Optimal Bi-Level Quantization of i.i.d. Sensor Observations for Binary Hypothesis Testing

Qian Zhang, *Member, IEEE*, Pramod K. Varshney, *Fellow, IEEE*, and Richard D. Wesel, *Senior Member, IEEE*

***Abstract*—We consider the problem of binary hypothesis testing using binary decisions from independent and identically distributed (i.i.d). sensors. Identical likelihood-ratio quantizers with threshold λ are used at the sensors to obtain sensor decisions. Under this condition, the optimal fusion rule is known to be a k-out-of-n rule with threshold k. For the Bayesian detection problem, we show that given k, the probability of error is a quasi-convex function of λ and has a single minimum that is achieved by the unique optimal $\lambda_{\rm opt}$. Except for the trivial situation where one hypothesis is always decided, we obtain a sufficient and necessary condition on $\lambda_{\rm opt}$, and show that $\lambda_{\rm opt}$ can be efficiently obtained via the SECANT algorithm. The overall optimal solution is obtained by optimizing every pair of (k, λ). For the Neyman–Pearson detection problem, we show that the use of the Lagrange multiplier method is justified for a given fixed k since the objective function is a quasi-convex function of λ. We further show that the receiver operating characteristic (ROC) for a fixed k is concave downward.**

***Index Terms*—Decision-making, multisensor systems, quantization, signal detection.**

I. Introduction

In this correspondence, we consider the problem of binary hypothesis testing using quantized sensor data. In our problem, n sensors observe an unknown hypothesis. The sensor observations are independent and identically distributed (i.i.d.) given the unknown hypothesis. Identical binary quantizers are employed at the sensors to yield n binary sensor decisions. These preliminary decisions are then transmitted to a fusion center. Based on received sensor decisions, the fusion center makes the final decision regarding the unknown hypothesis. Besides the distributed detection applications [1], this problem occurs in many other situations. In radar signal processing, a similar problem arises and is often referred to as the binary integration problem [2]–[4]. In communications, it is the 1-bit version of the generalized quantizer-detector design problem [5], [6].

A fundamental result is that when the sensor observations are conditionally independent, the optimal sensor quantizers are likelihood ratio threshold tests, normally using different thresholds [1]. However, the determination of these thresholds is generally difficult because multiple local optima may exist [1]. A person-by-person-optimal (PBPO) solution can be numerically obtained, for example, via the widely used Gauss-Seidel procedure, but the result may be only locally optimal.

In this correspondence, identical sensor quantizers are used, i.e., the same likelihood ratio threshold λ is used at all the sensors. Using identical sensor quantizers gives suboptimal result when the number of sensors is finite. However, in many cases the loss of optimality has been observed to be small. Moreover, the loss vanishes when the number of sensors goes to infinity [7], [8]. Using identical sensor quantizers significantly simplifies the problem. In this case, the optimum fusion rule is a k-out-of-n rule with threshold k [1].

We first consider the Bayesian problem where we seek the optimum (λ, k) that minimizes the average probability of error. We then consider the Neyman–Pearson problem in which we seek the optimum (λ, k) such that the probability of detection is maximized while the probability of false alarm is kept below a prescribed level.

A recent study by Shi, Sun, and Wesel [9] reveals an interesting property of this problem: quasi-convexity. They considered the problem of distributed detection of known signals in additive noise with identical sensors. Each sensor observation was a univariate random variable. Each sensor made a binary decision by comparing its observation to a threshold. For additive generalized Gaussian noise with any priors and for certain other additive noises with equal priors, they showed that the probability of error is a quasi-convex function of the sensor threshold, given a fixed k-out-of-n fusion rule.

The concept of quasi-convexity [10] can be quite useful since it eliminates the existence of multiple local optima. Hence, an optimum determined by any optimization method is the global optimum. Furthermore, it provides justification for the use of the Lagrange multiplier method for solving the Neyman–Pearson problem with identical sensors. This method is not applicable to the general distributed Neyman–Pearson problem due to a lack of convexity of the objective function. Quasi-convexity results obtained in this correspondence show that the Lagrange multiplier method can be used for a class of distributed Neyman–Pearson problems.

In this correspondence, we consider a much more general situation where sensor observations have an arbitrary probability distribution and sensor quantizers are defined in the likelihood ratio space rather than in the observation space as in [9], since the optimal sensor quantizer is a likelihood ratio threshold test [1]. For the Bayesian detection problem, we show that for arbitrary prior probabilities and for a fixed k-out-of-n fusion rule, the probability of error is a quasi-convex function of the sensor likelihood ratio threshold. We develop a method for computing the optimal threshold. For the Neyman–Pearson detection problem, we show that when the Lagrange multiplier method is used for a fixed k-out-of-n fusion rule, the objective function is a quasi-convex function of the sensor likelihood ratio threshold. The resulting receiver operating characteristic (ROC) is concave downward. The proof of these results is based on some fundamental properties of the sensor ROC. Based on the above results, one may first optimize the sensor likelihood ratio threshold, and then the k-out-of-n fusion rule.

In Section II, we present some notations and definitions. In Section III, we consider the Bayesian problem and prove the quasi-convexity property for the average probability of error. In Section IV, we consider the Neyman–Pearson problem and show that the Lagrange multiplier method is applicable. In Section V, we make some concluding remarks.

II. Notations and Definitions

Let us consider a fusion system that consists of n sensors and a fusion center. This system is used to determine whether an unknown hypothesis is H_0 or H_1. Let q_0 and q_1 denote the prior probabilities of H_0 and H_1.

