Robust Detection of Signals in Dependent Noise

GEORGE V. MOUSTAKIDES AND JOHN B. THOMAS, FELLOW, IEEE

Abstract—**The robust detection of signals in additive dependent noise is considered. The solution to the finite-sample problem is obtained when the Bayes risk is used as the performance measure. For the multivariate densities involved we assume that they belong to an ϵ-contamination model. The robust detection structure is shown to be optimum for the least-favorable density and is a censored version of the nominal likelihood ratio.**

I. Introduction

ROBUST DETECTION of signals in additive noise is considered in [1]. Under independent identically distributed (i.i.d.) noise, the min-max detector for the finite-sample strong signal case is found. Optimality is defined in a Neyman–Pearson sense. Specifically, it is required that the detector have a false alarm probability smaller than a certain level for every density inside an ϵ-contamination class, while the probability of detection satisfies a saddle-point condition. The solution is obtained using the results in [2]. This approach leads to a detector structure that is not optimum for an additive model for any noise density from the allowable class. In addition, even for reasonable values of the signals it can lead to a trivial detector structure. Other approaches consider asymptotically robust detection of weak signals in i.i.d. noise [1], [3], [4] or dependent noise [5]–[7]. Finally in [8], using the large deviations theory, the asymptotically robust detection of strong signals in i.i.d. noise is considered.

In this paper we find the robust detector for the finite-sample dependent observations case. We require that the multivariate densities involved satisfy an ϵ-contamination model. As a performance measure we use the Bayes risk. The detector structure is required to satisfy a saddle-point condition over the allowable class. Under our assumptions the optimum detector is always nontrivial and is optimum for a least-favorable density.

II. Preliminaries

We are given an observation vector $\{X_i\}_{i=1}^n$, and we would like to decide between the following two hypotheses:

$$H_0\colon\ X_i = N_i$$
$$H_1\colon\ X_i = N_i + s_i, \qquad i = 1,2,\cdots,n \tag{1}$$

where $\{N_i\}_{i=1}^n$ is a noise vector and $\{s_i\}_{i=1}^n$ is a known nonzero signal vector. As our performance measure we will use the Bayes risk. It is easy to see that for any reasonable cost assignment, the Bayes risk is equivalent to

$$R = P(D_0/H_1) + tP(D_1/H_0) \tag{2}$$

where $P(D_i/H_j)$ denotes the probability of deciding in favor of H_i given that H_j is true. The quantity t depends only on the prior probabilities and the decision costs; thus we will consider it to be a known constant. If we make the assumption that the risk of an erroneous decision is larger than the risk of a correct decision, then minimizing the total risk is equivalent to minimizing the expression in (2).

For simplicity let us denote random variables with upper case letters, vectors of random variables with upper case boldface letters, real variables with lower case letters, and vectors of real variables with lower case boldface letters. If V denotes the subset of R^n in which we decide in favor of H_1 and if we assume that the vector $\mathbf{N} = \{N_i\}_{i=1}^n$ has a multivariate density $f(\mathbf{x})$, then the performance measure defined in (2) can be written as

$$R(V,f) = \int_{V^c} f(\mathbf{x}-\mathbf{s})\,d\mathbf{x} + \int_V f(\mathbf{x})\,d\mathbf{x} \tag{3}$$

where V^c denotes the complement set of V. For a given density $f(\mathbf{x})$ the optimum set V that minimizes (3) is given by

$$V_0 = \left\{\mathbf{x}\colon \frac{f(\mathbf{x}-\mathbf{s})}{f(\mathbf{x})} \geq t\right\}. \tag{4}$$

The optimum set in (4) can be defined if we know the noise density $f(\mathbf{x})$. In the next section we will define an optimum set for the case where the density $f(\mathbf{x})$ is not exactly known, but it belongs to an ϵ-contamination class.

III. Robust Detection

As we can see from (4) the optimum detector structure that minimizes $R(V, f)$ can be easily defined if we know exactly the noise density $f(\mathbf{x})$. We now assume that this density is not known exactly and that it satisfies the following ϵ-contamination model:

$$f(\mathbf{x}) = (1-\epsilon)g(\mathbf{x}) + \epsilon h(\mathbf{x}). \tag{5}$$

Manuscript received May 22, 1985; revised February 27, 1986. This work was supported in part by the Office of Naval Research under Grant N00014-81-K-0146.

