Robust Multiple Classification of Known Signals in Additive Noise—An Asymptotic Weak Signal Approach

Ola Hössjer and Moncef Mettiji

Abstract—**The problem of extracting one out of a finite number of possible signals of known form given observations in an additive noise model is considered. Two approaches are studied: either the signal with shortest distance to the observed data or the signal having maximal correlation with some transformation of the observed data is chosen. With a weak signal approach, the limiting error probability is a monotone function of the Pitman efficacy and it is the same for both the distance-based and correlation-based detectors. Using the minimax theory of Huber, it is possible to derive robust choices of distance/correlation when the limiting error probability is used as performance criterion. This generalizes previous work in the area, from two signals to an arbitrary number of signals. We consider M-type and R-type distances and also one-dimensional as well as two-dimensional signals. Finally, some Monte Carlo simulations are performed to compare the finite sample size error probabilities with the asymptotic error probabilities.**

Index Terms— **Additive noise, asymptotic error probability, known signals, M-statistic, maximum correlation, minimum distance, multiple classification, R-statistic, robustness, weak signals.**

I. Introduction

CONSIDER the following multiple classification problem:

$$H_i: Y_k = \alpha c_{ik} + e_k, \quad k = 1, \cdots, n, \quad i = 1, \cdots, p. \tag{1.1}$$

The observed vector $\boldsymbol{Y}_n = (Y_1, \cdots, Y_n)$ is a sum of one of p possible vectors $\alpha \boldsymbol{c}_{1n}, \cdots, \alpha \boldsymbol{c}_{pn}$ and a noise vector $\boldsymbol{e}_n = (e_1, \cdots, e_n)$. The amplitude factor α and $\boldsymbol{c}_{in} = (c_{i1}, \cdots, c_{in})$ are assumed to be known while e_i represents independent identically distributed (i.i.d.) random variables with distribution F.

Let ϕ_n: $\mathbb{R}^n \to \{1, \cdots, P\}$ be a nonrandomized decision rule with Borel measurable decision regions $\Omega_{ni} = \phi_n^{-1}(i)$. Assume that the *a priori* probability of H_i is $\pi_i > 0$, $i = 1, \cdots, p$, with $\sum \pi_i = 1$. It is well known that the optimal Bayes (maximum *a posteriori*) decision rule minimizing the error probability

$$P_e(\phi_n, F) = \sum_{i=1}^{p} \pi_i P(\boldsymbol{Y}_n \notin \Omega_{ni} | H_i) \tag{1.2}$$

is given by

$$\phi_n(\boldsymbol{Y}_n) = \arg\min_i \left(-\sum_{k=1}^{n} \log f(Y_k - \alpha c_{ik}) - \log \pi_i \right), \tag{1.3}$$

where $f = F'$ (cf. [18, p. 48]).

When F is unknown to some degree it is important to find robust decision rules for which $P_e(\phi_n, F)$ is insensitive with respect to variations in F. A minimax solution for $P_e(\phi_n, F)$ is obtained in [13] when $p = 2$ and the distribution of the noise components is allowed to vary independently under both hypotheses. Unfortunately, this approach is difficult to generalize to arbitrary p. Instead, we will consider the asymptotic weak signal approach, where the signal amplitude α decreases with the sample size as

$$\alpha = \alpha_n = K/\sqrt{n}, \tag{1.4}$$

given some constant $K > 0$. For reasonable decision rules, the error probability then converges as $n \to \infty$ to a nonzero limit. An analysis of the limiting value of the error probability is relatively straightforward, since it is closely related to the Pitman efficacy of the detector test statistic. Therefore, the classical robust minimax approach introduced by Huber [7] may be used to analyze the detector. This connection between the asymptotic error probability and the efficacy has been utilized frequently in the literature when $p = 2$, using power-level type criteria, see for instance [2], [12], and [13] or [11] for an overview.

In this paper, we extend the correspondence between the efficacy and asymptotic error probability to $p > 2$, using P_e as performance measure, which is more appropriate in many communication problems. The asymptotic weak signal approach also makes it possible to find approximate expressions for $P_e(\phi_n, F)$ when n is fixed. The exact error probabilities are usually hard to calculate when F is not Gaussian.

A practical limitation of the weak signal approach is the decrease (1.4) of the signal amplitude with the sample size, since this does not correspond to collecting more and more samples of a fixed signal, but rather a new signal is considered

Manuscript received December 20, 1990; revised November 19, 1991. This work was supported by the Swedish Board for Technical Development under Contract 712-89-1073.

O. Hössjer is with the School of Operations Research and Industrial Engineering, Cornell University, Ithaca, NY 14853-3801.

M. Mettiji is with the Circuits and Systems Group, Department of Technology, Uppsala University, P.O. Box 534, S-751 21 Uppsala, Sweden.

IEEE Log Number 9203029.

for each n. On the other hand, it is interesting with a detector that classifies weak signals, since these are on the border of not being distinguishable, and therefore, the error probabilities are especially important to know in this case. In addition, the weak signal approach is mathematically more tractable than an asymptotic analysis based on large deviation theory.

We will consider two kinds of detector structures.

1) *Minimum Distance Detectors:* Given some distance measure in $\mathbb{R}^n$, choose the signal that is closest to the received data.
2) *Maximum Correlation Detectors:* Choose the signal that has maximal correlation with some transformation of the received data.

In Section II, we define in a general framework distance measures in $\mathbb{R}^n$. In particular, we will consider detectors based on M-distances and R-distances and we investigate the asymptotic properties of the corresponding detector test statistics in Section III. In Section IV, we show that the asymptotic error probability is essentially described by the Pitman efficacy. This is utilized in Section V in order to find robust detectors. The set-up is generalized in Section VI to two-dimensional signals. In Section VII, we compare the results with Monte Carlo simulations. Finally, some technical results from the asymptotic theory of the test statistics are collected in the appendix.

II. General Detector Structures

Let D_n: $\mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ be a distance measure between vectors in $\mathbb{R}^n$. We assume that D_n is translation invariant, i.e.,

$$D_n(\boldsymbol{x}+\boldsymbol{z}.\ \boldsymbol{y}+\boldsymbol{z}) = D_n(\boldsymbol{x}.\ \boldsymbol{y})$$

for any vectors $\boldsymbol{x}$, $\boldsymbol{y}$, and $\boldsymbol{z}$ in $\mathbb{R}^n$. Without ambiguity, we will write $D_n(\boldsymbol{x}.\ \boldsymbol{y}) = D_n(\boldsymbol{x}-\boldsymbol{y})$ in the sequel. We collect the signals into a $n \times p$ signal matrix $\boldsymbol{C}_n = (\boldsymbol{c}'_{1n}, \cdots, \boldsymbol{c}'_{pn})$, and define

$$D_n^i(\boldsymbol{x}) = -\left[\frac{\partial D_n(\boldsymbol{x}-\boldsymbol{\theta}\boldsymbol{C}'_n)}{\partial \theta_i}\right]_{\theta=0} = \sum_{k=1}^{n} c_{ik}\frac{\partial D_n(\boldsymbol{x})}{\partial x_k}, \qquad i=1,\cdots,p, \tag{2.1}$$

with $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_p)$. Our two detectors may now be specified as follows (see also Fig. 1), with $\boldsymbol{Y}_n = (Y_1, \cdots, Y_n)$.

1) *Minimum Distance Detector:*

$$\phi_n(\boldsymbol{Y}_n) = \arg\min_i (D_n(\boldsymbol{Y}_n - \alpha_n \boldsymbol{c}_{in}) - L\log\pi_i). \tag{2.2}$$

Note that if we introduce

$$T_{ni}^D(\boldsymbol{Y}_n) = D_n(\boldsymbol{Y}_n) - D_n(\boldsymbol{Y}_n - \alpha_n\boldsymbol{c}_{in}) \tag{2.3}$$

with $\boldsymbol{T}_n^D(\boldsymbol{Y}_n) = (T_{n1}^D, \cdots, T_{np}^D)$, then the minimum distance detector can be written as

$$\phi_n(\boldsymbol{Y}_n) = \arg\max_i (T_{ni}^D(\boldsymbol{Y}_n) + L\log\pi_i). \tag{2.4}$$

The reason for rewriting the minimum distance decision rule (2.2) into the form (2.4) is that the asymptotic behavior of the random vector $\boldsymbol{T}_n^D$ is more natural to study than that of the vector $(D_n(\boldsymbol{Y}_n - \alpha_n\boldsymbol{c}_{1n}), \cdots, D_n(\boldsymbol{Y}_n - \alpha_n\boldsymbol{c}_{pn}))$.

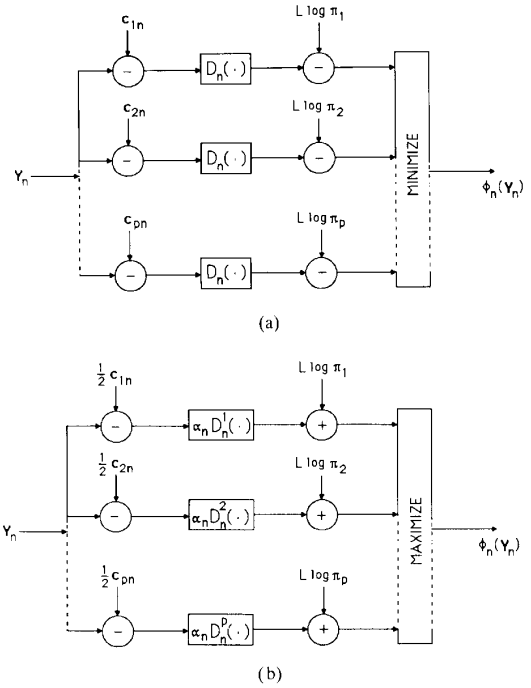

Fig. 1. General detector structures for (a) minimum distance detectors and (b) maximum correlation detectors. In (b) the first subtraction of $\frac{1}{2}\alpha_n\boldsymbol{c}_{in}$ from $\boldsymbol{Y}_n$ may be omitted if all signals have the same L^2-norm (see the remark after Lemma 1).

2) *Maximum Correlation Detector:*

$$\phi_n(\boldsymbol{Y}_n) = \arg\max_i \left(\alpha_n D_n^i\left(\boldsymbol{Y}_n - \frac{1}{2}\alpha_n\boldsymbol{c}_{in}\right) + L\log\pi_i\right). \tag{2.5}$$

Note that if we introduce

$$T_{ni}^C(\boldsymbol{Y}_n) = \alpha_n D_n^i\left(\boldsymbol{Y}_n - \frac{1}{2}\alpha_n\boldsymbol{c}_{in}\right) \tag{2.6}$$

with $\boldsymbol{T}_n^C(\boldsymbol{Y}_n) = (T_{n1}^C, \cdots, T_{np}^C)$, then the maximum correlation detector can be written as

$$\phi_n(\boldsymbol{Y}_n) = \arg\max_i (T_{ni}^C(\boldsymbol{Y}_n) + L\log\pi_i). \tag{2.7}$$

A first-order Taylor expansion of $D_n(\boldsymbol{Y}-\boldsymbol{\theta}\boldsymbol{C}'_n)$ with respect to θ_i around the point $\frac{1}{2}\alpha_n\boldsymbol{u}_i$ gives $T_{ni}^D(\boldsymbol{Y}_n) \approx T_{ni}^C(\boldsymbol{Y}_n)$, so we expect that the minimum distance and maximum correlation detectors perform similarly when n is large and $\alpha_n = K/\sqrt{n}$ small.

In order to make ϕ_n well defined in (2.2) and (2.5) when two or more of the quantities in brackets equal the maximum, we pick (say) the quantity with lowest index for definiteness. The norming constant L is specified from case to case. It is unity for the optimum detector (1.3) and the robust detectors in Sections V-B and VI-C. Note that the term $L\log\pi_i$ may be dropped when all *a priori* probabilities are equal, since this

does not change the decision rule. It follows from (2.1) that $D_n^i(\boldsymbol{Y}_n - \frac{1}{2}\alpha_n \boldsymbol{c}_{in})$ takes the form of a correlation between $\boldsymbol{c}_{in}$ and $\boldsymbol{Y}_n - \frac{1}{2}\alpha_n \boldsymbol{c}_{in}$, which explains the name of the detector (2.5).

Example 1) M-Distance: Given a function ρ: $\mathbb{R} \to \mathbb{R}$, we define the *M-distance* as

$$D_{M,n}(\boldsymbol{x}) = \sum_{k=1}^{n} \rho(x_k), \tag{2.8}$$

where $\boldsymbol{x} = (x_1, \cdots, x_n)$. For $\rho' = \psi$, we obtain

$$D_{M,n}^i(\boldsymbol{x}) = \sum_{k=1}^{n} c_{ik}\psi(x_k), \tag{2.9}$$

which may be regarded as a correlation between $\boldsymbol{c}_{in}$ and the vector $(\psi(x_1), \cdots, \psi(x_n))$. The canonical choice $\rho(x) = x^2$ yields the (squared) L^2-distance. Simple calculations show that (2.2) and (2.5) are actually algebraically equivalent for this distance. Huber's ρ-function,

$$\rho_a(x) = \begin{cases} x^2/2, & |x| \le a, \\ a|x| - a^2/2, & |x| > a, \end{cases} \tag{2.10}$$

defines a family of distances. Note that outliers of magnitude greater than a are weighted down by this distance. Since the *M*-distance is generally not scale invariant, we assume that a scale parameter of the error distribution has already been estimated for some training data. The L^2-distance corresponds to the case when $a = \infty$ and there is no protection against outliers. The L^1-distance corresponds to the case when $a \to 0+$. Here, $\rho(x) = |x|$ and there is high protection against outliers.