Let $\boldsymbol{X}_i$ denote the observation of the ith sensor, $i = 1, \ldots, n$. When H_j is true, $\boldsymbol{X}_i$ follows the probability distribution function $p_j(\boldsymbol{x}_i)$, $j =$

Manuscript received May 18, 2000; revised October 1, 2001. This work was supported in part by the Space and Naval Warfare Systems Center San Diego, under Grant N66001-99-1-8922.

Q. Zhang is with Analog Devices Inc., Norwood, MA 02062 USA (e-mail: Qian.Zhang@analog.com).

P. K. Varshney is with the Electrical Engineering and Computer Science Department, Syracuse University, Syracuse, NY 13244 USA (e-mail: varshney@syr.edu).

R. D. Wesel is with the Electrical Engineering Department, University of California, Los Angeles, Los Angeles, CA 90095-1594 USA (e-mail: wesel@ucla.edu).

Communicated by U. Madhow, Associate Editor for Detection and Estimation.

Publisher Item Identifier S 0018-9448(02)05154-4.

$0, 1$. Let u_i denote the decision of the ith sensor, which is the output of likelihood ratio threshold test

$$\frac{p_1(\boldsymbol{x}_i)}{p_0(\boldsymbol{x}_i)} \underset{u_i=0}{\overset{u_i=1}{\gtrless}} \lambda \tag{1}$$

where λ is the common threshold since identical sensor quantizers are employed.

The sensors transmit their decisions to the fusion center. Based on the received sensor decisions, the fusion center makes the final decision u_0. Let $u_0 = 0$ if the fusion center decides H_0. Let $u_0 = 1$ if the fusion center decides H_1. We recall that an optimum fusion rule is a k-out-of-n rule [1]

$$u_0 = \begin{cases} 1, & \text{if } u_1 + \cdots + u_n \geq k \\ 0, & \text{if } u_1 + \cdots + u_n < k \end{cases} \tag{2}$$

where k is an integer and $1 \leq k \leq n$.

For all the sensors, let P_F denote the identical false alarm probability and P_D the identical detection probability, where

$$P_F = \text{Prob}(u_1 = 1 | H_0)$$

and

$$P_D = \text{Prob}(u_1 = 1 | H_1).$$

The (P_F, P_D) curve is generally referred to as the ROC.

The quality of the fusion center decision u_0 is measured by the system false alarm probability Q_F and the system detection probability Q_D

$$Q_F = \sum_{i=k}^{n} \binom{n}{i} P_F^i (1 - P_F)^{n-i} \tag{3a}$$

$$Q_D = \sum_{i=k}^{n} \binom{n}{i} P_D^i (1 - P_D)^{n-i}. \tag{3b}$$

A number of performance criteria can be formulated based on Q_F and Q_D. In the next section, we use the minimum probability of error criterion. We also show that the result holds for the general Bayesian problem. In Section IV, we use the Neyman–Pearson criterion.

III. The Bayesian Detection Problem

In this section, we first consider a special case of the Bayesian detection problem and then comment on the general Bayesian problem. Using Q_F and Q_D, we express the probability of error P_e as

$$P_e = q_0 Q_F + q_1 (1 - Q_D). \tag{4}$$

P_e is a function of k, λ, q_0, and q_1. Our goal is to find λ and k that minimize P_e. Toward this goal, we first find the optimum λ that minimizes P_e for each k, where $1 \leq k \leq n$. We then choose the smallest of these minima and the corresponding values of λ and k. This systematic procedure is exhaustive and is guaranteed to result in a globally optimal solution. Let P_e^k denote the probability of error for a fixed k. The process is stated as follows:

$$P_{e,\min}^k = \min_{\lambda} P_e^k, \qquad 1 \leq k \leq n$$

$$P_{e,\min} = \min_{k} P_{e,\min}^k.$$

Next, we consider the minimization of P_e^k for each k, where $1 \leq k \leq n$. We show that P_e^k is a *quasi-convex* function of λ. By a *quasi-convex* function $f(\lambda)$ of λ, we mean that for some λ^*, $f(\lambda)$ is nonincreasing for $\lambda \leq \lambda^*$ and $f(\lambda)$ is nondecreasing for $\lambda \geq \lambda^*$ [10]. Throughout this correspondence, we assume that P_F and P_D have first-order derivatives with respect to λ. This assumption is not restrictive in practical situations.

Lemma 1: P_e^k is a quasi-convex function of λ.

Proof: To prove the lemma, it suffices to show that either $\frac{dP_e^k}{d\lambda} \leq 0$ (or $\frac{dP_e^k}{d\lambda} \geq 0$) for all λ, or $\frac{dP_e^k}{d\lambda} \leq 0$ when $\lambda \leq \lambda^*$ and $\frac{dP_e^k}{d\lambda} \geq 0$ when $\lambda \geq \lambda^*$ for some λ^*. Some fundamental properties of ROC are used in the proof.

Taking the derivative of both sides of (4) with respect to λ and using (3a), (3b), and $\frac{dP_D}{dP_F} = \lambda$, we obtain

$$\begin{aligned} \frac{dP_e^k}{d\lambda} &= q_0 \frac{dQ_F}{d\lambda} - q_1 \frac{dQ_D}{d\lambda} \\ &= n\binom{n-1}{k-1} \cdot (-P_F') \cdot \lambda q_1 P_D^{k-1} (1 - P_D)^{n-k} \\ &\quad - n\binom{n-1}{k-1} \cdot (-P_F') \cdot q_0 P_F^{k-1} (1 - P_F)^{n-k}. \end{aligned}$$

Define

$$g(\lambda, k) = n\binom{n-1}{k-1} \cdot (-P_F') \cdot [q_0 P_F^{k-1} \cdot (1 - P_F)^{n-k}]$$

$$\begin{aligned} r(\lambda, k) &= \ln \frac{q_1}{q_0} + \ln \lambda + (k-1) \cdot \ln \frac{P_D}{P_F} \\ &\quad + (n-k) \cdot \ln \frac{1 - P_D}{1 - P_F}. \end{aligned} \tag{5}$$