G. V. Moustakides is with the Institut de Recherche en Informatique et Systemes Aleatoires, Avenue du General Leclerc, 35042 Rennes, France.

J. B. Thomas is with the Department of Electrical Engineering and Computer Science, E-QUAD, Princeton University, Princeton, NJ 08544, USA.

IEEE Log Number 861092.

The parameter ϵ is assumed to be a known constant in the interval $[0,1)$. The functions $g(x)$ and $h(x)$ are multivariate densities. The density $g(x)$ is assumed to be known and $h(x)$ to be unknown. Following a saddle-point approach for our performance measure in (3), we would like to find a pair V_r, $f_l(x)$ that satisfies the following double inequality:

$$R(V,f_l) \geq R(V_r,f_l) \geq R(V_r,f) \tag{6}$$

for any $f(x)$ satisfying (5) and any set V. Notice that with the approach followed here the two error probabilities $P(D_0/H_1)$ and $P(D_1/H_0)$ cannot be considered independently as in the approach in [1] because the densities are not allowed to vary independently under the two hypotheses here, i.e., the additive-channel constraint is imposed. Thus the result of Huber [2, problem (iii)] cannot be applied here. With (6) we will necessarily have a single least-favorable density.

Before going to the solution let us first present our assumptions for the density $g(x)$. We assume that it is a well-defined nowhere-vanishing density, strongly unimodal in the direction of the signal vector $s = \{s_i\}_{i=1}^n$. Specifically, we assume that the function $-\ln(g(x+qs))$ is strictly convex with respect to the real q for every vector x. Thus if $L_g(x)$ denotes the nominal likelihood ratio, the function $L_g(x+qs)$ is continuous and strictly increasing in q for every fixed x. For $L_g(x)$ we also assume that constants $\alpha_V, \alpha_L \in [0,\infty]$ exist such that for every fixed vector x we have

$$\lim_{q\to+\infty} L_g(x+qs) = \alpha_V$$
$$\lim_{q\to-\infty} L_g(x+qs) = \alpha_L. \tag{7}$$

If $\alpha \in (\alpha_L, \alpha_V)$, we define as C_α the set

$$C_\alpha = \{x\colon L_g(x) = \alpha\}. \tag{8}$$

It is easy to see that the set C_α is nonempty because, from (7) and using the continuity of $L_g(x+qs)$ with respect to q, for any fixed x we can find a q to have $L_g(x+qs) = \alpha$. In general, the sets C_α will be $(n-1)$-dimensional surfaces. Notice now that, since $L_g(x+qs)$ is strictly increasing in q, we have that $\alpha_V > L_g(x) > \alpha_L$ for every x. Because $L_g(x)$ is a likelihood ratio vectors z and w exist that satisfy $L_g(z) \geq 1 \geq L_g(w)$; thus we conclude

$$\alpha_V > 1 > \alpha_L. \tag{9}$$

Finally, for $k = 0, \pm1, \pm2, \cdots$ let us define the following sets:

$$D_\alpha^k = \{x\colon x = x_c + qs\}, \qquad x_c \in C_\alpha, [q] = k-1 \tag{10}$$

where with [ ] we denote the integer part.

As noted earlier, the set C_α is generally an $(n-1)$-dimensional surface. The set D_α^k is the collection of points between two surfaces that are images of C_α after translation by the vectors $(k-1)s$ and ks. Since, because of (7) and the continuity of $L_g(x)$ along the direction of the vector s, we can always write $x = x_c + qs$ with $x_c \in C_\alpha$, we see that every x belongs to some set D_α^k. Thus using also the monotonicity of $L_g(x)$ along the direction of s, we have

$$A_\alpha = \{x\colon L_g(x) \geq \alpha\} = \bigcup_{k=1}^{\infty} D_\alpha^k$$
$$B_\alpha = \{x\colon L_g(x) < \alpha\} = \bigcup_{k=0}^{\infty} D_\alpha^{-k}. \tag{11}$$

We are now ready to define the optimum detector structure. The left inequality in (6) suggests that V_r must be the set

$$V_r = \left\{x\colon \frac{f_l(x-s)}{f_l(x)} \geq t\right\}. \tag{12}$$

Thus we need to specify only the least favorable density $f_l(x)$. It turns out that this density is not the same for every t. We must distinguish the cases $0 \leq t \leq 1$ and $t > 1$. With the following theorem we define the least-favorable density.