Example 2) R-Distance: Given $\boldsymbol{x} = (x_1, \cdots, x_n)$, let R_k^+ be the rank of $|x_k|$ among $|x_1|, \cdots, |x_n|$. The *R-distance* is defined as

$$D_{R,n}(\boldsymbol{x}) = \sum_{k=1}^{n} a_n(R_k^+)|x_k|, \tag{2.11}$$

where $a_n(1), \cdots, a_n(n)$ are appropriately chosen scores. The derivatives are given by

$$D_{R,n}^i(\boldsymbol{x}) = \sum_{k=1}^{n} c_{ik} \operatorname{sgn}(x_k) a_n(R_k^+), \tag{2.12}$$

which is a correlation between $\boldsymbol{c}_{in}$ and the vector $(\operatorname{sgn}(x_1)a_n(R_1^+), \cdots, \operatorname{sgn}(x_n)a_n(R_n^+))$ of signed ranks. The scores $a_n(\cdot)$ are generated from a function h: $(0, 1) \to \mathbb{R}$ according to

$$a_n(k) = h\left(\frac{k}{n+1}\right). \tag{2.13}$$

Examples of such scores are

a) Wilcoxon scores, $h(u) = u$;
b) van der Waerden normal scores, $h(u) = \Phi^{-1}((u+1)/2)$, where Φ is the c.d.f. of the standard normal distribution;
c) truncated normal scores, $h_b(u) = \min(\Phi^{-1}((u+1)/2), b)$, given some $b > 0$, introduced by Rousseeuw ([15], [16]);
d) Jaeckel's minimax solution in [9], $\hat{h}(u) = \min(\Phi^{-1}((u+1-\epsilon)/2(1-\epsilon)), b)$.

Also note that $h \equiv 1$ yields (up to scaling) the L^1-norm. Since the *R*-distance is scale equivariant, there is no need for a preliminary estimate of scale, as for the *M*-distance. For an account of *R*-distances, see [6].

Remark: We close the section by remarking that our signal model can be imbedded into a linear regression model,

$$\boldsymbol{Y}_n = \boldsymbol{\theta}\boldsymbol{C}_n' + \boldsymbol{e}_n,$$

where $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_p)$ is the unknown regression parameter and $\boldsymbol{e}_n = (e_1, \cdots, e_n)$ the noise vector. The hypothesis testing problem (1.1) is now written as

$$H_i: \boldsymbol{\theta} = \alpha_n \boldsymbol{u}_i, \qquad i = 1, \cdots, p,$$

with $\boldsymbol{u}_i$ being the unit vector in $\mathbb{R}^p$ with one in position i. A detector may then be constructed by first estimating $\boldsymbol{\theta}$ from the data according to

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} D_n(\boldsymbol{Y}_n - \boldsymbol{\theta}\boldsymbol{C}_n')$$

and then make a decision according to

$$\begin{aligned} \phi_n(\boldsymbol{Y}_n) &= \arg\min_i \left\| (\hat{\boldsymbol{\theta}} - \alpha_n \boldsymbol{u}_i)\boldsymbol{C}_n' \right\| \\ &= \arg\min_i \left\| \hat{\boldsymbol{\theta}}\boldsymbol{C}_n' - \alpha_n \boldsymbol{c}_{in} \right\|, \end{aligned}$$

where $\|\cdot\|$ is the Euclidean norm in $\mathbb{R}^n$. However, the computational burden of estimating $\boldsymbol{\theta}$ is often quite large, so we will prefer the other two detector structures.

III. Asymptotic Properties

In this section, we investigate the asymptotic behavior of the statistics $\boldsymbol{T}_n^D$ and $\boldsymbol{T}_n^C$ introduced in (2.3) and (2.6), and start with the following regularity conditions.

1) The distribution F of the random variables e_i is symmetric with an absolutely continuous density f such that the Fisher information $I(f) = \int_{-\infty}^{\infty} (f'^2(x)/f(x))\,dx$ is finite.
2) The signal matrices $\boldsymbol{C}_n$ satisfy $\lim_{n\to\infty} \boldsymbol{C}_n'\boldsymbol{C}_n/n = \boldsymbol{\Sigma}$, where $\boldsymbol{\Sigma} = (\sigma_{ij})_{i,j=1}^p$ is a symmetric, positive semidefinite $p \times p$ matrix. Let also $\boldsymbol{\sigma}_1, \cdots, \boldsymbol{\sigma}_p$ be the row vectors of $\boldsymbol{\Sigma}$ and $\boldsymbol{\sigma} = (\sigma_{11}, \cdots, \sigma_{pp})$ the vector of diagonal elements.

A. M-Detectors

Assume that $D_n = D_{M,n}$ (cf. (2.8)), with ρ absolutely continuous and a derivative ψ that is skew symmetric[1] and satisfies

$$0 < A_M(\psi, F) = \int \psi^2(x)\,dF(x) < \infty, \tag{3.1}$$

$$\int (\psi(x+h) - \psi(x))^2\,dF(x) = o(1) \quad \text{as } h \to 0 \tag{3.2}$$

and

$$\int \psi(x+h)\,dF(x) = hB_M(\psi, F) + o(h), \tag{3.3}$$

[1] It is actually only necessary that $\int \psi(x)\,dF(x) = 0$ in Lemma 1, but in Section V, we need this stronger requirement.

for some $B_M(\psi, F) > 0$. If, for instance, ψ is absolutely continuous with ψ' bounded, it follows that $B_M(\psi, F) = \int \psi' \, dF$. Assume also that

$$\int (\rho(x+h+q) - \rho(x+h-q) - 2q\psi(x+h))^2 \, dF(x) = o(h^2+q^2), \quad (3.4)$$

and

$$\int (\rho(x+h+q) - \rho(x+h-q) - 2q\psi(x+h)) \, dF(x) = o(h^2+q^2). \quad (3.5)$$

as $h, q \to 0$. Conditions (3.2)–(3.5) are satisfied if ψ is discontinuous at most at a finite number of points and absolutely continuous with a bounded derivative ψ' outside the points of discontinuity. Conditions (3.2)–(3.3) are the same as those in [1].

We may now prove asymptotic normality for the test statistics $\boldsymbol{T}_{M,n}^D$ and $\boldsymbol{T}_{M,n}^C$ introduced in (2.3) and (2.6), where the subscript M indicates that the distance measure is of M-type.

Lemma 1: Suppose that D_n is the M-distance defined in (2.8). Then, under *(F), (C),* and (3.1)–(3.3),

$$\boldsymbol{T}_{M,n}^C \xrightarrow{d} N(\boldsymbol{\mu}_i, K^2 A_M(\psi, F)\boldsymbol{\Sigma}) \quad (3.6)$$

under H_i, where the mean vector is given by

$$\boldsymbol{\mu}_i = K^2 B_M(\psi, F)\left(\boldsymbol{\sigma}_i - \frac{1}{2}\boldsymbol{\sigma}\right), \quad (3.7)$$

and with $\boldsymbol{\sigma}_i$ and $\boldsymbol{\sigma}$ defined in (*C*). If ρ satisfies (3.4)–(3.5) as well, (3.6)–(3.7) also hold with $\boldsymbol{T}_{M,n}^D$ in place of $\boldsymbol{T}_{M,n}^C$.

Proof: See the Appendix. □

Remark: In Lemma 1, T_{nj}^C is defined by differentiating $D_n(\boldsymbol{Y}_n - \boldsymbol{\theta}\boldsymbol{C}_n')$ with respect to θ_j around the point $\boldsymbol{\theta} = \frac{1}{2}\alpha_n \boldsymbol{u}_j$. When all *a priori* probabilities are equal, this choice of differentiation point always yields asymptotically optimal decision regions among all decision rules based on $\boldsymbol{T}_n^C$, see Proposition 2 and the remark following. If instead $\boldsymbol{\theta} = 0$ is chosen as the differentiation point, it follows as in the proof of Lemma 1 that (3.6) holds with $\boldsymbol{\mu}_i = K^2 B_M(\psi, F)\boldsymbol{\sigma}_i$. Hence, the mean vector under H_i, $\boldsymbol{\mu}_i$ is translated in a direction proportional to $\boldsymbol{\sigma} = (\sigma_{11}, \cdots, \sigma_{pp})$ for $i = 1, \cdots, p$. Since the decision rule picks the maximum of $T_{nj}^C + L \log \pi_j$ it is easily seen that the asymptotic error probability is unaffected after the change of differentiation point only when $\sigma_{11} = \cdots = \sigma_{pp}$. This corresponds to (cf. (C)) that the signals $\alpha_n \boldsymbol{c}_{1n}, \cdots, \alpha_n \boldsymbol{c}_{pn}$ have equal asymptotic L^2-norm. This remark is also valid for R-statistics and two-dimensional signals.

B. R-Detectors

Assume that $D_n = D_{R,n}$ (cf. (2.11)), with scores $a_n(i)$ generated from a function h according to (2.13). Also suppose that h is nonnegative, nondecreasing,

$$0 < A_R(h) = \int_0^1 h(u)^2 \, du < \infty, \quad (3.8)$$

$$0 < B_R(h, F) = \int_0^1 h(u) h_F(u) \, du < \infty, \quad (3.9)$$

where

$$h_F(u) = -\frac{f'\left(F^{-1}\left(\frac{u+1}{2}\right)\right)}{f\left(F^{-1}\left(\frac{u+1}{2}\right)\right)}. \quad (3.10)$$

As for M-detectors, we may prove asymptotic normality for the test statistics $\boldsymbol{T}_{R,n}^D$ and $\boldsymbol{T}_{R,n}^C$, with the subscript R indicating that the distance measure is generated from ranks.

Lemma 2: Suppose that D_n is the R-distance defined in (2.11) with scores generated according to (2.13). Then, under (*F*), (*C*), and (3.8)–(3.10),

$$\boldsymbol{T}_{R,n}^C \xrightarrow{d} N(\boldsymbol{\mu}_i, K^2 A_R(h)\boldsymbol{\Sigma}) \quad (3.11)$$

under H_i, where the mean vector satisfies

$$\boldsymbol{\mu}_i = K^2 B_R(h, F)\left(\boldsymbol{\sigma}_i - \frac{1}{2}\boldsymbol{\sigma}\right), \quad (3.12)$$

with $\boldsymbol{\sigma}_i$ and $\boldsymbol{\sigma}$ defined in (*C*). Furthermore, (3.11)–(3.12) hold with $\boldsymbol{T}_{R,n}^D$ in place of $\boldsymbol{T}_{R,n}^C$.

Proof: See the Appendix. □

IV. Asymptotic Error Probabilities

Having derived the asymptotic properties of the test statistics in Section III, we are now ready to analyze the asymptotic error probabilities. Let $\phi = \{\phi_n\}_{n=1}^{\infty}$ symbolize a sequence of decision rules. For such a sequence, we let

$$\overline{P}_e(\phi, F) = \limsup_{n\to\infty} P_e(\phi_n, F), \quad (4.1)$$

the lim sup of the error probabilities, be our performance criterion.

We are interested in decision rules generated from M-distances/R-distances according to (2.2)–(2.7). It is clear from Lemmas 1–2 that the maximum correlation and minimum distance test statistics may be treated within a unified framework, as may M-statistics and R-statistics. For this reason, we simply write $\boldsymbol{T}_n$ in the sequel referring either to $\boldsymbol{T}_{M,n}^D$, $\boldsymbol{T}_{M,n}^C$, $\boldsymbol{T}_{R,n}^D$, or $\boldsymbol{T}_{R,n}^C$. Similarly, A may stand for $A_M(\psi, F)$ or $A_R(h)$ and B for either $B_M(\psi, F)$ or $B_R(h, F)$.

We say that a sequence of decision rules ϕ is generated from $\{\boldsymbol{T}_n\}$ if there exists a decomposition of $\mathbb{R}^p$ into p disjoint Borel measurable subsets $G_1, \cdots, G_p$ such that

$$\phi_n^{-1}(i) = \Omega_{ni} = \boldsymbol{T}_n^{-1}(G_i), \qquad i = 1, \cdots, p, \quad (4.2)$$

so that H_i is chosen whenever $\boldsymbol{T}_n$ falls into G_i.