Now we have

$$\frac{dP_e^k}{d\lambda} = g(\lambda, k) \cdot \left(e^{r(\lambda, k)} - 1\right). \tag{6}$$

Noting that P_F decreases as λ increases, i.e., $\frac{dP_F}{d\lambda} \leq 0$, we have $g(\lambda, k) \geq 0$. Therefore, the sign of $\frac{dP_e^k}{d\lambda}$ is determined by the value of $r(\lambda, k)$. Now it suffices to show that $r(\lambda, k)$ is either always negative (positive), or $r(\lambda, k) \leq 0$ for $\lambda \leq \lambda^*$ and $r(\lambda, k) \geq 0$ for $\lambda \geq \lambda^*$ for some λ^*.

Taking the derivative of both sides of (5) with respect to λ, we obtain

$$\begin{aligned} \frac{dr(\lambda, k)}{d\lambda} &= \frac{1}{\lambda} + \left(-\frac{dP_F}{d\lambda}\right) \\ &\quad \cdot \left[\frac{k-1}{P_D} \cdot \left(\frac{P_D}{P_F} - \lambda\right) + \frac{n-k}{1-P_D} \cdot \left(\lambda - \frac{1-P_D}{1-P_F}\right)\right]. \end{aligned} \tag{7}$$

Since (P_F, P_D) is a point on a ROC curve, which is concave downward as shown in Fig. 1, we have

$$\frac{P_D}{P_F} \geq \lambda \geq \frac{1 - P_D}{1 - P_F}.$$

Combining this result with the fact that $1 \leq k \leq n$, we know that the term in the square bracket in the right-hand side of (7) is nonnegative. Recalling that $\frac{dP_F}{d\lambda} \leq 0$, we have $\frac{dr(\lambda, k)}{d\lambda} > 0$, i.e., $r(\lambda, k)$ is a monotone increasing function of λ. Hence, $r(\lambda, k)$ intersects the λ-axis at most once. This proves the lemma. □

In many situations, use of the log-likelihood ratio is preferred. Let $\tau = \ln \lambda$. Because τ is a monotone increasing function of λ, Lemma 1 holds when λ is replaced by τ.

Corollary 1: Given k, P_e^k is a quasi-convex function of τ.

The proof of Lemma 1 implies that P_e^k has a single minimum that is achieved by a unique λ. This result is stated in the following theorem.

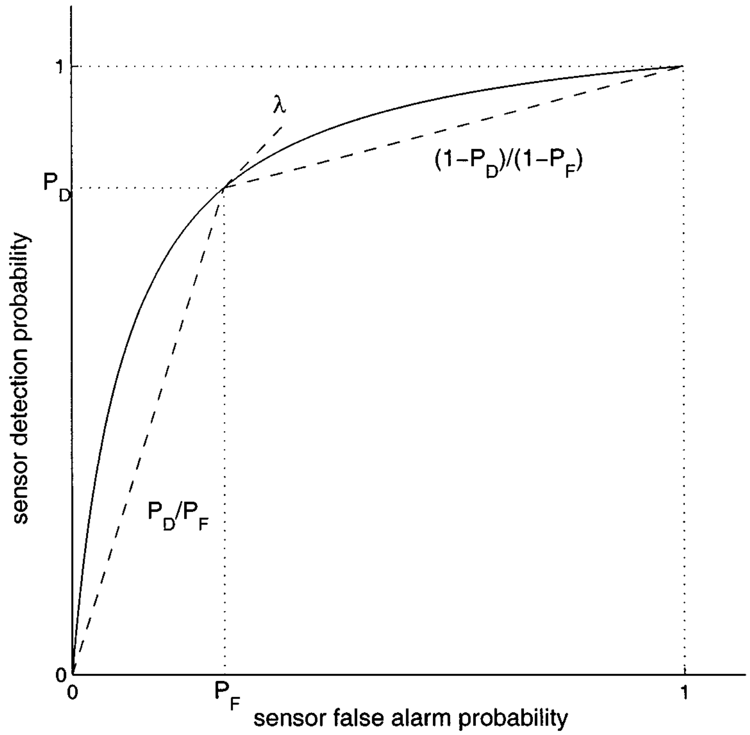

Fig. 1. Sensor ROC.

Theorem 1: P_e^k has a single minimum, which is achieved by a unique λ.

The proof of Lemma 1 suggests that if λ satisfies $r(\lambda, k) = 0$, then λ minimizes P_e^k. This is a sufficient condition stated in the following theorem.

Theorem 2: For a given k, λ minimizes P_e^k if it satisfies

$$\ln \frac{q_1}{q_0} + \ln \lambda + (k-1) \ln \frac{P_D}{P_F} + (n-k) \ln \frac{1-P_D}{1-P_F} = 0. \tag{8}$$

It is shown later that for nontrivial solutions, i.e., when the fusion system does not always decide one hypothesis, $r(\lambda, k) = 0$ is also a necessary condition on the optimal λ.

To use Theorem 2 to find the optimal λ, $r(\lambda, k) = 0$ must have a positive root. This condition is satisfied for a class of sensors. On the sensor ROC, let $\lambda_{0,0}$ denote the slope at the point $(0, 0)$, and $\lambda_{1,1}$ the slope at the point $(1, 1)$. The following theorem shows that if $\lambda_{0,0} = \infty$ and $\lambda_{1,1} = 0$, then $r(\lambda, k) = 0$ has a positive root for any nonzero prior probabilities. This result can be obtained using simple calculus and the properties of ROC that

$$\lim_{\lambda \to 0} \frac{1}{\lambda} \cdot \frac{1-P_D}{1-P_F} = 1 \quad \text{and} \quad \lim_{\lambda \to +\infty} \frac{1}{\lambda} \cdot \frac{P_D}{P_F} = 1.$$

Theorem 3: For a given k, $1 \le k \le n$, if $\lambda_{0,0} = \infty$, $\lambda_{1,1} = 0$, and $q_0, q_1 > 0$, then $r(\lambda, k) = 0$ has a unique positive root.