Theorem: A density $f_l(x)$ that satisfies the saddle-point condition defined in (6) is given by the following.

Case 1: For $0 \leq t \leq 1$,

$$f_l(x) = \begin{cases} (1-\epsilon)g(x), & x \in A_\xi \\ (1-\epsilon)g(x+ks)\xi^k, & x \in D_\xi^{-k}, k = 0,1,2,\cdots \end{cases} \tag{13}$$

where ξ is a constant less than unity and selected to satisfy the total mass constraint. Notice that for $x \in D_\xi^{-k}$ we have $x + ks \in D_\xi^0$; thus ξ satisfies

$$\int_{A_\xi} g(x)\,dx + \frac{1}{1-\xi}\int_{D_\xi^0} g(x)\,dx = \frac{1}{1-\epsilon}. \tag{14}$$

Case 2: For $t > 1$,

$$f_l(x) = \begin{cases} (1-\epsilon)g(x), & x \in B_\zeta \\ (1-\epsilon)g(x-ks)\zeta^{-k}, & x \in D_\zeta^k, k = 1,2,\cdots \end{cases} \tag{15}$$

where ζ is a constant greater than unity. To satisfy the mass constraint, we have

$$\int_{B_\zeta} g(x)\,dx + \frac{1}{\zeta-1}\int_{D_\zeta^0} g(x)\,dx = \frac{1}{1-\epsilon}. \tag{16}$$

Proof: The proof of this theorem with proofs of existence and uniqueness of solutions to (14) and (16) is given in the Appendix.

Let us consider the form of the likelihood ratio for the density $f_l(x)$ since, because of (12), this defines the optimum set V_r. Denoting the likelihood ratio of the least favorable density by $L_r(x)$, its form is given by the following lemma.

Lemma: The likelihood ratio of $f_l(x)$ defined in (13) and (15) is given by the following.

Case 1: For $0 \le t \le 1$,

$$L_r(x) = \begin{cases} L_g(x), & x \in A_\xi \text{ (or when } L_g(x) \ge \xi) \\ \xi, & x \in \bigcup_{k=0}^{\infty} D_\xi^{-k} \text{ (or when } L_g(x) < \xi). \end{cases} \tag{17}$$

Case 2: For $t > 1$,

$$L_r(x) = \begin{cases} \zeta, & x \in \bigcup_{k=1}^{\infty} D_\zeta^{k} \text{ (or when } L_g(x) \ge \zeta) \\ L_g(x), & x \in B_\zeta \text{ (or when } L_g(x) < \zeta). \end{cases} \tag{18}$$

Proof: We will only show (17), since (18) follows similarly. From (10) and (11) we see that for $x \in A_\xi$ we have $x - s \in A_\xi \cup D_\xi^0$. Thus from (13) for $x \in A_\xi$ we have that $f_l(x) = (1-\epsilon)g(x)$ and $f_l(x - s) = (1-\epsilon)g(x - s)$; thus (17) is satisfied. When $x \in D_\xi^{-k}$ for some k from the set $\{0, 1, 2, \cdots\}$, we have that $x - s \in D_\xi^{-(k+1)}$, and using (13) yields

$$L_r(x) = \frac{(1-\epsilon)g(x - s + (k+1)s)\xi^{k+1}}{(1-\epsilon)g(x + ks)\xi^k} = \xi. \tag{19}$$

This concludes the proof.