In order to facilitate the computation of $\overline{P}_e(\phi, F)$ when ϕ is generated from $\{\boldsymbol{T}_n\}$, it is appropriate to transform $\boldsymbol{T}_n$ so that the covariance matrix becomes radially symmetric. In particular, this makes it easy to see how $\overline{P}_e$ depends on A and B. According to Lemmas 1–2 we have $\boldsymbol{T}_n \xrightarrow{d} \boldsymbol{Z}$, where

$$\boldsymbol{Z} \sim N(\boldsymbol{\mu}_i, K^2 A\boldsymbol{\Sigma}) \quad (4.3)$$

under H_i and $\boldsymbol{\mu}_i = K^2 B(\boldsymbol{\sigma}_i - \frac{1}{2}\boldsymbol{\sigma})$. The rank-$q$ matrix $\boldsymbol{\Sigma}$ may be diagonalized as

$$\boldsymbol{\Sigma} = \boldsymbol{Q}'\boldsymbol{D}\boldsymbol{Q}, \quad (4.4)$$

where $\boldsymbol{Q}$ is a $p \times p$ orthogonal matrix and $\boldsymbol{D} = \operatorname{diag}(d_1, \cdots, d_p)$, a diagonal matrix with diagonal elements $d_1 \geq \cdots \geq d_q > 0$ and $d_{q+1} = \cdots = d_p = 0$. Then define the affine transformation $\boldsymbol{\Lambda}$: $\mathbb{R}^p \to \mathbb{R}^q$ by

$$\boldsymbol{\Lambda}(\boldsymbol{x}) = \frac{1}{KB}\left(\boldsymbol{x} + \frac{1}{2}K^2 B\boldsymbol{\sigma}\right)\boldsymbol{H}, \tag{4.5}$$

where

$$\boldsymbol{H} = \boldsymbol{Q}'\begin{pmatrix} 1/\sqrt{d_1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1/\sqrt{d_q} \\ 0 & \cdots & 0 \\ \vdots & \vdots & \vdots \\ 0 & \cdots & 0 \end{pmatrix}$$

(i.e., $\boldsymbol{H} = \boldsymbol{\Sigma}^{-1/2}$ when $q = p$). Transforming $\boldsymbol{Z}$ we obtain $\tilde{\boldsymbol{T}}_n(\boldsymbol{Y}_n) = \boldsymbol{\Lambda}(\boldsymbol{T}_n(\boldsymbol{Y}_n)) \xrightarrow{d} \tilde{\boldsymbol{Z}}$, where

$$\tilde{\boldsymbol{Z}} = \boldsymbol{\Lambda}(\boldsymbol{Z}) \sim N(\tilde{\boldsymbol{\mu}}_i, \mathcal{E}^{-1}\boldsymbol{I}_q) \tag{4.6}$$

under H_i,

$$\tilde{\boldsymbol{\mu}}_i = \boldsymbol{\Lambda}(\boldsymbol{\mu}_i) = K\boldsymbol{u}_i\boldsymbol{\Sigma}\boldsymbol{H}, \tag{4.7}$$

$\boldsymbol{I}_q$ is the $q \times q$ identity matrix and

$$\mathcal{E} = B^2/A \tag{4.8}$$

is the Pitman efficacy. This quantity $\mathcal{E}$ only depends on ψ and F (or h and F) and it measures the concentration of the random vector $\tilde{\boldsymbol{Z}}$ around $\tilde{\boldsymbol{\mu}}_i$. On the other hand, the Euclidean distances between the center of gravity points under different hypotheses depend on the signal characteristics only, since

$$\begin{aligned}\|\tilde{\boldsymbol{\mu}}_i - \tilde{\boldsymbol{\mu}}_j\|^2 &= K^2(\sigma_{ii} - 2\sigma_{ij} + \sigma_{jj}) \\ &= \lim_{n\to\infty} \frac{K^2}{n}\sum_{k=1}^{n}(c_{ik} - c_{jk})^2,\end{aligned} \tag{4.9}$$

which is the asymptotic Euclidean distance between $\alpha_n \boldsymbol{c}_{in}$ and $\alpha_n \boldsymbol{c}_{jn}$.[2] Consequently, the signals are separated from the detector and noise characteristics (F and ψ or h) in terms of means and covariance matrices of $\tilde{\boldsymbol{Z}}$, respectively. Since the signal characteristics are fixed, so are all $\tilde{\boldsymbol{\mu}}_i$. We will see in Section V that this simplifies the analysis of $\overline{P}_e(\phi, F)$.

It follows from (4.3)–(4.4) that $P(Z \in M) = 1$ under $H_1, \cdots, H_p$, where

$$M = \operatorname{Span}\{\boldsymbol{u}_1\boldsymbol{Q}, \cdots, \boldsymbol{u}_q\boldsymbol{Q}\} - \frac{1}{2}K^2 B\boldsymbol{\sigma} \tag{4.10}$$

is a q-dimensional hyperplane in $\mathbb{R}^p$. Given a decomposition $\{G_i\}_{i=1}^p$ of $\mathbb{R}^p$, let

$$\tilde{G}_i = \boldsymbol{\Lambda}(G_i \cap M). \tag{4.11}$$

When $p = q$ we have $M = \mathbb{R}^p$, $\tilde{G}_i = \boldsymbol{\Lambda}(G_i)$, and since $\boldsymbol{\Lambda}$ is then bijective it follows that $\phi_n^{-1}(i) = \tilde{\boldsymbol{T}}_n^{-1}(\tilde{G}_i)$, so that H_i is chosen whenever $\tilde{\boldsymbol{T}}_n \in \tilde{G}_i$ and the sets $\tilde{G}_i$ correspond to decision regions in $\mathbb{R}^q$. This may *not* be true when $q < p$ since $\boldsymbol{T}_n$ need not be concentrated to M for a fixed n and $\boldsymbol{\Lambda}$ is not injective outside M. Actually, $\boldsymbol{\Lambda}(G_i) = \mathbb{R}^q$ may hold in this case, and therefore the image of different G_i under $\boldsymbol{\Lambda}$ overlap. However, the following proposition implies that the sets $\tilde{G}_i$ correspond asymptotically to decision regions in $\mathbb{R}^q$ as far as computation of error probabilities is concerned, even when $q < p$.

Proposition 1: Given a decomposition $\{G_i\}_{i=1}^p$ of $\mathbb{R}^p$, let ϕ be the sequence of decision rules generated from $\{\boldsymbol{T}_n\}$ according to (4.2) and suppose that[3]

$$P(\boldsymbol{Z} \in \partial G_i | H_i) = 0, \qquad i = 1, \cdots, p. \tag{4.12}$$

Then,

$$\begin{aligned}\overline{P}_e(\phi, F) &= \lim_{n\to\infty} P_e(\phi_n, F) \\ &= \sum_{i=1}^{p} \pi_i P(\tilde{\boldsymbol{Z}} \notin \tilde{G}_i | H_i),\end{aligned} \tag{4.13}$$

with $\tilde{G}_i$ defined in (4.11).

Proof: Since $\boldsymbol{T}_n \xrightarrow{d} \boldsymbol{Z}$, it follows from (4.12) and the relation $P(\boldsymbol{Z} \in M) = 1$ (cf. (4.10)) that

$$\begin{aligned}&\lim_{n\to\infty} \sum_{i=1}^{p} \pi_i P(\boldsymbol{T}_n \notin G_i | H_i) \\ &= \sum_{i=1}^{p} \pi_i P(\boldsymbol{Z} \notin G_i | H_i) \\ &= \sum_{i=1}^{p} \pi_i P(\boldsymbol{Z} \in G_i^c \cap M | H_i).\end{aligned} \tag{4.14}$$

Since $\boldsymbol{\Lambda}$ is surjective and

$$\boldsymbol{\Lambda}^{-1}(\boldsymbol{\Lambda}(\boldsymbol{x})) = \boldsymbol{x} + \operatorname{Span}\{\boldsymbol{u}_{q+1}\boldsymbol{Q}, \cdots, \boldsymbol{u}_p\boldsymbol{Q}\},$$

it follows that the restriction $\boldsymbol{\Lambda}_M$: $M \to \mathbb{R}^q$ of $\boldsymbol{\Lambda}$ to M is a bijection and hence

$$\boldsymbol{\Lambda}_M^{-1}(\tilde{G}_i^c) = G_i^c \cap M \tag{4.15}$$

because of (4.11). But (4.15) implies that

$$\begin{aligned}P(\boldsymbol{Z} \in G_i^c \cap M | H_i) &= P(\boldsymbol{Z} \in \boldsymbol{\Lambda}_M^{-1}(\tilde{G}_i^c) | H_i) \\ &= P(\tilde{\boldsymbol{Z}} \in \tilde{G}_i^c | H_i),\end{aligned}$$

which together with (4.14) proves the proposition. □

Fig. 2 illustrates the decision regions in the spaces $\mathbb{R}^n$, $\mathbb{R}^p$, and $\mathbb{R}^q$ for decision rules generated from $\{\boldsymbol{T}_n\}$ according to (4.2). Note that the actual decision is made in $\mathbb{R}^p$, and the transformation $\boldsymbol{\Lambda}$ is performed in order to facilitate the computation of asymptotic error probabilities.

The decision rules (2.4) (and (2.7)) are generated from $\{\boldsymbol{T}_n\}$, with

$$\begin{aligned}G_i &= G_i(\boldsymbol{\pi}, L) \\ &= \left\{\boldsymbol{x} \in \mathbb{R}^p; i = \arg\max_j (x_j + L \log \pi_j)\right\}\end{aligned} \tag{4.16}$$

and $\boldsymbol{\pi} = (\pi_1, \cdots, \pi_p)$.

[2] The choice of $\boldsymbol{\Lambda}$ is not unique if two of the nonzero eigenvalues of $\boldsymbol{\Sigma}$ coincide, since $\boldsymbol{Q}$ is not unique in this case. However, this does not affect the covariance matrix of $\tilde{\boldsymbol{Z}}$ and the pairwise distances between $\tilde{\boldsymbol{\mu}}_i$ and $\tilde{\boldsymbol{\mu}}_j$.

[3] In (4.12), $\partial G_i = cl(G_i) \cap cl(G_i^c)$ denotes the boundary of G_i, ^{c} is the set-theoretic complement and $cl(\cdot)$ the closure of the set.

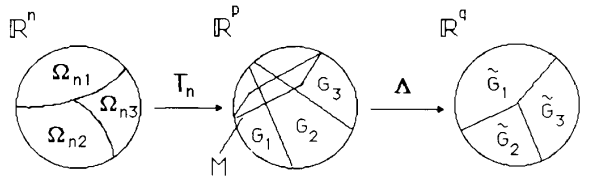

Fig. 2. Decision regions in the spaces $\mathbb{R}^n$, $\mathbb{R}^p$, and $\mathbb{R}^q$ for a decision rule ϕ_n generated from $\boldsymbol{T}_n$ according to (4.2).

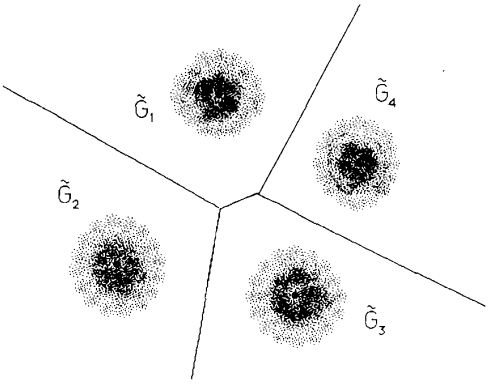

Fig. 3. Distribution of $\tilde{Z}$ under different hypotheses and the regions $\tilde{G}_i$ (when $\pi_1 = \cdots = \pi_p = 1/p$).

Proposition 2: Suppose $G_i = G_i(\boldsymbol{\pi}, L)$ is given by (4.16). Then the transformed regions $\tilde{G}_i = \tilde{G}_i(\boldsymbol{\pi}, L)$ in $\mathbb{R}^q$ according to (4.11) are given by (cf. Fig. 3)

$$\tilde{G}_i(\boldsymbol{\pi}, L) = \left\{\boldsymbol{y}\in\mathbb{R}^q; i = \arg\min_j \left(\frac{1}{2}\|\boldsymbol{y} - \tilde{\boldsymbol{\mu}}_j\|^2 - \frac{L}{B}\log \pi_j\right)\right\}. \quad (4.17)$$

Moreover, (4.12) is valid, so that $\overline{P}_e$ is given by (4.13).

Proof: One may verify directly that $\boldsymbol{\Lambda}$ transforms $G_i \cap M$ to the set $\tilde{G}_i$ in (4.17). We prefer to show this indirectly in the following way: Let $\boldsymbol{z}$ be an observation of the random vector $\boldsymbol{Z}$, whose multivariate normal distribution under H_i is given by (4.3). We know that $P(\boldsymbol{Z}\in M) = 1$ and that the density of $\boldsymbol{Z}$ under H_i is

$$f_i(\boldsymbol{x}) = \frac{1}{(2\pi)^{q/2}\left(K^2 A\prod_1^q d_j\right)^{1/2}} \exp\left(-\frac{1}{2K^2 A}(\boldsymbol{x} - \boldsymbol{\mu}_i)\boldsymbol{\Sigma}^{-1}(\boldsymbol{x} - \boldsymbol{\mu}_i)'\right), \quad \boldsymbol{x}\in M, \quad (4.18)$$

where

$$\boldsymbol{\Sigma}^{-1} = \boldsymbol{Q}'\,\mathrm{diag}\,(d_1^{-1}, \cdots, d_q^{-1}, 0, \cdots, 0)\boldsymbol{Q}$$

is a (pseudo)inverse of $\boldsymbol{\Sigma}$. Suppose now that given $\boldsymbol{z}$ we want to determine H_i and that the cost of making an error is C_i under H_i. The optimal decision regions F_i that minimize the average cost

$$\overline{C} = \sum_{i=1}^p C_i \pi_i P(\boldsymbol{z}\notin F_i | H_i)$$

are given by (cf. [18, p. 48])

$$F_i = \left\{\boldsymbol{x}\in M; i = \arg\max_j C_j\pi_j f_j(\boldsymbol{x})\right\} = \left\{\boldsymbol{x}\in M; i = \arg\max_j \left(x_j + \frac{A}{B}\log(C_j\pi_j)\right)\right\}. \quad (4.19)$$

where the last equality follows from (4.18). On the other hand, we may make the decision after observing $\tilde{\boldsymbol{z}} = \boldsymbol{\Lambda}(\boldsymbol{z})$. Here, $\tilde{\boldsymbol{z}}$ is an observation of $\tilde{\boldsymbol{Z}}$, which is distributed according to (4.6) under H_i. Let $\tilde{f}_i(\boldsymbol{y})$ denote the corresponding density. In analogy with (4.19), the optimal choices of decision regions when observing $\tilde{\boldsymbol{z}}$ are

$$\tilde{F}_i = \left\{\boldsymbol{y}\in\mathbb{R}^q; i = \arg\max_j C_j\pi_j\tilde{f}_j(\boldsymbol{y})\right\} = \left\{\boldsymbol{y}\in\mathbb{R}^q; i = \arg\min_j \left(\frac{1}{2}\|\boldsymbol{y} - \tilde{\boldsymbol{\mu}}_j\|^2 - \frac{A}{B^2}\log(C_j\pi_j)\right)\right\}. \quad (4.20)$$

Since F_i and $\tilde{F}_i$ are unique and $\boldsymbol{\Lambda}_M$, the restriction of $\boldsymbol{\Lambda}$ to M, is bijective, it follows that $\tilde{F}_i = \boldsymbol{\Lambda}(F_i)$, $i = 1, \cdots, p$. If now the costs are chosen according to $C_i = \pi_i^{LB/A-1}$, inspection shows that $G_i \cap M = F_i$ and $\tilde{G}_i = \tilde{F}_i$ when $\tilde{G}_i$ is given by (4.17). Hence, $\tilde{G}_i = \boldsymbol{\Lambda}(G_i \cap M)$, which proves the first assertion of the proposition.