Putting

$$\lim_{\lambda \to 0} \frac{P_D}{P_F} = 1$$
$$\lim_{\lambda \to 0} \frac{1}{\lambda} \cdot \frac{1-P_D}{1-P_F} = 1$$
$$\lim_{\lambda \to +\infty} \frac{1-P_D}{1-P_F} = 1$$

and

$$\lim_{\lambda \to +\infty} \frac{1}{\lambda} \cdot \frac{P_D}{P_F} = 1$$

into (5), and using $\tau = \ln \lambda$, we have

$$\lim_{\tau \to -\infty} \frac{r(\tau, k)}{(n-k+1) \cdot \tau} = 1 \quad \text{and} \quad \lim_{\tau \to +\infty} \frac{r(\tau, k)}{k \cdot \tau} = 1.$$

Therefore, $r(\tau, k)$ is approximately a linear function of τ at $\pm\infty$. Using this property, we can approximate the $r(\tau, k)$ curve by two straight lines and use the SECANT algorithm [11] to find the root of $r(\tau, k) = 0$. This algorithm is stated as follows.

1. Choose $\varepsilon > 0$. Arbitrarily choose τ_1, τ_2. Let $r_1 = r(\tau_1, k)$ and $r_2 = r(\tau_2, k)$. Set $i = 3$.
2. Let
$$\tau_i = \frac{r_{i-1} \cdot \tau_{i-2} - r_{i-2} \cdot \tau_{i-1}}{r_{i-1} - r_{i-2}}.$$
Let $r_i = r(\tau_i, k)$.
3. If $|r_i| \le \varepsilon$, stop; otherwise, let $i = i + 1$, go to step 2.

When $\lambda_{0,0} = \infty$ and $\lambda_{1,1} = 0$, the algorithm converges quickly because $r(\tau, k)$ is well approximated by two straight lines, as shown in Fig. 2.

Since $r(\tau, k)$ is a monotone increasing function of λ, $r(\tau, k) = 0$ has a root if and only if $r(\lambda_{0,0}, k) \ge 0$ and $r(\lambda_{1,1}, k) \le 0$. When these conditions are satisfied, the uniqueness of the optimal λ (Theorem 1) implies that $r(\lambda, k) = 0$ is a necessary condition on the optimal λ. On the other hand, if $r(\lambda_{0,0}, k) < 0$, H_0 is always decided; if $r(\lambda_{1,1}, k) > 0$, H_1 is always decided. These are trivial solutions.

From Theorems 2 and 3, we observe that the optimal value of λ is intimately related to k, q_0, and q_1. Here we present some results on these relationships. Suppose $r(\lambda, k) = 0$ has a positive root λ_k for each k, where $1 \le k \le n$.

Lemma 2: λ_k is a decreasing function of k for fixed q_1 and q_0, except when $P_D = P_F$, or when the sensors always decide H_1.

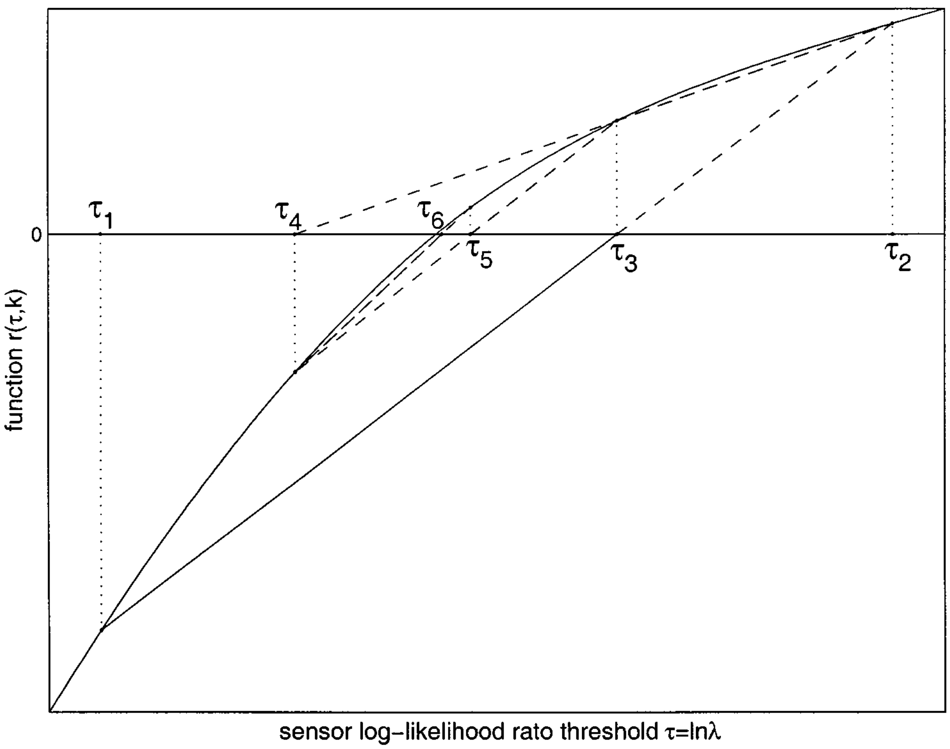

Fig. 2. The SECANT algorithm.