Notice that $L_r(x)$ defined in (17) and (18) can be written as $L_r(x) = b_i(L_g(x))$, where the univariate functions $b_i(z)$, $i = 1, 2$ are defined as follows:

$$b_1(z) = \begin{cases} z, & z \ge \xi \\ \xi, & z < \xi \end{cases} \tag{20}$$

for the case $0 \le t \le 1$ and

$$b_2(z) = \begin{cases} \zeta, & z \ge \zeta \\ z, & z < \zeta \end{cases} \tag{21}$$

for the case $t > 1$. It turns out that we can combine the two cases. Since the important thing is the set V_r, we can see that this set does not change if instead of the two functions $b_i(z)$ we use the following function:

$$b(z) = \begin{cases} \zeta, & z \ge \zeta \\ z, & \xi \le z < \zeta. \\ \xi, & z < \xi \end{cases} \tag{22}$$

The foregoing statement is true because from the theorem we have that $\xi < 1 < \zeta$. We see that the robust detector uses a censored version of the nominal likelihood ratio. Basically, what this result means is that if our prior knowledge (expressed by t) exceeds certain levels (ξ or ζ), then because of the contamination in the nominal density the prior knowledge is more important than the posterior (expressed by $L_g(x)$) and should be the only one to use for the decision.

TABLE I
VALUES OF ζ FOR DIFFERENT VALUES OF ϵ AND SAMPLE SIZE n

ϵ	$n = 1$	$n = 5$	$n = 10$	$n = 20$	$n = 50$	$n = \infty$
0.001	10.40	99.92	304.27	704.32	991.58	1000.00
0.005	6.49	38.56	89.33	161.05	199.11	200.00
0.010	5.22	24.98	51.52	84.24	99.67	100.00
0.050	3.04	8.61	13.64	18.25	19.97	20.00
0.100	2.38	5.32	7.55	9.37	9.99	10.00
0.500	1.30	1.66	1.85	1.96	2.00	2.00

IV. EXAMPLE

We now present the solution to the robust problem for the Gaussian nominal case. Let

$$g(x) = \left[(2\pi)^n |Q|\right]^{-1/2} \exp\left\{-\tfrac{1}{2}x^T Q^{-1} x\right\} \tag{23}$$

where Q is a nonsingular covariance matrix and $|Q|$ its determinant. The nominal likelihood ratio becomes

$$L_g(x) = \exp\left\{s^T Q^{-1} x - \tfrac{1}{2} s^T Q^{-1} s\right\}. \tag{24}$$

It is easy to see that $g(x)$ satisfies the unimodality assumptions and that $L_g(x)$ satisfies (7) with $\alpha_V = \infty$ and $\alpha_L = 0$. The sets C_α and D_α^k take the form

$$C_\alpha = \left\{x: s^T Q^{-1} x = \tfrac{1}{2} s^T Q^{-1} s + \ln\alpha\right\}$$
$$D_\alpha^k = \left\{x: \left(k - \tfrac{1}{2}\right) s^T Q^{-1} s + \ln\alpha \le s^T Q^{-1} x < \left(k + \tfrac{1}{2}\right) s Q^{-1} s + \ln\alpha\right\}. \tag{25}$$

The multiple integrals in (14) and (16) can be easily computed and take the form

$$\frac{1}{1-\xi}\Phi\left(\frac{\sigma - 2\ln\xi}{2\sqrt{\sigma}}\right) - \frac{\xi}{1-\xi}\Phi\left(\frac{-\sigma - 2\ln\zeta}{2\sqrt{\sigma}}\right) = \frac{1}{1-\epsilon} \tag{26}$$

and

$$\frac{\zeta}{\zeta-1}\Phi\left(\frac{\sigma + 2\ln\zeta}{2\sqrt{\sigma}}\right) - \frac{1}{\zeta-1}\Phi\left(\frac{-\sigma + 2\ln\zeta}{2\sqrt{\sigma}}\right) = \frac{1}{1-\epsilon} \tag{27}$$

where $\Phi(x)$ is the normalized Gaussian cumulative distribution and $\sigma = s^T Q^{-1} s$. Comparing (26) and (27) we can see that $\xi = 1/\zeta$. Notice also that the solutions to (26) and (27) depend only on σ. Letting $\sigma \to \infty$ the solution ζ satisfies $\zeta \to 1/\epsilon$. Let us now consider the case where Q is the identity matrix and all signals are equal to unity. This corresponds to an i.i.d. Gaussian noise sequence under nominal conditions. For this case $\sigma = n$. In Table I we give values of ζ for several values of ϵ and n.