We observe finally that

$$P(\boldsymbol{Z}\in\partial G_i | H_i) = P(\boldsymbol{Z}\in\partial G_i \cap M | H_i) = 0,$$

because of (4.18) and (4.19). □

Remark: We say that a sequence of decision rules is asymptotically optimal if it minimizes $\overline{P}_e(\phi, F)$ given F. The asymptotically optimal sequence of decision rules among all based on $\{\boldsymbol{T}_n\}$ (cf. (4.2)) have

$$\tilde{G}_i = \left\{\boldsymbol{y}\in\mathbb{R}^q; i = \arg\min_j \left(\frac{1}{2}\|\boldsymbol{y} - \tilde{\boldsymbol{\mu}}_j\|^2 - \frac{A}{B^2}\log\pi_j\right)\right\}.$$

This follows from (1.3), (4.6), and Proposition 1. The set $\tilde{G}_i$ always agrees with $\tilde{G}_i(\boldsymbol{\pi}, L)$ in Proposition 2 when $\pi_1 = \cdots = \pi_p$. This is also true if $\rho = -\log f$ or $h = h_F$, $L = 1$ and $\boldsymbol{\pi}$ is arbitrary, since in these cases $A = B$. In fact, the latter two sequences of decision rules are asymptotically optimal given F among all decision rules.

V. Robust Detectors

We now turn to the problem when F is not known exactly, but rather belongs to some appropriately chosen class $\mathcal{F}$ of distributions. Our goal is to find a sequence of decision rules that *minimize*

$$\sup_{F\in\mathcal{F}} \overline{P}_e(\phi, F). \tag{5.1}$$

We found in Section IV that $\overline{P}_e(\phi, F)$ is given by (4.13) and (4.17) when ϕ corresponds to a minimum distance/maximum correlation M- or R-detector. A problem is that the regions $\tilde{G}_i(\boldsymbol{\pi}, L)$ depend on B in general, which in turn depends on F. An important special case when $\tilde{G}_i(\boldsymbol{\pi}, L)$ is independent of F is when $\boldsymbol{\pi} = (1/p, \cdots, 1/p)$, which corresponds to a uniform *a priori* information.

A. Equal a priori Probabilities

Assume that $\boldsymbol{\pi} = (1/p, \cdots, 1/p)$. Let F_0 be the nominal distribution, assumed to have a symmetric, strongly unimodal and twice continuously differentiable density f_0 with finite Fisher information. We consider the gross-error neighborhood

$$\mathcal{F}_\epsilon = \{F = (1-\epsilon)F_0 + \epsilon H; H\in\mathcal{M}\} \tag{5.2}$$

introduced by Huber in [7], where $0 < \epsilon \leq 1$ and $\mathcal{M}$ is the set of all symmetric probability measures with finite Fisher information. First we show that $\overline{P}_e$ is equivalent to the efficacy $\mathcal{E}$ as performance criterion.

Lemma 3: Suppose $\boldsymbol{\pi} = (1/p, \cdots, 1/p)$, that $\phi = \{\phi_n\}$ is generated from a sequence of M- or R-statistics according to (2.4) or (2.7). Then, $\overline{P}_e(\phi, F)$ is a decreasing function of the efficacy, $\mathcal{E}$.

Proof: It follows from Propositions 1–2 that $\overline{P}_e = \sum_{i=1}^p \pi_i P(\tilde{\boldsymbol{Z}}\notin\tilde{G}_i|H_i)$, with

$$\tilde{G}_i = \left\{\boldsymbol{y}\in\mathbb{R}^q; i = \arg\min_j \|\boldsymbol{y} - \tilde{\boldsymbol{\mu}}_j\|\right\}. \tag{5.3}$$

Let $\tilde{\boldsymbol{Z}}_0$ be a random vector with distribution $N(0, \boldsymbol{I}_q)$. It follows then from (4.6) that

$$\overline{P}_e = \sum_{i=1}^p \pi_i P(\tilde{\boldsymbol{Z}}_0\notin\sqrt{\mathcal{E}}(\tilde{G}_i - \tilde{\boldsymbol{\mu}}_i)|H_i).$$

Moreover, 0 is an interior point of the convex set $\tilde{G}_i - \tilde{\boldsymbol{\mu}}_i$ for each i, and hence, $a(\tilde{G}_i - \tilde{\boldsymbol{\mu}}_i) \subset b(\tilde{G}_i - \tilde{\boldsymbol{\mu}}_i)$, whenever $0 < a < b$, and this proves the lemma.[4] □

In order to find minimax decision rules, we introduce $\hat{F}$ as the element of $\mathcal{F}_\epsilon$ that minimizes the Fisher information. Its density is given by

$$\hat{f}(x) = \begin{cases} (1-\epsilon)f_0(x), & |x| \leq a, \\ (1-\epsilon)f_0(a)\exp(-b(|x|-a)), & |x| > a, \end{cases} \tag{5.4}$$

for some a and b depending on F_0 and ϵ, chosen so that $\hat{f}$ becomes a p.d.f. and $-\hat{f}'/\hat{f}$ is continuous (cf. [7]). Define

$$\hat{\rho}(x) = -\log \hat{f}(x)$$

and

$$\hat{h}(u) = h_{\hat{F}}(u)$$

(cf. (3.10)). In particular, when $F_0 = \Phi$, the standard normal distribution, $\hat{\rho}$ equals Huber's ρ-function (2.10) and $\hat{h}$ gives Jaeckel's scores defined in Example 2 of Section II. It is shown by Huber ([7], M-statistics) and Jaeckel ([9], R-statistics) that the following saddle-point conditions are satisfied (with $\hat{\psi} = \hat{\rho}'$):

$$\begin{aligned}\sup_\psi \mathcal{E}_M(\psi, \hat{F}) &= \mathcal{E}_M(\hat{\psi}, \hat{F}) \\ &= \inf_{F\in\mathcal{F}_\epsilon} \mathcal{E}_M(\hat{\psi}, F), \end{aligned}\tag{5.5}$$

$$\begin{aligned}\sup_h \mathcal{E}_R(h, \hat{F}) &= \mathcal{E}_R(\hat{h}, \hat{F}) \\ &= \inf_{F\in\mathcal{F}_\epsilon} \mathcal{E}_R(\hat{h}, F), \end{aligned}\tag{5.6}$$

and moreover,

$$\mathcal{E}_M(\hat{\psi}, \hat{F}) = \mathcal{E}_R(\hat{h}, \hat{F}), \tag{5.7}$$

which follows from the definition of A_M, B_M, A_R, and B_R. Let $\hat{\phi}_M$ and $\hat{\phi}_R$ denote sequences of M-detectors and R-detectors respectively, either of minimum distance type (2.4) or maximum correlation type (2.7) and with $\rho = \hat{\rho}$ or $h = \hat{h}$. (The choice of L is arbitrary since $\boldsymbol{\pi} = (1/p, \cdots, 1/p)$.) We then have the following minimax result.

Theorem 1: Let $\boldsymbol{\pi} = (1/p, \cdots, 1/p)$. Then the sequences of decision rules $\hat{\phi}_M$ and $\hat{\phi}_R$ defined above satisfy the relations

$$\begin{aligned}\inf_\phi \sup_{F\in\mathcal{F}_\epsilon} \overline{P}_e(\phi, F) &= \sup_{F\in\mathcal{F}_\epsilon} \overline{P}_e(\hat{\phi}_M, F) \\ &= \sup_{F\in\mathcal{F}_\epsilon} \overline{P}_e(\hat{\phi}_R, F), \end{aligned}\tag{5.8}$$

where the uncertainty class $\mathcal{F}_\epsilon$ is defined in (5.2) and the infimum is taken over all sequences of decision rules.

Proof: It follows from Lemma 3 and (5.5)–(5.7) that

$$\begin{aligned}\sup_{F\in\mathcal{F}_\epsilon} \overline{P}_e(\hat{\phi}_M, F) &= \overline{P}_e(\hat{\phi}_M, \hat{F}) = \overline{P}_e(\hat{\phi}_R, \hat{F}) \\ &= \sup_{F\in\mathcal{F}_\epsilon} \overline{P}_e(\hat{\phi}_R, F). \end{aligned}\tag{5.9}$$

Since $\hat{\rho}(x) = -\log \hat{f}(x)$, the minimum distance M-detector coincides with the optimal Bayes decision rule (1.3) at $\hat{F}$. Hence, $\hat{\phi}_M$ minimizes $P_e(\phi_n, F)$ for each n and

$$\inf_\phi \overline{P}_e(\phi, \hat{F}) = \overline{P}_e(\hat{\phi}_M, \hat{F}). \tag{5.10}$$

(5.9) and (5.10) together prove the theorem. □

[4] More generally, let $\tilde{\boldsymbol{\mu}}_i(\mathcal{E})$ and $\tilde{\Sigma}(\mathcal{E})$ denote mean vector and covariance matrix of $\tilde{\boldsymbol{Z}}$ as a function of $\mathcal{E}$ and suppose also that the transformed decision regions $\tilde{G}_i(\mathcal{E})$ depend on $\mathcal{E}$. Then a sufficient condition for Lemma 3 to remain valid is that $P(\tilde{\boldsymbol{Z}}_0\in\Sigma(\mathcal{E})^{-1/2}(\tilde{G}_i(\mathcal{E}) - \tilde{\boldsymbol{\mu}}_i(\mathcal{E})))$ is an increasing function of $\mathcal{E}$, for $i = 1, \cdots, p$.

B. General Case

In this subsection, we impose no restrictions on $\boldsymbol{\pi}$. As mentioned previously, the regions $\hat{G}_i(\boldsymbol{\pi}, L)$ then depend on B, which in turn depends on F. We will, therefore, restrict the uncertainty class to

$$\mathcal{F}'_\epsilon = \{F = (1-\epsilon)F_0 + \epsilon H; H \in \mathcal{M}'\}, \tag{5.11}$$

where $\mathcal{M}'$ is the subclass of $\mathcal{M}$ (cf. (5.2)) of probability measures with $H([-a, a]) = 0$, $a = a(\epsilon, \mathcal{F}_0)$ being the constant in (5.4). $\mathcal{F}'_\epsilon$ is thus a subclass of $\mathcal{F}_\epsilon$ where contamination is restricted to the tails. The class $\mathcal{F}'_\epsilon$ is not a genuine neighborhood of F_0 since F_0 itself does not belong to $\mathcal{F}'_\epsilon$. On the other hand, $\hat{F}$—the least favorable distribution of $\mathcal{F}_\epsilon$—belongs to $\mathcal{F}'_\epsilon$.

Lemma 4: Let $\boldsymbol{\pi}$ be arbitrary and let $\hat{\phi}_M$ and $\hat{\phi}_R$ be defined as in Section V-A with L arbitrary. Then, $\overline{P}_e(\hat{\phi}_M, F)$ and $\overline{P}_e(\hat{\phi}_R, F)$ are identical and constant over $\mathcal{F}'_\epsilon$.

Proof: According to Propositions 1–2, $\overline{P}_e(\hat{\phi}_M, F)$ depends only on $A_M(\hat{\psi}, F)$ and $B_M(\hat{\psi}, F)$ and $\overline{P}_e(\hat{\phi}_R, F)$ only on $A_R(\hat{h})$ and $B_R(\hat{h}, F)$. Since $\hat{\psi} = -\hat{f}'/\hat{f}$ is skew-symmetric, a change of variables $u = 2\hat{F}(x) - 1, x > 0$, shows that

$$A_M(\hat{\psi}, \hat{F}) = A_R(\hat{h}).$$

Moreover, since $\hat{\psi}(x)$ is constant when $|x| > a(\epsilon, F_0)$,

$$A_M(\hat{\psi}, F) = A_M(\hat{\psi}, \hat{F}), \qquad \forall F \in \mathcal{F}'_\epsilon$$

follows. Similarly,

$$B_M(\hat{\psi}, F) = B_M(\hat{\psi}, \hat{F}), \qquad \forall F \in \mathcal{F}'_\epsilon.$$

A change of variables plus integration by parts shows that

$$B_R(\hat{h}, F) = 2\int_0^\infty \frac{d}{dx}(\hat{h}(2F(x) - 1))f(x)\,dx$$

where $F \in \mathcal{F}'_\epsilon$. In particular, putting $F = \hat{F}$ gives

$$B_R(\hat{h}, \hat{F}) = 2\int_0^\infty \hat{\psi}'(x)\hat{f}(x)\,dx = B_M(\hat{\psi}, \hat{F}).$$

since $\hat{\psi}'$ is symmetric. The facts that $F(x) = \hat{F}(x)$ when $|x| \le a(\epsilon, F_0)$ and $\hat{h}'(2F(x) - 1) = 0$ when $x > a(\epsilon, F_0)$ finally imply

$$B_R(\hat{h}, F) = B_R(\hat{h}, \hat{F}), \qquad \forall F \in \mathcal{F}'_\epsilon.$$

□

We may now state the minimax robustness result for general $\boldsymbol{\pi}$. The difference from Theorem 1 is only that the uncertainty class $\mathcal{F}'_\epsilon$ is narrower.

Theorem 2: Let $\boldsymbol{\pi}$ be arbitrary and let $\hat{\phi}_M$ and $\hat{\phi}_R$ be defined as in Lemma 3 with $L = 1$. Then,

$$\begin{aligned}\inf_\phi \sup_{F \in \mathcal{F}'_\epsilon} \overline{P}_e(\phi, F) &= \sup_{F \in F'_\epsilon} \overline{P}_e(\hat{\phi}_M, F) \\ &= \sup_{F \in \mathcal{F}'_\epsilon} \overline{P}_e(\hat{\phi}_R, F),\end{aligned} \tag{5.12}$$

with the uncertainty class $\mathcal{F}'_\epsilon$ given by (5.11) and the infimum taken over all sequences of decision rules.