Proof: It suffices to show $\frac{d\lambda_k}{dk} \le 0$. Using (8), we have

$$\frac{d\lambda_k}{dk} = -\left[\ln\frac{P_D}{P_F} - \ln\frac{1-P_D}{1-P_F}\right] \Big/ \left[\frac{1}{\lambda_k} + \frac{k-1}{P_D}\cdot\left(-\frac{dP_F}{d\lambda_k}\right)\cdot\left(\frac{P_D}{P_F}-\lambda_k\right) + \frac{n-k}{1-P_D}\cdot\left(-\frac{dP_F}{d\lambda_k}\right)\cdot\left(\lambda_k - \frac{1-P_D}{1-P_F}\right)\right].$$

Since

$$\frac{P_D}{P_F} \ge \lambda_k \ge \frac{1-P_D}{1-P_F}$$

$\frac{dP_F}{d\lambda_k} \le 0$, and $1 \le k \le n$, we have $\frac{d\lambda_k}{dk} \le 0$. The equality holds only if

$$\frac{P_D}{P_F} = \lambda_k = \frac{1-P_D}{1-P_F}$$

or $\lambda_k = 0$. To satisfy the first condition, the sensors must have $P_D = P_F$, i.e., they cannot distinguish H_0 and H_1. To satisfy the second condition, the sensors must always decide H_1. These are trivial situations and will not be considered in this correspondence. □

Lemma 3: λ_k is a decreasing function of q_1/q_0 for fixed k, except when the sensors always decide H_1.

Proof: Let $s = q_1/q_0$. It suffices to show that $\frac{d\lambda_k}{ds} \le 0$. Using (8), we have

$$\frac{d\lambda_k}{ds} = -\frac{1}{s} \Big/ \left[\frac{1}{\lambda_k} + \frac{k-1}{P_D}\cdot\left(-\frac{dP_F}{d\lambda_k}\right)\cdot\left(\frac{P_D}{P_F}-\lambda_k\right) + \frac{n-k}{1-P_D}\cdot\left(-\frac{dP_F}{d\lambda_k}\right)\cdot\left(\lambda_k - \frac{1-P_D}{1-P_F}\right)\right].$$

Since

$$\frac{P_D}{P_F} \ge \lambda_k \ge \frac{1-P_D}{1-P_F}$$

$\frac{dP_F}{d\lambda_k} \le 0$, and $1 \le k \le n$, we have $\frac{d\lambda_k}{ds} \le 0$. Equality holds if and only if $s = +\infty$, i.e., $q_1 = 1$, or $\lambda_k = 0$. To satisfy these conditions, the sensors must always decide H_1. This is a trivial situation and will not be considered in this correspondence. □

Next, we illustrate the theoretical results in an example. We consider the detection of known signals in Gaussian noise. The sensor observation is $x_i = s_i + z_i$, where $s_i = \pm d$ is the transmitted signal and z_i is a Gaussian random variable with zero mean and unit variance.

Define $H_0 \equiv \{s_i = -d\}$ and $H_0 \equiv \{s_i = +d\}$. The log-likelihood ratio τ_i for this problem is given by $\tau_i = 2dx_i$. The sensor false alarm and detection probabilities are

$$P_F = Q\left(\frac{\tau}{2d} + d\right) \quad \text{and} \quad P_D = Q\left(\frac{\tau}{2d} - d\right)$$

where τ is the log-likelihood ratio threshold and

$$Q(z) = \int_z^\infty \frac{1}{\sqrt{2\pi}}\, e^{-\frac{x^2}{2}}\, dx.$$

With fixed $n = 5$ and $d = 1$, we observe the relationship between P_e^k, τ, k, q_0, and q_1.

In Fig. 3, with $q_0 = q_1 = 0.5$, P_e^k is plotted against τ for each value of k. We can see that for any given k, P_e^k is a quasi-convex function of τ and has a single minimum achieved by a unique value of τ. These results agree with the main results of Section III. We also notice that the optimal τ decreases with k, as suggested in Lemma 2.

In Fig. 4, with $q_0 = 0.25$ and $q_1 = 0.75$, P_e^k is plotted against τ for each value of k. Comparing this figure to Fig. 3, we notice that the optimal τ for each value of k has decreased. These decreases are due