V. CONCLUSION

We have presented the solution to the problem of detecting signals in additive dependent noise for the finite-sample case under the assumption that the noise density belongs to an ϵ-contamination model. It was shown that the robust detection scheme uses a censored version of the nominal likelihood ratio to make a decision. Even though the censored version is not the likelihood ratio of the least-favorable density, it nevertheless satisfies the saddle-point condition. This means that it achieves the same performance as the optimum detection structure for the least-favorable density. Obviously, this is not contradictory with the optimum Bayes theory because the censored version achieves the optimum performance only for a limited range of values of the threshold t ($t < 1$ or $t \geq 1$). The assumption that the observations are dependent is crucial. As we can see from the example, the observations are dependent under the least-favorable density even if they are independent under nominal conditions. For the i.i.d. nominal case we see that the robust detector structure censors the total likelihood ratio rather than each observation. This is basically a consequence of our contamination model in (5). What this model means is that either all observations are "good" (probability $1-\epsilon$) or all of them are "bad" (probability ϵ). This approach treats the observation vector as one unity and not its components separately. Clearly, a more realistic situation exists if at every instant we can have good or bad observations independently of the kinds of observations we have at the other instances. The resulting contaminated density for such a case under dependency has a much more complicated form than the one in (5).

Appendix
Proof of the Theorem

We first show existence of a solution to (14); similarly, one can show the existence of a solution to (16). Let ξ in (14) tend to α_L; we then have that A_ξ tends to the whole space and D_ξ^0 becomes the empty set. Thus the left side of (14) tends to unity, a quantity which is less than the right side. Notice now that because of (9) the set C_1 is nonempty, and thus the set D_1^0 has positive g-measure. Letting $\xi \to 1$, the left side of (14) tends to infinity. Using continuity, a $\xi \in [0,1)$ exists that satisfies (14).

To show uniqueness, call $\omega(\xi)$ the left side of (14). We will show that its derivative with respect to ξ is positive for $0 \leq \xi < 1$. The derivative has the form

$$\omega'(\xi) = \frac{d}{d\xi}\int_{A_\xi} g(x)\,dx + \frac{1}{1-\xi}\frac{d}{d\xi}\int_{D_\xi^0} g(x)\,dx + \frac{1}{(1-\xi)^2}\int_{D_\xi^0} g(x)\,dx. \tag{28}$$

If we call $\varphi(\xi)$ the probability under $g(x)$ of the event $\{L_g(x) \geq \xi\}$, then the first term in (28) becomes $\varphi'(\xi)$. For the second term, using the monotonicity of $L_g(x)$ at the direction of s, we can write

$$\begin{aligned}\int_{D_\xi^0} g(x)\,dx &= \int_{\substack{L_g(x)<\xi \\ L_g(x+s)\geq\xi}} g(x)\,dx \\ &= \int_{L_g(x)\geq\xi} \frac{g(x-s)}{g(x)} g(x)\,dx - \int_{L_g(x)\geq\xi} g(x)\,dx \\ &= \int_\xi^\infty (1-\tau)\varphi'(\tau)\,d\tau.\end{aligned} \tag{29}$$

Using (29), the second term in (28) is equal to $-\varphi'(\xi)$, and thus the derivative of $\omega(\xi)$ is equal to the third term in (28), which is positive.

Now we show that the density defined by (13) belongs to the ϵ-contamination class; similarly, we can show this for the density in (15). It is enough to show that

$$f_l(x) \geq (1-\epsilon)g(x) \tag{30}$$

for every vector x. When $x \in A_\xi \cup D_\xi^0$, we have equality in (30). For $x \in D_\xi^{-k}$ using (13) and the definition of the set D_ξ^{-k}, we must show that

$$g(x_c - qs)\xi^k \geq g(x_c - (k+q)s) \tag{31}$$

where $0 \leq q < 1$, $x_c \in C_\xi$ and $k \geq 1$. The inequality in (31) is a direct consequence of the convexity of $-\ln g(x)$ and the fact that $\xi = g(x_c - x)/g(x_c)$.