Proof: According to Lemma 4,

$$\begin{aligned}\sup_{F \in \mathcal{F}'_\epsilon} \overline{P}_e(\hat{\phi}_R, F) &= \sup_{F \in \mathcal{F}'_\epsilon} \overline{P}_e(\hat{\phi}_M, F) \\ &= \overline{P}_e(\hat{\phi}_M, \hat{F}).\end{aligned} \tag{5.13}$$

Since $\hat{\phi}_M$ yields the optimum decision rule for each n (5.10) holds, which together with (5.13) proves the theorem. □

VI. Two-Dimensional Signals

A. Problem Statement

Let us generalize our hypothesis testing problem to the case when the observed data are two-dimensional vectors:

$$H_i\colon \boldsymbol{Y}_k = \alpha_n \boldsymbol{c}_{ik} + \boldsymbol{e}_k, \quad k = 1, \cdots, n, \quad i = 1, \cdots, p, \tag{6.1}$$

where $\boldsymbol{Y}_k = (Y_{k1}, Y_{k2})$, $\boldsymbol{c}_{ik} = (c_{ik1}, c_{ik2})$, and $\boldsymbol{e}_k = (e_{k1}, e_{k2})$. We also assume that the errors $\boldsymbol{e}_k$ are i.i.d. random vectors with distribution $F(d\boldsymbol{x}) = f(\|\boldsymbol{x}\|)\,d\boldsymbol{x}$, which is radially symmetric, so that $2\pi r f(r)$ becomes the p.d.f. of the noise envelope $\|\boldsymbol{e}\|$. As before, the norming constant α_n approaches zero as $n \to \infty$ according to (1.4). Note that the model (6.1) reduces to (1.1) if we assume that e_{k1} and e_{k2} are independent. However, this is only true when the marginal distributions of $\boldsymbol{e}$ are Gaussian. In many radar and communication problems, (6.1) corresponds to detection of bandpass signals in bandpass noise.[5]

We restrict ourselves to generalize the M-detector from the one-dimensional case. Much of the analysis is analogous, so we will be rather brief. Let $\underline{\boldsymbol{x}} = (\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n)$ and $\underline{\boldsymbol{y}} = (\boldsymbol{y}_1, \cdots, \boldsymbol{y}_n)$ be given $2 \times n$-matrices and $\rho\colon \mathbb{R} \to \mathbb{R}$ a given function. Define the M-distance

$$D_n(\underline{\boldsymbol{x}} - \underline{\boldsymbol{y}}) = \sum_{k=1}^n \rho(\|\boldsymbol{x}_k - \boldsymbol{y}_k\|) \tag{6.2}$$

and the corresponding minimum distance detector

$$\phi_n(\underline{\boldsymbol{Y}}_n) = \arg\min_i (D_n(\underline{\boldsymbol{Y}}_n - \alpha_n \underline{\boldsymbol{c}}_{in}) - L \log \pi_i), \tag{6.3}$$

where $\underline{\boldsymbol{Y}}_n = (\boldsymbol{Y}_1, \cdots, \boldsymbol{Y}_n)$ and $\underline{\boldsymbol{c}}_{in} = (\boldsymbol{c}_{i1}, \cdots, \boldsymbol{c}_{in})$. Introducing $\boldsymbol{T}_n^D(\underline{\boldsymbol{Y}}_n)$, a vector with components $T_{ni}^D(\underline{\boldsymbol{Y}}_n) = D_n(\underline{\boldsymbol{Y}}_n) - D_n(\underline{\boldsymbol{Y}}_n - \alpha_n \underline{\boldsymbol{c}}_{in})$, $i = 1, \cdots, p$, we may rewrite (6.3) as

$$\phi_n(\underline{\boldsymbol{Y}}_n) = \arg\max_i (T_{ni}^D(\underline{\boldsymbol{Y}}_n) + L \log \pi_i). \tag{6.4}$$

The maximum correlation detector takes the form

$$\phi_n(\underline{\boldsymbol{Y}}_n) = \arg\max_i (T_{ni}^C(\underline{\boldsymbol{Y}}_n) + L \log \pi_i), \tag{6.5}$$

where

$$T_{ni}^C(\underline{\boldsymbol{Y}}_n) = \alpha_n \sum_{k=1}^n \boldsymbol{c}_{ik} \cdot \boldsymbol{w}\left(\boldsymbol{Y}_k - \frac{1}{2}\alpha_n \boldsymbol{c}_{ik}\right), \tag{6.6}$$

[5]The set-up may easily be generalized to observations in any dimension. For notational simplicity, we will confine ourselves to two dimensions. This case is also most interesting for applications.

are the components of the p-dimensional vector $\boldsymbol{T}_n^C(\underline{\boldsymbol{Y}}_n)$, $\cdot$ denotes the dot product and $\boldsymbol{w}$: $\mathbb{R}^2 \to \mathbb{R}^2$ is given by

$$\boldsymbol{w}(\boldsymbol{x}) = \frac{\psi(\|\boldsymbol{x}\|)}{\|\boldsymbol{x}\|}\boldsymbol{x}, \tag{6.7}$$

with $\psi = \rho'$. We may interpret this detector as one that chooses $\underline{\boldsymbol{c}}_{in}$ in order to maximize the correlation with a transformation of $\underline{\boldsymbol{Y}}_n - \frac{1}{2}\alpha_n\underline{\boldsymbol{c}}_{in}$. The remark following Lemma 1 holds even in two dimensions: If the L^2-norms of $\alpha_n\boldsymbol{c}_{1n}, \cdots, \alpha_n\boldsymbol{c}_{pn}$ are the same, we may replace $\boldsymbol{w}(\underline{\boldsymbol{Y}}_n - \frac{1}{2}\alpha_n\underline{\boldsymbol{c}}_{in})$ by $\boldsymbol{w}(\underline{\boldsymbol{Y}}_n)$ in (6.6).

B. Asymptotics

As in Section III, we first consider the asymptotic behavior of the test statistics $\boldsymbol{T}_n^D$ and $\boldsymbol{T}_n^C$. First, we impose a set of regularity conditions: Suppose that the limits

$$\tilde{\sigma}_{ij} = \lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}\boldsymbol{c}_{ik}\cdot\boldsymbol{c}_{jk} \tag{6.8}$$

exist and put $\tilde{\boldsymbol{\Sigma}} = (\tilde{\sigma}_{ij})_{i,j=1}^{p}$, $\tilde{\boldsymbol{\sigma}}_i = (\tilde{\sigma}_{i1}, \cdots, \tilde{\sigma}_{ip})$, and $\tilde{\boldsymbol{\sigma}} = (\tilde{\sigma}_{11}, \cdots, \tilde{\sigma}_{pp})$. Assume also that ρ is absolutely continuous with $\rho' = \psi$,

$$\begin{aligned} 0 < \tilde{A}(\psi, F) &= \int_0^\infty \pi r f(r)\psi(r)^2\, dr \\ &= \frac{1}{2}E\|\boldsymbol{w}(\boldsymbol{e})\|^2 = E(w_1(\boldsymbol{e}))^2 < \infty \end{aligned} \tag{6.9}$$

(with $\boldsymbol{w} = (w_1, w_2)$), and

$$\lim_{\boldsymbol{h}\to 0} E\|\boldsymbol{w}(\boldsymbol{e}+\boldsymbol{h}) - \boldsymbol{w}(\boldsymbol{e})\|^2 = 0. \tag{6.10}$$

Further, there exists a $\tilde{B}(\psi, F)$, $0 < \tilde{B}(\psi, F) < \infty$, such that

$$E\boldsymbol{w}(\boldsymbol{e}+\boldsymbol{h}) = \tilde{B}(\psi, F)\boldsymbol{h} + o(\|\boldsymbol{h}\|) \tag{6.11}$$

holds. Finally,

$$\begin{aligned} &E(\rho(\|\boldsymbol{e}+\boldsymbol{h}+\boldsymbol{q}\|) - \rho(\|\boldsymbol{e}+\boldsymbol{h}-\boldsymbol{q}\|) \\ &\qquad - 2\boldsymbol{q}\cdot\boldsymbol{w}(\boldsymbol{e}+\boldsymbol{h}))^2 = o(\|\boldsymbol{h}\|^2 + \|\boldsymbol{q}\|^2) \end{aligned} \tag{6.12}$$

and

$$\begin{aligned} &E(\rho(\|\boldsymbol{e}+\boldsymbol{h}+\boldsymbol{q}\|) - \rho(\|\boldsymbol{e}+\boldsymbol{h}-\boldsymbol{q}\|) \\ &\qquad - 2\boldsymbol{q}\cdot\boldsymbol{w}(\boldsymbol{e}+\boldsymbol{h})) = o(\|\boldsymbol{h}\|^2 + \|\boldsymbol{q}\|^2) \end{aligned} \tag{6.13}$$

hold as $\|h\|$ and $\|q\| \to 0$. What is the explicit form of $\tilde{B}(\psi, F)$? First we note that the derivative

$$D\boldsymbol{w}(\boldsymbol{x}) = \frac{\psi(\|\boldsymbol{x}\|)}{\|\boldsymbol{x}\|}\boldsymbol{I}_2 + \frac{\psi'(\|\boldsymbol{x}\|)\|\boldsymbol{x}\| - \psi(\|\boldsymbol{x}\|)}{\|\boldsymbol{x}\|^3}\boldsymbol{x}'\boldsymbol{x},$$

if ψ is differentiable. If also ψ' is bounded, we can differentiate under the integral sign in (6.11) to obtain

$$\begin{aligned} [D_{\boldsymbol{h}}E\boldsymbol{w}(\boldsymbol{e}+\boldsymbol{h})]_{\boldsymbol{h}=0} &= ED\boldsymbol{w}(\boldsymbol{e}) \\ &= \frac{1}{2}\left(E\frac{\psi(\|\boldsymbol{e}\|)}{\|\boldsymbol{e}\|} + E\psi'(\|\boldsymbol{e}\|)\right)\boldsymbol{I}_2 \\ &= \tilde{B}(\psi, F)\boldsymbol{I}_2. \end{aligned}$$

We now proceed and prove asymptotic normality for the test statistics $\boldsymbol{T}_n^D$ and $\boldsymbol{T}_n^C$.

Lemma 5: Let $\boldsymbol{T}_n^C$ and $\boldsymbol{T}_n^D$ be defined as in Section VI-A and suppose that (6.8)–(6.11) are valid. Then,

$$\boldsymbol{T}_n^C \xrightarrow{d} N(\boldsymbol{\mu}_i, K^2\tilde{A}(\psi, F)\tilde{\boldsymbol{\Sigma}}) \tag{6.14}$$

under H_i, where

$$\boldsymbol{\mu}_i = K^2\tilde{B}(\psi, F)\left(\tilde{\boldsymbol{\sigma}}_i - \frac{1}{2}\tilde{\boldsymbol{\sigma}}\right). \tag{6.15}$$

If also (6.12)–(6.13) are satisfied, (6.14) holds with $\boldsymbol{T}_n^D$ in place of $\boldsymbol{T}_n^C$.

Proof: As in the proof of Lemma 1, we make the decompositions $\boldsymbol{T}_n^C = \boldsymbol{T}_n^{(1)} + \boldsymbol{T}_n^{(2)} + \boldsymbol{T}_n^{(3)}$ and $\boldsymbol{T}_n^D = \boldsymbol{T}_n^{(1)} + \boldsymbol{T}_n^{(2)} + \boldsymbol{T}_n^{(3)} + \boldsymbol{T}_n^{(4)}$, where the jth component of $\boldsymbol{T}_n^{(1)}, \cdots, \boldsymbol{T}_n^{(4)}$ under H_i is given by

$$T_{nj}^{(1)} = \alpha_n\sum_{k=1}^{n}\boldsymbol{c}_{jk}\cdot\boldsymbol{w}(\boldsymbol{e}_k),$$

$$\begin{aligned} T_{nj}^{(2)} = \alpha_n\sum_{k=1}^{n}\boldsymbol{c}_{jk}\cdot\Bigg(&\boldsymbol{w}\left(\boldsymbol{e}_k + \alpha_n\left(\boldsymbol{c}_{ik} - \frac{1}{2}\boldsymbol{c}_{jk}\right)\right) \\ &-\boldsymbol{w}(\boldsymbol{e}_k) - E\boldsymbol{w}\left(\boldsymbol{e}_k + \alpha_n\left(\boldsymbol{c}_{ik} - \frac{1}{2}\boldsymbol{c}_{jk}\right)\right)\Bigg), \end{aligned}$$

$$T_{nj}^{(3)} = \alpha_n\sum_{k=1}^{n}\boldsymbol{c}_{jk}\cdot E\boldsymbol{w}\left(\boldsymbol{e}_k + \alpha_n\left(\boldsymbol{c}_{ik} - \frac{1}{2}\boldsymbol{c}_{jk}\right)\right),$$

and finally

$$\begin{aligned} T_{nj}^{(4)} = \sum_{k=1}^{n}\Bigg(&(\rho(\|\boldsymbol{e}_k + \alpha_n\boldsymbol{c}_{ik}\|) \\ &- \rho(\|\boldsymbol{e}_k + \alpha_n(\boldsymbol{c}_{ik} - \boldsymbol{c}_{jk})\|) \\ &-\boldsymbol{c}_{jk}\cdot\boldsymbol{w}\left(\boldsymbol{e}_k + \alpha_n\left(\boldsymbol{c}_{ik} - \frac{1}{2}\boldsymbol{c}_{jk}\right)\right)\Bigg). \end{aligned}$$