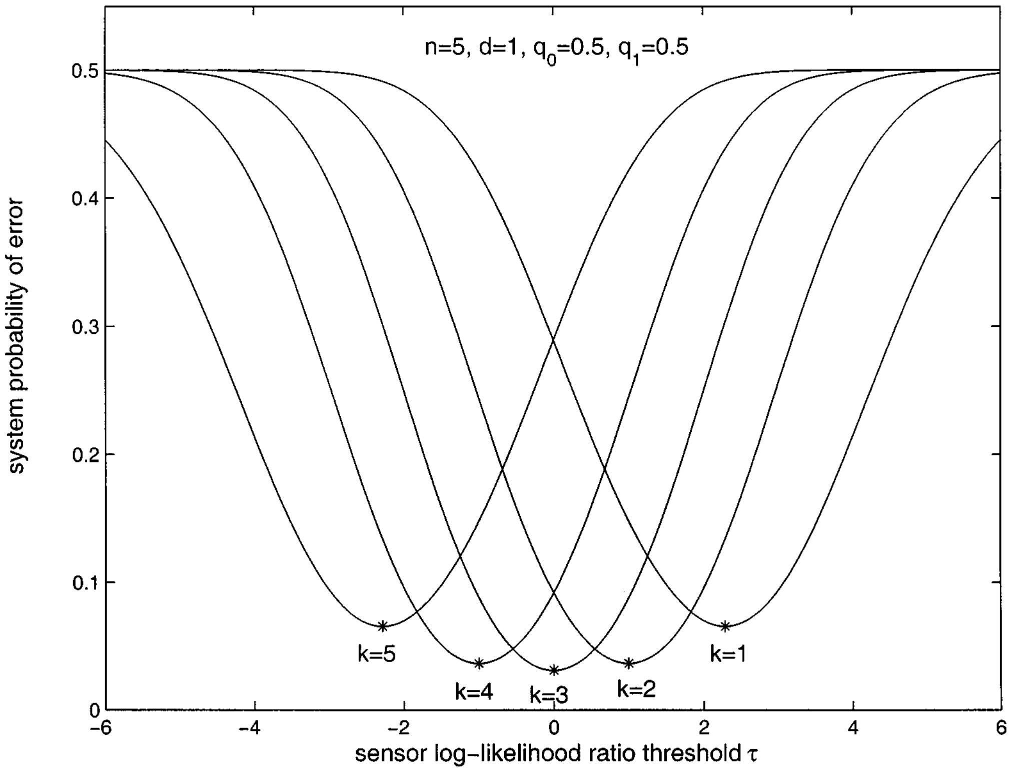

Fig. 3. Probability of error with equal prior probabilities.

to the increase in q_1 and of course the decrease in q_0, as implied by Lemma 3.

In Fig. 5, with $q_0 = q_1 = 0.5$, $r(\tau, k)$ is plotted against τ for each value of k. We can see that $r(\tau, k)$ is a monotonically increasing function of τ for any given k. We also notice that a $r(\tau, k)$ curve can be well approximated by two concatenated straight lines.

The above results for P_e^k can be easily extended to the more general case by taking into account the Bayesian costs. The result is available in [14].

IV. The Neyman–Pearson Detection Problem

In this section, we consider the Neyman–Pearson problem. We show that this problem can be formulated as a minimum probability of error problem that we considered in Section III.

In a Neyman–Pearson detection problem, the goal is to maximize Q_D while keeping Q_F below a prescribed level α. Because the sensor decisions are discrete random variables, fusion center may need to use randomized fusion rules to meet the required probability of false alarm [12], [13]. However, randomized fusion rules introduce an undesired degree of freedom and require additional computational resources for synchronization. In this correspondece, we only consider fixed fusion rules, i.e., k-out-of-n fusion rules.

We break the original Neyman–Pearson problem into a set of Neyman–Pearson problems for each value of k. For a given k, our goal is to maximize Q_D while keeping Q_F below a prescribed value α. We then choose the solution that yields the maximal Q_D.

We are particularly interested in solving this problem via the Lagrange multiplier method. This method has seen limited use in solving the distributed Neyman–Pearson problem for the general case with nonidentical sensors [1]. The difficulty is inherent in that the objective function may not be convex, even with a fixed fusion rule, and multiple local optima may exist. Fortunately for our problem, we show that for a fixed k-out-of-n fusion rule, the objective function is a quasi-convex function and therefore the Lagrange multiplier method is suitable. Furthermore, we show that for a fixed k-out-of-n fusion rule, the ROC is concave downward.

Define $L_k(\lambda, s) = s(Q_F - \alpha) - Q_D$, where s is the Lagrange multiplier and $s \geq 0$. Now our goal is to minimize $L_k(\lambda, s)$ with respect to s and λ.

Lemma 4: $L_k(\lambda, s) =$ is a quasi-convex function of λ.

Proof: Through simple arithmetic manipulation, we have

$$L_k(\lambda, s) = \frac{1}{q_1} P_e^k - \left(\frac{q_0}{q_1}\alpha + 1\right)$$

where $q_0 = s/(1+s)$ and $q_1 = 1/(1+s)$. Since P_e^k is a quasi-convex function of λ, so is $L_k(\lambda, s)$. □

Using this result, one may set $\frac{\partial L_k(\lambda, s)}{\partial \lambda}$ equal to zero to obtain the solution. Noticing $s = q_0/q_1$, one will obtain the solution by solving the following equations for each value of k:

$$Q_F - \alpha = 0$$
$$\ln\frac{1}{s} + \ln\lambda + (k-1)\cdot\ln\frac{P_D}{P_F} + (n-k)\cdot\ln\frac{1-P_D}{1-P_F} = 0.$$

Since $L_k(\lambda, s)$ is a quasi-convex function of λ, it has a single minimum and any optimization technique can be used to determine it. To find the globally optimal solution, one can repeat the above proceduref or all possible values of k, and then choose the pair of k and λ that gives the largest Q_D.

In general, the resulting ROC, i.e., the (Q_F, Q_D) curve, is not concave downward as in centralized detection. However for a fixed k-out-of-n fusion rule, the ROC is concave.

