 encourage the use of identical sensors.

References

[1] W. Baek, S. Bommareddy, Optimal M-ary data fusion with distributed sensors, IEEE Transactions on Aerospace and Electronic Systems 31 (3) (1995) 1150–1152.
[2] Z. Chair, P.K. Varshney, Optimal data fusion in multiple sensor detection systems, IEEE Transactions on Aerospace and Electronic Systems 22 (1) (1986) 98–101.
[3] B.V. Dasarathy, Operationally efficient architecture for fusion of binary-decision sensors in multidecision environments, Optical Engineering 36 (3) (1997) 632–641.
[4] R.O. Duda, P.E. Hart, Pattern Classification and Scene Analysis, Wiley, New York, 1973.
[5] I.Y. Hoballah, P.K. Varshney, Distributed Bayesian signal detection, IEEE Transactions on Information Theory 35 (5) (1989) 995–1000.
[6] W.W. Irving, J.N. Tsitsiklis, Some properties of optimal thresholds in decentralized detection, IEEE Transactions on Automatic Control 39 (4) (1994) 835–838.
[7] M. Kam, W. Chang, Q. Zhu, Hardware complexity of binary distributed detection systems with isolated local Bayesian detectors, IEEE Transactions on Systems Man and Cybernetics 21 (3) (1991) 713–725.
[8] W. Liu, Y. Lu, J.S. Fu, Data fusion of multiradar system by using genetic algorithm, IEEE Transactions on Aerospace and Electronic Systems 38 (2) (2002) 601–612.
[9] A. Papoulis, Probability, Random Variables, and Stochastic Processes, third ed., McGraw-Hill, New York, 1991.
[10] A.R. Reibman, L.W. Nolte, Design and performance comparison of distributed sensor systems, IEEE Transactions on Aerospace and Electronic Systems 23 (6) (1987) 789–797.
[11] A.R. Reibman, L.W. Nolte, Optimal detection and performance of distributed sensor systems, IEEE Transactions on Aerospace and Electronic Systems 23 (1) (1987) 24–30.
[12] F.A. Sadjadi, Hypothesis testing in a distributed environment, IEEE Transactions on Aerospace and Electronic Systems 22 (2) (1986) 134–137.
[13] K.S. Tang, K.F. Man, Q. Kwong, S. He, Genetic algorithms and their applications, IEEE Signal Processing Magzine 13 (6) (1996) 22–37.
[14] Z.B. Tang, K.R. Pattipati, D.L. Kleinman, A distributed M-ary hypothesis testing problem with correlated observations, IEEE Transactions on Automatic Control 37 (7) (1992) 1042–1046.
[15] R.R. Tenney, N.R. Sandell, Detection with distributed sensors, IEEE Transactions on Aerospace and Electronic Systems 17 (4) (1981) 501–509.
[16] S.C. Thomopoulos, R. Viswanathan, D.K. Bougoulias, Optimal decision fusion in multiple sensor systems, IEEE Transactions on Aerospace and Electronic Systems 23 (5) (1987) 644–653.
[17] S.C. Thomopoulos, R. Viswanathan, D.K. Bougoulias, Optimal distributed decision fusion, IEEE Transactions on Aerospace and Electronic Systems 25 (5) (1989) 761–765.
[18] J.N. Tsitsiklis, M. Athans, On the complexity of decentralized decision making and detection problems, IEEE Transactions on Automatic Control 30 (5) (1985) 440–446.
[19] H.L. Van Trees, Detection Estimation and Modulation Theory, vol. 1. Wiley, New York, 1969.
[20] Q. Zhang, P.K. Varshney, Decentralized M-ary detection via hierarchical binary decision fusion, Information Fusion 2 (2001) 3–16.
[21] Y. Zhu, R.S. Blum, Z. Luo, K.M. Wong, Unexpected properties and optimum-distributed sensor detectors for dependent observation cases, IEEE Transactions on Automatic Control 45 (1) (2000) 62–72.