By similar arguments as in Lemma 1, it may be shown that

$$\boldsymbol{T}_n^{(1)} \xrightarrow{d} N(0, K^2\tilde{A}(\psi, F)\tilde{\boldsymbol{\Sigma}}),$$

$$\boldsymbol{T}_n^{(2)} \xrightarrow{d} 0,$$

$$\boldsymbol{T}_n^{(3)} \to K^2\tilde{B}(\psi, F)\left(\tilde{\boldsymbol{\sigma}}_i - \frac{1}{2}\tilde{\boldsymbol{\sigma}}\right),$$

and

$$\boldsymbol{T}_n^{(4)} \xrightarrow{d} 0,$$

under H_i, from which the lemma follows. □

C. Robust Detectors

Given a sequence of decision rules $\phi = \{\phi_n\}_{n=1}^{\infty}$, let $P(\phi_n, F)$ symbolize the error probability as in (1.2) and $\overline{P}_e(\phi, F)$ the lim sup of the error probabilities as in (4.1). The properties of $\overline{P}_e$ in Section IV carry over directly, since the asymptotic behavior of $\boldsymbol{T}_n^C$ and $\boldsymbol{T}_n^D$ is analogous in the one-and two-dimensional cases according to Lemmas 1 and 5. The only difference is that the quantities A, B, $\boldsymbol{\Sigma}$, etc. are replaced by their two-dimensional analogies $\tilde{A}$, $\tilde{B}$, and $\tilde{\boldsymbol{\Sigma}}$. We also introduce the two-dimensional efficacy

$$\tilde{\mathcal{E}}(\psi, F) = \tilde{B}(\psi, F)^2/\tilde{A}(\psi, F). \tag{6.16}$$

The distribution of the transformed limiting random vector $\tilde{\boldsymbol{Z}}$ is then given by (cf. (4.6))

$$\tilde{\boldsymbol{Z}} = \boldsymbol{\Lambda}(\boldsymbol{Z}) \sim N(\tilde{\boldsymbol{\mu}}_i, \tilde{\mathcal{E}}^{-1}\boldsymbol{I}_q). \tag{6.17}$$

under H_i, with $\tilde{\boldsymbol{\mu}}_i$ given by (4.7) (replacing $\boldsymbol{\Sigma}$ by $\tilde{\boldsymbol{\Sigma}}$). Since Propositions 1–2 carry over in particular, we see that $\overline{P}_e(\phi, F)$ only depends on $\tilde{A}(\psi, F)$ and $\tilde{B}(\psi, F)$ for general *a priori* probability vectors $\boldsymbol{\pi}$. In the special case when $\boldsymbol{\pi} = (1/p, \cdots, 1/p)$, $\overline{P}_e(\phi, F)$ only depends on $\tilde{\mathcal{E}}(\psi, F)$.

Let us define the ϵ-neighborhood

$$\mathcal{F}_\epsilon'' = \{F = (1-\epsilon)F_0 + \epsilon H : H \in \mathcal{M}''\}$$

as uncertainty class, where F_0 and H are probability measures on $\mathbb{R}^2$, with $F_0(d\boldsymbol{x}) = f_0(\|\boldsymbol{x}\|)\,d\boldsymbol{x}$ radially symmetric and absolutely continuous and with each $H \in \mathcal{M}''$ radially symmetric and supported outside a circle of radius a, where a is a constant to be chosen below in (6.19). In this way, only tail contamination is allowed. This implies in particular that $F_0 \notin \mathcal{F}_\epsilon''$, so $\mathcal{F}_\epsilon''$ is not a genuine neighborhood of F_0 (as in Section V-B). The least favorable distribution in $\mathcal{F}_\epsilon''$ with minimal Fisher information does not have exponential tails ($f(r) = Ce^{-kr}$) as in the one-dimensional case, and the optimal detector for this distribution would have a rather untractable form (cf. [8, pp. 229–230]). We rather propose to start out from $\hat{F}(d\boldsymbol{x}) = \hat{f}(\|\boldsymbol{x}\|)\,d\boldsymbol{x}$, which has exponentially decaying tails according to

$$\hat{f}(r) = \begin{cases} (1-\epsilon)f_0(r), & 0 \le r \le a, \\ (1-\epsilon)f_0(a)e^{-b(r-a)}, & r > a. \end{cases}$$

This is done by Kassam in [10] when $p = 2$. The constants $a = a(\epsilon, f_0)$ and $b = b(\epsilon, f_0)$ are chosen in order to make $\hat{f}$ a p.d.f. and $\hat{f}'/\hat{f}$ continuous at a. Then, define

$$\hat{\rho}(r) = -\log \hat{f}(r), \tag{6.18}$$

$$\begin{aligned} \hat{\psi}(r) = \hat{\rho}'(r) &= -\frac{\hat{f}'(r)}{\hat{f}(r)} \\ &= \begin{cases} -f_0'(r)/f_0(r), & 0 \le r \le a, \\ -f_0'(a)/f_0(a), & r > a, \end{cases} \end{aligned} \tag{6.19}$$

and let $\hat{\phi}^D$ and $\hat{\phi}^C$ be the sequences of minimum distance/maximum correlation decision rules with $L = 1$, $\rho = \hat{\rho}$, and $\psi = \hat{\psi}$.

It is easy to see that $\tilde{A}(\hat{\psi}, F) = \frac{1}{2}E_F\hat{\psi}(\|\boldsymbol{e}\|)^2$ is constant over $\mathcal{F}_\epsilon''$. (Here, E_F denotes expectation when F is the underlying distribution.) However, as F varies over $\mathcal{F}_\epsilon''$,

$$\tilde{B}(\hat{\psi}, F) = (1-\epsilon)\tilde{B}(\hat{\psi}, F_0) + \frac{1}{2}\epsilon E_H\left(\frac{\hat{\psi}(\|\boldsymbol{e}\|)}{\|\boldsymbol{e}\|}\right).$$

varies between $(1-\epsilon)\tilde{B}(\hat{\psi}, F_0)$ and $(1-\epsilon)\tilde{B}(\hat{\psi}, F_0) + \epsilon\hat{\psi}(a)/2a$, so $\tilde{B}$ is *not* constant over the uncertainty class. We can, therefore, not proceed along the lines of Section V-B to prove that $\hat{\phi}^D$ and $\hat{\phi}^C$ are minimax decision rules in terms of asymptotic error probability. Not even when the *a priori* probabilities are equal and $\overline{P}_e$ depends monotonically on the efficacy $\tilde{\mathcal{E}}$ do $\hat{\phi}^D$ and $\hat{\phi}^C$ constitute a minimax detector, since a saddle-point condition of the kind (5.5) is not valid.

Anyhow, $\hat{\phi}^D$ and $\hat{\phi}^C$ are intuitively appealing as robust detectors because of the limiting form of $\hat{\psi}$. Moreover, when ϵ is small, $\tilde{B}$ is *nearly* constant over $\mathcal{F}_\epsilon''$, so that $\hat{\phi}^D$ and $\hat{\phi}^C$ are nearly minimax.

VII. Monte Carlo Simulations

In this section, we compare some Monte Carlo estimates of the error probabilities $P_e(\phi_n, F)$ with the asymptotic formulas derived in Section IV. We performed 30 000 simulations for each combination of signal configuration, noise distribution and detector. The following signal configuration/noise distribution/detectors were used.

1) Signal Configurations: One-dimensional signals, $p = 2$, $c_{1k} = 0$,

$$c_{2k} = \frac{1}{\sqrt{1+\mu^2}}(1-(-1)^k\mu), \quad \mu = 0.5. \tag{S1}$$

One-dimensional signals, $p = 4$,

$$c_{ik} = \sqrt{2}\,\sin\left(\frac{2\pi k}{n} + \frac{i\pi}{2}\right). \tag{S2}$$

Two-dimensional signals, $p = 2$, $\boldsymbol{c}_{1k} = (0, 0)$,

$$\boldsymbol{c}_{2k} = \left(\cos\frac{2\pi k}{n}, \sin\frac{2\pi k}{n}\right). \tag{S3}$$

Two-dimensional signals, $p = 4$,

$$\boldsymbol{c}_{ik} = \left(\cos\left(\frac{2\pi k}{n} + \frac{i\pi}{2}\right), \sin\left(\frac{2\pi k}{n} + \frac{i\pi}{2}\right)\right). \tag{S4}$$

2) Noise Distributions: The simulated noise distributions are a mixture of normal distributions ($CN(\epsilon, \sigma^2)$-*C*ontaminated *N*ormal) with density functions

$$f(x) = (1-\epsilon)\varphi(x) + \epsilon\varphi(x/\sigma) \tag{7.1}$$

in one dimension and

$$f(\boldsymbol{x}) = (1-\epsilon)\varphi(x_1)\varphi(x_2) + \epsilon\varphi(x_1/\sigma)\varphi(x_2/\sigma) \tag{7.2}$$

in two dimensions, φ being the p.d.f. of the standard normal distribution.

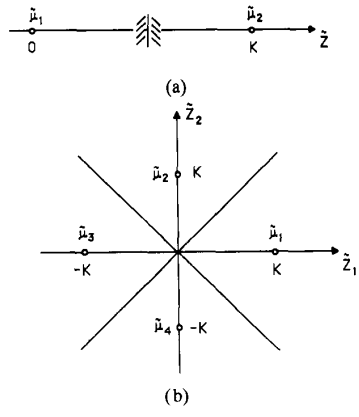

Fig. 4. Mean vectors $\tilde{\mu}_i$ and decision regions $\tilde{G}_i$ for signal configurations. (a) (S1) and (S3). (b) (S2) and (S4).

3) Detectors: The detectors considered are grouped into four categories: MD, MC, RD, and RC. The first letter indicates distance type, the second stands for either minimum *D*istance or maximum *C*orrelation. For MD detectors the Huber functions ρ_a (cf. (2.10)) have been used and for MC detectors $\psi_a = \rho_a'$. For $a = 0$ we write $|x|$ for ρ and sgn for ψ and for $a = \infty$ we let x^2 denote ρ. The score functions utilized for *R*-distances have either been the van der Waerden normal scores ($h = \Phi^{-1}$), truncated normal scores ($h = h_b$) or Wilcoxon scores ($h = h_w$). (See Example 2.)

Next, we describe the calculation of asymptotic error probabilities. The error probability $P_e(\phi_n, F)$ may be approximated by $\overline{P}_e(\phi, F)$, which is given by (4.13), with decision regions $\tilde{G}_i(\boldsymbol{\pi}, L)$ according to (4.17) and the distribution of the limit random vector $\tilde{\boldsymbol{Z}}$ under each hypothesis given by (4.6) in the one-dimensional case and by (6.17) in two dimensions. We suppose equal *a priori* probabilities, so we may put $L = 0$ in the detector (2.2), (2.5), (6.3), or (6.5). Assuming that the signals repeat themselves in periods equal to the length of the simulated signals, it is easy to see (cf. (*C*)) that

$$\boldsymbol{\Sigma} = \frac{1}{n}\boldsymbol{C}_n'\boldsymbol{C}_n \tag{7.3}$$

in the one-dimensional case with n equal to the length of the simulated signals. A similar truncation of (6.8) is possible for two-dimensional signals when $\tilde{\boldsymbol{\Sigma}}$ is to be calculated. Hence, we obtain

$$\overline{P}_e(\phi, F) = \frac{1}{p}\sum_{i=1}^{p} P(\tilde{\boldsymbol{Z}} \notin \tilde{G}_i(\boldsymbol{\pi}, 0)), \tag{7.4}$$

with $\boldsymbol{\Sigma}$ (or $\tilde{\boldsymbol{\Sigma}}$) as given above. We illustrate the calculation of $\overline{P}_e$ for each of the studied signal configurations (S1)–(S4) next and in Fig. 4.

(S1) In this case, $\boldsymbol{\Sigma} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$, $q = 1$, $\tilde{\mu}_1 = K$, and $\tilde{\mu}_2 = 0$, from which it follows that

$$\overline{P}_e(\phi, F) = Q(K\sqrt{\mathcal{E}}/2), \tag{7.5}$$

where $Q(x) = 1 - \Phi(x)$ the upper tail probability of a random variable with standard normal distribution and $\mathcal{E}$ is the efficacy (cf. (4.8)).

(S2) For this signal configuration,

$$\boldsymbol{\Sigma} = \begin{pmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 \\ -1 & 0 & 1 & 0 \\ 0 & -1 & 0 & 1 \end{pmatrix} \tag{7.6}$$

and $q = 2$. The two nonzero eigenvalues of $\boldsymbol{\Sigma}$ are equal to two, so the choice of $\boldsymbol{\Lambda}$ is not unique as mentioned in Section IV. One choice gives $\tilde{\mu}_1 = K(1, 0)$, $\tilde{\mu}_2 = K(0, 1)$, $\tilde{\mu}_3 = K(-1, 0)$, and $\tilde{\mu}_4 = K(0, -1)$; so that

$$\overline{P}_e(\phi, F) = 1 - (1 - Q(K\sqrt{\mathcal{E}/2}))^2 \tag{7.7}$$

follows.

(S3) In this case $\tilde{\boldsymbol{\Sigma}} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$, $q = 1$, $\tilde{\mu}_1 = K$, and $\tilde{\mu}_2 = 0$, from which

$$\overline{P}_e(\phi, F) = Q(K\sqrt{\tilde{\mathcal{E}}}/2) \tag{7.8}$$

follows, with $\tilde{\mathcal{E}}$ equal to the efficacy in two dimensions (cf. (6.16)).