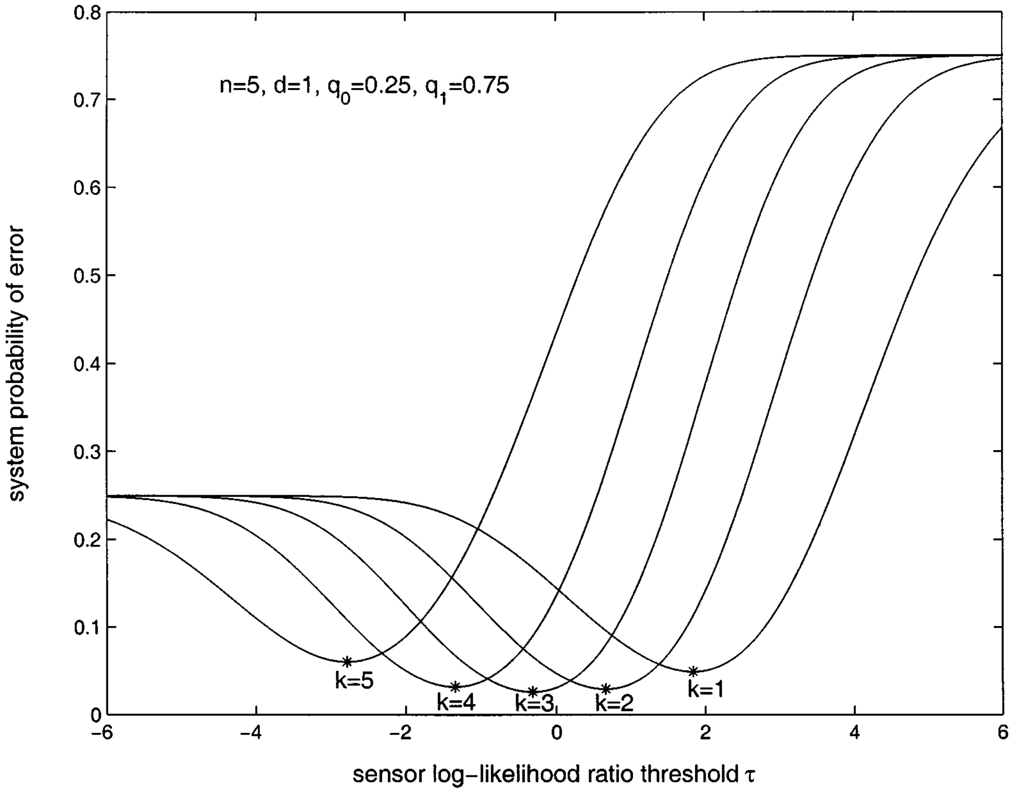

Fig. 4. Probability of error with unequal prior probabilities.

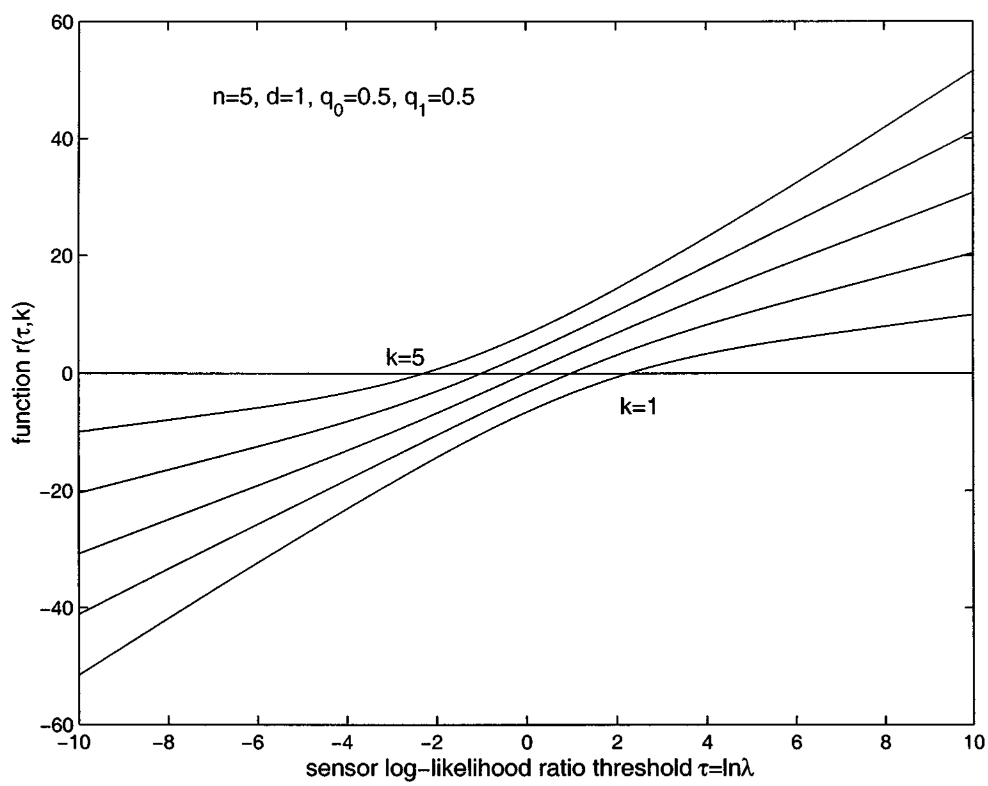

Fig. 5. $r(\tau, k)$ curves and the two-line approximation.

Lemma 5: For a given k, Q_D is a concave function of Q_F.

Proof: It suffices to show that $\frac{d^2 Q_D}{dQ_F^2} \leq 0$. Recalling from the proof of Lemma 1, we have

$$\frac{dQ_D}{dQ_F} = \frac{dQ_D}{d\lambda} \Bigg/ \frac{dQ_F}{d\lambda} = e^{r(\lambda,\, k)}.$$

Therefore,

$$\frac{d^2 Q_D}{dQ_F^2} = \frac{de^{r(\lambda,\, k)}}{dQ_F} = \frac{de^{r(\lambda,\, k)}}{d\lambda} \Bigg/ \frac{dQ_F}{d\lambda} = e^{r(\lambda,\, k)} \cdot \frac{dr(\lambda,\, k)}{d\lambda} \Bigg/ \frac{dQ_F}{d\lambda}.$$

Since $\frac{dP_F}{d\lambda} \leq 0$, we have $\frac{dQ_F}{d\lambda} \leq 0$. Recalling that $\frac{dr(\lambda,\, k)}{d\lambda} \geq 0$, we have $\frac{d^2 Q_D}{dQ_F^2} \leq 0$. □

The concavity of the ROC for a fixed k-out-of-n fusion rule ensures that the Lagrange multiplier method can be used to uniquely determine the optimal threshold in the case considered here.