The last thing to be shown is that the density defined in (13) satisfies the saddle-point condition (6) and, more specifically, the right inequality. Remember that we are in the case $0 \leq t \leq 1$. To show the right inequality in (6), it is enough to show that

$$\int_{V_r^c} f_l(x-s)\,dx + t\int_{V_r} f_l(x)\,dx \geq \int_{V_r^c} f(x-s)\,dx + t\int_{V_r} f(x)\,dx \tag{32}$$

where V_r^c denotes the complement set of V_r. Inequality (32) is equivalent to

$$\int_{V_r}[f(x-s) - f_l(x-s)]\,dx \geq t\int_{V_r}[f(x) - f_l(x)]\,dx. \tag{33}$$

For $t \leq \xi$, (33) is trivially satisfied because V_r becomes the whole space. Now let $\xi < t \leq 1$. Notice that $V_r = \{x\colon L_g(x) \geq t\} \subset \{x\colon L_g(x) \geq \xi\} = A_\xi$. Since for $x \in A_\xi$ we have $x - s \in A_\xi \cup D_\xi^0$, from (13) we conclude that for $x \in V_r$ we have $f_l(x) = (1-\epsilon)g(x)$ and $f_l(x-s) = (1-\epsilon)g(x-s)$. Thus (33) is equivalent to

$$\int_{L_g(x)\geq t} h(x-s)\,dx \geq t\int_{L_g(x)\geq t} h(x)\,dx. \tag{34}$$

Changing variables in the first integral yields

$$\int_{L_g(x+s)\geq t} h(x)\,dx \geq t\int_{L_g(x)\geq t} h(x)\,dx. \tag{35}$$

Inequality (35) is true because we are in the case $t \leq 1$ and because $L_g(x)$ is increasing in the direction of s; thus the first integral is taken over a larger set. This concludes the proof.

Notice that the densities that achieve the lowest performance bound are those that satisfy (35) with equality. For the case where $t < 1$ we can have equality in (35) only when the two sides are equal to zero, in other words, if $h(x)$ places no mass on the set $\{x\colon L_g(x+s) \geq t\}$.

References

[1] R. D. Martin and S. C. Schwartz, "Robust detection of a known signal in nearly Gaussian noise," *IEEE Trans. Inform. Theory*, vol. IT-17, pp. 50–56, Jan. 1971.

[2] P. J. Huber, "A robust version of the probability ratio test," *Ann. Math. Statist.*, vol. 36, pp. 1753–1758, Dec. 1965.

[3] S. A. Kassam and J. B. Thomas, "Asymptotically robust detection of a known signal in contaminated non-gaussian noise," *IEEE Trans.*

Inform. Theory, vol. IT-22, pp. 22–26, Jan. 1976.
[4] S. A. Kassam, G. V. Moustakides, and J. G. Shin, "Robust detection of known signals in asymmetric noise," *IEEE Trans. Inform. Theory*, vol. IT-26, pp. 84–91, Jan. 1982.
[5] H. V. Poor, "Signal detection in the presence of weakly dependent noise—Part II: Robust detection," *IEEE Trans. Inform. Theory*, vol. IT-28, pp. 744–752, Sept. 1982.
[6] G. V. Moustakides and J. B. Thomas, "Min–max detection of weak signals in φ-mixing noise," *IEEE Trans. Inform. Theory*, vol. IT-30, pp. 529–537, May 1984.
[7] ——, "Optimum detection of a weak signal with minimal knowledge of dependency," *IEEE Trans. Inform. Theory*, vol. IT-32, pp. 97–102, Jan. 1986.
[8] G. V. Moustakides, "Robust detection of signals: A large deviations approach," *IEEE Trans. Inform. Theory*, vol. IT-31, pp. 822–825, Nov. 1985.