(S4) Here, $\tilde{\boldsymbol{\Sigma}}$ is given by (7.6) and $\tilde{\mu}_1 - \tilde{\mu}_4$ are the same as for (S2). This gives

$$\overline{P}_e(\phi, F) = 1 - (1 - Q(K\sqrt{\tilde{\mathcal{E}}/2}))^2. \tag{7.9}$$

The results of the Monte Carlo study are shown in Tables I–VI and Fig. 5. The simulated error probabilities are given in % with 95% confidence intervals. Table I gives an example that the enhanced performance with robust detectors in impulsive noise is more evident for two-dimensional signals than for one-dimensional signals. Table II gives an example of the effect of increasing the number of possible signals p from two to four. For the signal configurations (S1) and (S2), the error probability is almost doubled when the nearest neighbor distance (between different $\tilde{\boldsymbol{\mu}}_i$) is the same. Table III compares the performance of minimum distance and maximum correlation detectors for various signal configurations. The minimum distance detectors seem to be slightly better. This is clear for MD $|x|$ compared to MC sgn, but for the other detector pairs the difference is not so obvious. Table IV and V compare the performance of the detectors in various types of noise. The MD $\rho_{1.5}$ detector has the best performance, closely followed by MC $\psi_{1.5}$. The Euclidean distance MD x^2 has only a little better performance for Gaussian noise but catastrophic characteristics for impulsive noise. The rank-based detectors behave relatively well, especially the Wilcoxon detector, which has better performance than the other rank-based detectors in impulsive noise. The MD $|x|$ detector has relatively low error probabilities for impulsive noise and finally MC sgn has relatively bad performance for all noise distributions.

Let us now discuss the agreement between the simulated finite sample error probabilities and the asymptotic error probabilities. About half of the 106 asymptotic error probabilities lie outside the 95% confidence intervals, so there is certainly

TABLE I
ERROR PROBABILITIES (IN %) FOR ONE- AND TWO-DIMENSIONAL SIGNALS

Detector		Two signals 1 dim (S1) $K = \sqrt{15}$	Two signals 2 dim (S3) $K = \sqrt{15}$	Four signals 1 dim (S2) $K = \sqrt{20}$	Four signals 2 dim (S4) $K = \sqrt{20}$
MD x^2	Theor	9.83	9.83	11.06	11.06
	Sim	9.52 ± 0.33	9.63 ± 0.33	10.81 ± 0.35	11.03 ± 0.35
MD $\rho_{1.5}$	Theor	4.78	4.20	4.08	3.41
	Sim	5.06 ± 0.25	4.14 ± 0.23	4.24 ± 0.23	3.80 ± 0.22
MD $!x!$	Theor	6.85	4.93	6.74	4.27
	Sim	5.80 ± 0.26	4.53 ± 0.24	4.77 ± 0.24	3.98 ± 0.22

Positioning of $\tilde{\mu}_1, \cdots, \tilde{\mu}_p$ is the same for (S1) and (S3) and also the same for (S2) and (S4). $F = CN(0.25, 3^2)$ and $n = 30$.

TABLE II
ERROR PROBABILITIES (IN %) FOR ONE-DIMENSIONAL SIGNALS WHEN $p = 2$ AND $p = 4$, RESPECTIVELY

Detector		Two signals (S1) $K = \sqrt{20}$	Four signals (S2) $K = \sqrt{10}$
MD x^2	Theor	9.83	18.70
	Sim	9.52 ± 0.33	17.80 ± 0.43
MD $\rho_{1.5}$	Theor	4.78	9.33
	Sim	5.06 ± 0.25	9.73 ± 0.34
RD $h_{1.5}$	Theor	5.20	10.14
	Sim	5.67 ± 0.26	11.47 ± 0.36

The amplitude K is $\sqrt{2}$ times larger for (S1), so that $\min_{i \neq j} \|\tilde{\mu}_i - \tilde{\mu}_j\|$ is the same for both signal configurations. Hence, $P_{e,(S2)} = 1 - (1 - P_{e,(S1)})^2 \approx 2P_{e,(S1)}$. $F = CN(0.25, 3^2)$ and $n = 30$.

TABLE III
ERROR PROBABILITIES (IN %) FOR MINIMUM DISTANCE AND MAXIMUM CORRELATION DETECTORS THAT THEORETICALLY HAVE THE SAME ASYMPTOTIC ERROR PROBABILITY.

Detector		(S1) $K = \sqrt{20}$	(S2) $K = \sqrt{15}$	(S3) $K = \sqrt{20}$	(S4) $K = \sqrt{15}$
MD $\rho_{1.5}$	Theor	1.90	1.10	1.86	1.07
	Sim	1.92 ± 0.16	0.94 ± 0.11	2.04 ± 0.16	1.15 ± 0.12
MC $\psi_{1.5}$	Theor	1.90	1.10	1.86	1.07
	Sim	1.95 ± 0.16	1.12 ± 0.12	2.46 ± 0.18	1.15 ± 0.15
MD $!x!$	Theor	4.22	3.44	2.77	1.89
	Sim	3.00 ± 0.19	1.82 ± 0.15	2.62 ± 0.18	1.54 ± 0.14
MC sgn	Theor	4.22	3.44	2.77	1.89
	Sim	4.35 ± 0.23	3.13 ± 0.20	3.62 ± 0.21	1.76 ± 0.15
RD h_w	Theor	1.94	1.14	—	—
	Sim	1.88 ± 0.16	1.17 ± 0.12	—	—
RC h_w	Theor	1.94	1.14	—	—
	Sim	1.98 ± 0.16	1.20 ± 0.12	—	—

a significant deviation from the asymptotic values for some detectors/noise distributions/signal configurations. This is due to the finite signal length, which in most cases was 30, except

TABLE IV
ERROR PROBABILITIES (IN %) FOR VARIOUS DETECTORS AND NOISE DISTRIBUTIONS

Detector		Noise distribution $CN(\epsilon, \sigma^2)$ $\epsilon = 0$ $\sigma = 1$	$\epsilon = 0.05$ $\sigma = 3$	$\epsilon = 0.25$ $\sigma = 3$	$\epsilon = 0.1$ $\sigma = 10$
MD x^2	Theor	0.62	2.06	11.06	36.54
	Sim	0.67 ± 0.09	2.20 ± 0.17	10.81 ± 0.35	31.64 ± 0.53
MD $\rho_{1.5}$	Theor	0.72	1.10	4.08	2.28
	Sim	0.78 ± 0.10	0.94 ± 0.11	4.24 ± 0.23	2.45 ± 0.17
MC $\psi_{1.5}$	Theor	0.72	1.10	4.08	2.28
	Sim	0.72 ± 0.10	1.12 ± 0.12	4.64 ± 0.24	2.60 ± 0.18
MD $!x!$	Theor	2.87	3.44	6.74	4.62
	Sim	1.31 ± 0.13	1.82 ± 0.15	4.77 ± 0.24	2.92 ± 0.19
MC sgn	Theor	2.87	3.44	6.74	4.62
	Sim	2.38 ± 0.17	3.13 ± 0.20	6.54 ± 0.28	4.42 ± 0.23
RD Φ^{-1}	Theor	0.62	1.29	6.21	4.79
	Sim	0.67 ± 0.09	1.27 ± 0.13	6.92 ± 0.29	4.98 ± 0.25
RD $h_{1.5}$	Theor	0.72	1.09	4.60	2.46
	Sim	0.69 ± 0.09	1.13 ± 0.12	5.85 ± 0.27	3.60 ± 0.21
RD h_w	Theor	0.74	1.14	4.41	2.51
	Sim	0.68 ± 0.09	1.17 ± 0.12	5.04 ± 0.25	3.24 ± 0.20

Signal configuration (S2) is used with $K = \sqrt{15}$ and $n = 30$.

TABLE V
ERROR PROBABILITIES (IN %) FOR VARIOUS DETECTORS AND NOISE DISTRIBUTIONS

Detector		Noise distribution $CN(\epsilon, \sigma^2)$ $\epsilon = 0$ $\sigma = 1$	$\epsilon = 0.05$ $\sigma = 3$	$\epsilon = 0.25$ $\sigma = 3$	$\epsilon = 0.1$ $\sigma = 10$
MD x^2	Theor	0.64	2.06	11.06	36.54
	Sim	0.60 ± 0.09	2.32 ± 0.17	11.03 ± 0.35	33.28 ± 0.53
MD $\rho_{1.5}$	Theor	0.76	1.07	3.41	1.83
	Sim	0.75 ± 0.10	1.15 ± 0.12	3.80 ± 0.22	2.09 ± 0.16
MC $\psi_{1.5}$	Theor	0.76	1.07	3.41	1.83
	Sim	0.74 ± 0.10	1.15 ± 0.12	3.56 ± 0.21	1.94 ± 0.16
MD $!x!$	Theor	1.52	1.89	4.27	2.70
	Sim	1.22 ± 0.12	1.54 ± 0.14	3.98 ± 0.22	2.46 ± 0.18
MC sgn	Theor	1.52	1.89	4.27	2.70
	Sim	1.41 ± 0.13	1.76 ± 0.15	4.30 ± 0.23	2.77 ± 0.19

Signal configuration (S4) is used with $K = \sqrt{15}$ and $n = 30$.

in Table VI. This table shows how the error probabilities depend on n for signal configuration (S2) and various detectors in impulsive noise ($CN(0.25, 3^2)$). The convergence to the asymptotic error probability is quite clear but slow for some detectors, MD $|x|$ and RD $h_{1.5}$. Next, Fig. 5 shows that the agreement with the asymptotic values is greater for smaller amplitudes K, which is natural since the convergence to a normal distribution for test statistics is usually slower in the tails. Moreover, the relative error of the estimates is roughly proportional to $1/\sqrt{P_e}$ for low error probabilities. The simulated error probabilities are further away from the asymptotic

TABLE VI
SIMULATED ERROR PROBABILITIES (IN %) FOR VARIOUS SIGNAL LENGTHS COMPARED TO THE (THEORETICALLY) ASYMPTOTIC

Detector	Signal length (n)				
	10	30	60	100	∞
MD x^2	10.72 ± 0.38	11.05 ± 0.35	10.94 ± 0.35	10.92 ± 0.35	11.06
MD $\rho_{1.5}$	5.58 ± 0.26	4.44 ± 0.23	4.46 ± 0.23	4.31 ± 0.23	4.08
MD $\|x\|$	5.34 ± 0.25	4.76 ± 0.24	5.23 ± 0.25	5.35 ± 0.25	6.74
MD $h_{1.5}$	7.30 ± 0.29	5.51 ± 0.26	5.27 ± 0.25	5.07 ± 0.25	4.60

Signal configuration (S2) is used with $K = \sqrt{15}$ and $F = CN(0.25, 3^2)$.

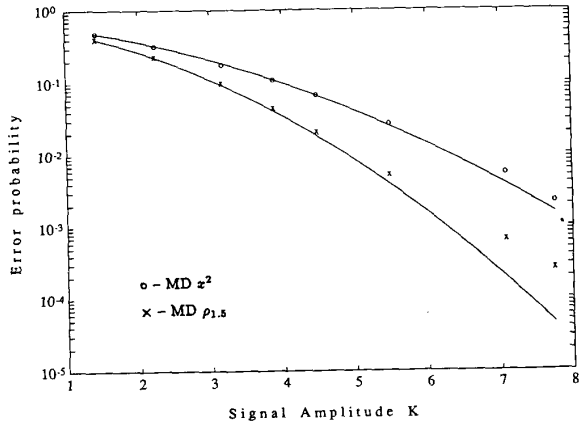

Fig. 5. Asymptotic (solid lines) and simulated error probabilities for various signal amplitudes K. The signal configuration is (S2) with $n = 30$ and $F = CN(0.25, 3^2)$.

values for impulsive noise ($CN(0.1, 10^2)$ and $CN(0.25, 3^2)$). This is particularly clear for the rank-based detectors. As far as detector types are concerned, the agreement between simulated and asymptotic values is relatively good for all detectors except one—the MD $|x|$-detector. This discrepancy is actually beneficial, since the simulated error probabilities are constantly lower than expected. The reason for this is probably that the statistic $\boldsymbol{T}^{(4)}$ converges in distribution to 0 rather slowly because of the discontinuity of $\rho'(x)$ at zero.

APPENDIX

Proof of Lemma 1: Let i indicate that H_i is true, and define first the vectors $\boldsymbol{T}_n^{(l)} = (T_{n1}^{(l)}, \cdots, T_{np}^{(l)})$, $l = 1, \cdots, 4$, by

$$T_{nj}^{(1)} = \frac{K}{\sqrt{n}} \sum_{k=1}^{n} c_{jk} \psi(e_k),$$

$$T_{nj}^{(2)} = \frac{K}{\sqrt{n}} \sum_{k=1}^{n} c_{jk} \left(\psi\left(e_k + \frac{K}{\sqrt{n}} \left(c_{ik} - \frac{1}{2} c_{jk} \right) \right) - \psi(e_k) - E\psi\left(e_k + \frac{K}{\sqrt{n}} \left(c_{ik} - \frac{1}{2} c_{jk} \right) \right) \right),$$

$$T_{nj}^{(3)} = \frac{K}{\sqrt{n}} \sum_{k=1}^{n} c_{jk} E\psi\left(e_k + \frac{K}{\sqrt{n}} \left(c_{ik} - \frac{1}{2} c_{jk} \right) \right),$$

and

$$T_{nj}^{(4)} = \sum_{k=1}^{n} \left(\rho\left(e_k + \frac{K}{\sqrt{n}} c_{ik} \right) - \rho\left(e_k + \frac{K}{\sqrt{n}} (c_{ik} - c_{jk}) \right) - \frac{K}{\sqrt{n}} c_{jk} \psi\left(e_k + \frac{K}{\sqrt{n}} \left(c_{ik} - \frac{1}{2} c_{jk} \right) \right) \right),$$

for $j = 1, \cdots, p$. It follows then from the definitions of $\boldsymbol{T}_{M,n}^{C}$ and $\boldsymbol{T}_{M,n}^{D}$ that $\boldsymbol{T}_{M,n}^{C} = \boldsymbol{T}_n^{(1)} + \cdots + \boldsymbol{T}_n^{(3)}$ and $\boldsymbol{T}_{M,n}^{D} = \boldsymbol{T}_n^{(1)} + \cdots + \boldsymbol{T}_n^{(4)}$.