V. Summary

We considered the problem of distributed binary hypothesis testing with independent identical sensors. The goal was to find the optimal k-out-of-n fusion rule and the optimal likelihood ratio threshold test for the sensors according to a performance criterion.

For the Bayesian detection problem, we showed that the objective function possesses the property of quasi-convexity, which secures the unique (global) optimum. We then developed a SECANT type of algorithm to efficiently compute the optimum. For Neyman–Pearson detection problem, we showed that the quasi-convexity exists and gives good reason for the use of the Lagrange multiplier method.

References

[1] P. K. Varshney, *Distributed Detection and Data Fusion*. New York: Springer-Verlag, 1997.

[2] J. Han, P. K. Varshney, and R. Srinivasan, "Distributed binary integration," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 29, pp. 2–8, Jan. 1993.

[3] F. Gini, F. Lombardini, and L. Verrazzani, "Decentralized CFAR detection with binary integration in Weibull clutter," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 33, pp. 396–407, Apr. 1997.

[4] J. V. DiFranco and W. L. Rubin, *Radar Detection*. Englewood Cliffs, NJ: Prentice-Hall, 1968.

[5] S. A. Kassam, "Optimum quantization for signal detection," *IEEE Trans. Commun.*, vol. COM-25, pp. 479–484, May 1977.

[6] H. V. Poor and J.B. Thomas, "Applications of Ali–Silvey distance measures in the design of generalized quantizers for binary decision systems," *IEEE Trans. Commun.*, vol. COM-25, pp. 893–900, Sept. 1977.

[7] J. N. Tsitsiklis, "Decentralized detection by a large number of sensors," *Math. Contr., Signals, Syst.*, vol. 1, no. 2, pp. 167–182, 1988.

[8] P. Chen and A. Papamarcou, "New asymptotic results in parallel distributed detection," *IEEE Trans. Inform. Theory*, vol. 39, pp. 1847–1863, Nov. 1993.

[9] W. Shi, T. W. Sun, and R. D. Wesel, "Quasiconvexity and optimal binary fusion for distributed detection with identical sensors in generalized Gaussian noise," *IEEE Trans. Inform. Theory*, vol. 47, pp. 446–450, Jan. 2001.

[10] L. Hörmander, *Notions of Convexity*. Boston, MA: Birkhäuser, 1994.

[11] W. H. Press *et al.*, *Numerical Recipes: The Art of Scientific Computing*. New York: Cambridge Univ. Press, 1986.

[12] J. N. Tsitsiklis, "Decentralized detection," *Adv. Statist. Signal Processing*, vol. 2, pp. 297–344, 1993.

[13] Y. I. Han and T. Kim, "Randomized fusion rules can be optimal in distributed Neyman–Pearson detectors," *IEEE Trans. Inform. Theory*, vol. 43, pp. 1281–1288, July 1997.

[14] Q. Zhang, "Efficient computational algorithms for the design of distributed detection networks," Ph.D. dissertation, Syracuse Univ., Syracuse, NY, 2000.

A Note on Robust Hypothesis Testing

Luc Devroye, László Györfi, *Fellow, IEEE*, and Gábor Lugosi, *Member, IEEE*

***Abstract*—We introduce a simple new hypothesis testing procedure, which, based on an independent sample drawn from a certain density, detects which of k nominal densities is the true density closest to, under the total variation (L_1) distance. We obtain a density-free uniform exponential bound for the probability of false detection.**

***Index Terms*—Robust detection, robust hypotheses testing.**

I. Result

A model of robust hypothesis testing may be formulated as follows: let $f^{(1)}, \ldots, f^{(k)}$ be fixed densities on $\mathbb{R}^d$ which are the nominal densities under k hypotheses. We observe independent and identically distributed (i.i.d.) random vectors $X_1, \ldots, X_n$ according to a common density f. Under the hypothesis H_j $(j = 1, \ldots, k)$ the density f is a distorted version of $f^{(j)}$. This notion may be formalized in various ways. In this correspondence, we assume that the true density lies within a certain total variation distance of the underlying nominal density. More precisely, we assume that there exists a positive number ϵ such that for some $j \in \{1, \ldots, k\}$

$$\left\| f - f^{(j)} \right\| \leq \Delta_j - \epsilon$$

where

$$\Delta_j \stackrel{\text{def}}{=} (1/2) \min_{i \neq j} \|f^{(i)} - f^{(j)}\|.$$

Here $\|f - g\| = \int |f - g|$ denotes the L_1 distance between two densities. Thus, we formally define the k hypotheses by (see Fig. 1)

$$H_j = \left\{ f\colon \left\| f - f^{(j)} \right\| \leq \Delta_j - \epsilon \right\}, \qquad j = 1, \ldots, k.$$

The goal is to construct tests which, with high probability, assign to the observed sample the index j of the correct nominal density, that is, determines to which hypothesis H_j the density f belongs to.

Manuscript received August 1, 2000; revised July 1, 2001. The work of G. Lugosi was supported by DGI under Grant BMF2000-0807.

L. Devroye is with the School of Computer Sciences, McGill University, Montreal, QC H3A 2K6, Canada (e-mail: luc@kriek.cs.mcgill.ca).

L. Györfi is with the Department of Computer Science and Information Theory, Technical University of Budapest, Budapest, Hungary (e-mail: gyorfi@inf.bme.hu).

G. Lugosi is with the Department of Economics, Pompeu Fabra University, 08005 Barcelona, Spain (e-mail: lugosi@upf.es).

Communicated by U. Madhow, Associate Editor for Detection and Estimation.

Publisher Item Identifier S 0018-9448(02)05147-7.