From Lemma A.1 and (C) we obtain

$$\boldsymbol{T}_n^{(1)} \xrightarrow{d} N(0, K^2 A_M(\psi, F)\boldsymbol{\Sigma}). \tag{A.1}$$

The components of $\boldsymbol{T}_n^{(2)}$ satisfy (since $E\psi(e) = 0$)

$$E(T_{nj}^{(2)})^2 = \operatorname{var} T_{nj}^{(2)} \le \frac{K^2}{n} \sum_{k=1}^{n} \left(c_{jk}^2 \int (\psi(x + h_{k,n}) - \psi(x))^2 \, dF(x) \right),$$

where

$$h_{k,n} = \frac{K}{\sqrt{n}} \left(c_{ik} - \frac{1}{2} c_{jk} \right).$$

Since (C) implies that

$$\lim_{n \to \infty} \max_{1 \le k \le n} \frac{|c_{ik}|}{\sqrt{n}} = 0, \qquad i = 1, \cdots, p \tag{A.2}$$

(cf. [5, Theorem A10]), it follows from (C) and (3.2) that $E(T_{nj}^{(2)})^2 \to 0$ as $n \to \infty$ and hence,

$$\boldsymbol{T}_n^{(2)} \xrightarrow{p} 0, \tag{A.3}$$

by Chebyshev's inequality. Next, we may expand the components of $\boldsymbol{T}_n^{(3)}$ as

$$T_{nj}^{(3)} = \frac{K}{\sqrt{n}} \sum_{k=1}^{n} c_{jk} \int \psi(x + h_{k,n}) \, dF(x) \tag{A.4}$$

and it follows from (C), (3.3), and (A.2) that

$$T_{nj}^{(3)} \to K^2 B_M(\psi, F) \left(\sigma_{ij} - \frac{1}{2} \sigma_{jj} \right)$$

as $n \to \infty$, or equivalently

$$\boldsymbol{T}_n^{(3)} \to K^2 B_M(\psi, F) \left(\boldsymbol{\sigma}_i - \frac{1}{2} \boldsymbol{\sigma} \right) = \boldsymbol{\mu}_i. \tag{A.5}$$

Finally, for $\boldsymbol{T}_n^{(4)}$ we treat the variance and mean separately, yielding

$$\operatorname{var} T_{nj}^{(4)} \le \sum_{k=1}^{n} \int (\rho(x + h_{k,n} + q_{k,n}) - \rho(x + h_{k,n} - q_{k,n}) - 2q_{k,n}\psi(x + h_{k,n}))^2 \, dF(x),$$

where

$$q_{k,n} = \frac{1}{2} \frac{K}{\sqrt{n}} c_{jk}$$

and

$$ET_{nj}^{(4)} = \sum_{k=1}^{n} \int (\rho(x + h_{k,n} + q_{k,n}) - \rho(x + h_{k,n} - q_{k,n}) - 2q_{k,n}\psi(x + h_{k,n}))\, dF(x).$$

It follows from (C), (3.4), and (3.5) that $\operatorname{var} T_{nj}^{(4)} \to 0$ and $ET_{nj}^{(4)} \to 0$ as $n \to \infty$ and hence, by Chebyshev's inequality

$$\boldsymbol{T}_n^{(4)} \xrightarrow{p} 0. \tag{A.6}$$

Slutsky's Lemma now yields the asserted asymptotic normality for $\boldsymbol{T}_{M,n}^{C}$ and $\boldsymbol{T}_{M,n}^{D}$. □

Proof of Lemma 2: Observe that

$$\boldsymbol{T}_{R,n}^{C} = K(S_{n1}(\boldsymbol{\theta}_1), \cdots, S_{np}(\boldsymbol{\theta}_p)), \tag{A.7}$$

where $S_{nj}(\boldsymbol{\theta}_j)$ is the jth component of $\boldsymbol{S}_n(\boldsymbol{\theta}_j)$ in (A.18) and $\boldsymbol{\theta}_j = -K(\boldsymbol{u}_i - \frac{1}{2}\boldsymbol{u}_j)$ under H_i. It then follows from Corollary A.1 and (C) that

$$\boldsymbol{T}_{R,n}^{C} - K\boldsymbol{S}_n(0) \xrightarrow{p} K^2 B_R(h, F)\left(\boldsymbol{\sigma}_i - \frac{1}{2}\boldsymbol{\sigma}\right) = \boldsymbol{\mu}_i. \tag{A.8}$$

Since

$$K\boldsymbol{S}_n(0) \xrightarrow{d} N(0, K^2 A_R(h)\boldsymbol{\Sigma}). \tag{A.9}$$

again according to Corollary A.1, the convergence result for $\boldsymbol{T}_{R,n}^{C}$ follows from Slutsky's Lemma. In order to prove asymptotic normality for $\boldsymbol{T}_{R,n}^{D}$, we introduce

$$Q_n(\boldsymbol{e}_n - \boldsymbol{\theta}\boldsymbol{C}_n') = D_n(\boldsymbol{e}_n) - \sqrt{n}\boldsymbol{\theta}\boldsymbol{S}_n(0)' + \frac{n}{2}B_R(h, F)\boldsymbol{\theta}\boldsymbol{\Sigma}\boldsymbol{\theta}', \tag{A.10}$$

which is a quadratic approximation of D_n around 0, and

$$\boldsymbol{T}_n^{*} = (Q_n(\boldsymbol{Y}_n) - Q_n(\boldsymbol{Y}_n - \alpha_n\boldsymbol{c}_{1n}), \cdots, Q_n(\boldsymbol{Y}_n) - Q_n(\boldsymbol{Y}_n - \alpha_n\boldsymbol{c}_{pn})). \tag{A.11}$$

Inserting (A.10) into (A.11) yields

$$\boldsymbol{T}_n^{*} = K\boldsymbol{S}_n(0) + \boldsymbol{\mu}_i$$

under H_i, from which

$$\boldsymbol{T}_n^{*} \xrightarrow{d} N(\boldsymbol{\mu}_i, K^2 A_R(h)\boldsymbol{\Sigma}) \tag{A.12}$$

follows (Corollary A.1). It remains to show that $\boldsymbol{T}_n^{*}$ is close to $\boldsymbol{T}_{R,n}^{D}$, which is a consequence of Q_n being close to D_n. Actually,

$$\boldsymbol{T}_n^{*} - \boldsymbol{T}_{R,n}^{D} \xrightarrow{p} 0 \tag{A.13}$$

follows easily from applying Lemma A.3, with $\boldsymbol{\theta}$ taking the values $-K\boldsymbol{u}_i$, $-K(\boldsymbol{u}_i - \boldsymbol{u}_1), \cdots, -K(\boldsymbol{u}_i - \boldsymbol{u}_p)$, respectively. The asymptotic normality of $\boldsymbol{T}_{R,n}^{D}$ follows then from (A.12) and (A.13). □

Lemma A.1: Let e_i, $i = 1, 2, \cdots$ be i.i.d. random variables with $Ee_i = 0$, $Ee_i^2 = \sigma^2$ and put

$$\boldsymbol{S} = (S_1, \cdots, S_p) = \frac{1}{\sqrt{n}}\boldsymbol{e}_n\boldsymbol{C}_n,$$

where $\boldsymbol{e}_n = (e_1, \cdots, e_n)$ and $\boldsymbol{C}_n$ satisfies (C). Then, there exists a random vector $\boldsymbol{Z}$ such that

$$\boldsymbol{S} \xrightarrow{d} \boldsymbol{Z} \sim N(0, \sigma^2\boldsymbol{\Sigma}) \tag{A.14}$$

as $n \to \infty$.

Proof: According to the "Cramer–Wold device," it is enough to show that

$$\boldsymbol{\lambda}\boldsymbol{S}' \xrightarrow{d} \boldsymbol{\lambda}\boldsymbol{Z}' \sim N(0, \sigma^2\boldsymbol{\lambda}\boldsymbol{\Sigma}\boldsymbol{\lambda}') \tag{A.15}$$

as $n \to \infty$ for each $\boldsymbol{\lambda} \in \mathbb{R}^p$. When $\boldsymbol{\lambda}\Sigma\lambda' > 0$, (A.15) follows from the Lindeberg–Feller central limit theorem, see for instance [5], Theorem A.12. When $\boldsymbol{\lambda}\boldsymbol{\Sigma}\boldsymbol{\lambda}' = 0$ we have

$$E(\boldsymbol{\lambda}\boldsymbol{S}')^2 = \sigma^2\boldsymbol{\lambda}\frac{\boldsymbol{C}_n'\boldsymbol{C}_n}{n}\boldsymbol{\lambda}' \to 0, \qquad \text{as } n \to \infty,$$

because of (C). Chebyshev's inequality then gives (A.15). □

Lemma A.2: Let $\{x_k\}$ and $\{z_k\}$ be sequences of real numbers such that

$$\lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} x_k^2 = \lambda_1,$$

$$\lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} z_k^2 = \lambda_2,$$

and

$$\lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} x_k z_k = \lambda_3$$

hold with finite limits. Put

$$S_n(K) = \frac{1}{\sqrt{n}}\sum_{k=1}^{n} x_k \operatorname{sgn}\left(e_k + \frac{K}{\sqrt{n}}z_k\right) a_n(R_{nk}^{+}(K)),$$

with $R_{nk}^{+}(K)$ being the rank of $|e_k + Kz_k/(\sqrt{n})|$ and $a_n(\cdot)$ defined as in Example 2. Then,

$$S_n(0) \xrightarrow{d} N(0, A_R(h)\lambda_1) \tag{A.16}$$

and

$$S_n(K) - S_n(0) \xrightarrow{p} K\lambda_3 B_R(h, F) \tag{A.17}$$

as $n \to \infty$, with A_R and B_R as defined in (3.8)–(3.9).

Proof: Formula (A.16) may be proved exactly as in [3, Theorem V.1.7] (where the case $x_k \equiv 1$ is treated), and (A.17) follows from e.g., [17, Theorem 3.2]. □

Corollary A.1: Let

$$\boldsymbol{S}_n(\boldsymbol{\theta}) = \frac{1}{\sqrt{n}}\sum_{k=1}^{n}\boldsymbol{c}_k \operatorname{sgn}\left(e_k - \frac{1}{\sqrt{n}}\boldsymbol{\theta}\cdot\boldsymbol{c}_k\right)a_n(R_{nk}^{+}(\boldsymbol{\theta})), \tag{A.18}$$

where $\boldsymbol{c}_k = (c_{1k}, \cdots, c_{pk})$, R_{nk}^{+} is the rank of $|e_k - \boldsymbol{\theta}\cdot\boldsymbol{c}_k/\sqrt{n}|$ and $a_n(\cdot)$ is defined as in Example 2. Then,

$$\boldsymbol{S}_n(0) \xrightarrow{d} N(0,\ A_R(h)\boldsymbol{\Sigma}) \tag{A.19}$$

and

$$\boldsymbol{S}_n(\boldsymbol{\theta}) - \boldsymbol{S}_n(0) \xrightarrow{p} -B_R(h,\ F)\boldsymbol{\theta}\boldsymbol{\Sigma}. \tag{A.20}$$

Proof: According to the "Cramer–Wold device," (A.19) follows once we show

$$\begin{aligned}\boldsymbol{\lambda}\boldsymbol{S}_n(0)' &= \frac{1}{\sqrt{n}}\sum_{k=1}^{n} d_k \operatorname{sgn}(e_k)a_n(R_{nk}^{+}(0)) \\ &\xrightarrow{d} N(0,\ A_R(h)\boldsymbol{\lambda}\boldsymbol{\Sigma}\boldsymbol{\lambda}'),\end{aligned}$$

for each $\boldsymbol{\lambda}\in\mathbb{R}^p$ and with $d_k = \boldsymbol{\lambda}\boldsymbol{c}_k'$. But this is immediate from (A.16), since

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n} d_k^2 = \boldsymbol{\lambda}\boldsymbol{\Sigma}\boldsymbol{\lambda}'.$$

In the same manner, (A.20) follows from (A.17). □

Lemma A.3 Let Q_n (cf. (A.10)) be the quadratic approximation of the rank-distance $D_{R,n}$ (cf. (2.11)). Then,

$$D_{R,n}\left(\boldsymbol{e}_n - \frac{1}{\sqrt{n}}\boldsymbol{\theta}\boldsymbol{C}_n'\right) - Q_n\left(\boldsymbol{e}_n - \frac{1}{\sqrt{n}}\boldsymbol{\theta}\boldsymbol{C}_n'\right) \xrightarrow{p} 0, \tag{A.21}$$

for each $\boldsymbol{\theta}\in\mathbb{R}^p$.

Proof: Put

$$\operatorname{Rem}(\boldsymbol{\theta}) = D_{R,n}\left(\boldsymbol{e}_n - \frac{1}{\sqrt{n}}\boldsymbol{\theta}\boldsymbol{C}_n'\right) - Q_n\left(\boldsymbol{e}_n - \frac{1}{\sqrt{n}}\boldsymbol{\theta}\boldsymbol{C}_n'\right).$$

Then, according to (A.20)

$$\frac{\partial \operatorname{Rem}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}} = -\boldsymbol{S}_n(\boldsymbol{\theta}) + \boldsymbol{S}_n(0) - B_R(h,\ F)\boldsymbol{\theta}\boldsymbol{\Sigma} \xrightarrow{p} 0, \tag{A.22}$$

for each $\boldsymbol{\theta}\in\mathbb{R}^p$, with $\boldsymbol{\Sigma}$ defined as in Corollary A.1. It follows as in [4, proof of Lemma 3.1], that (A.21) and (A.22) are actually equivalent. The proof makes use of the fact that $D_{R,n}(\boldsymbol{e}_n - 1/\sqrt{n}\boldsymbol{\theta}\boldsymbol{C}_n')$ is a convex function of $\boldsymbol{\theta}$, which in turn is a consequence of that $D_{R,n}$ defines a norm in $\mathbb{R}^n$ whenever the scores $a_n(1), \cdots, a_n(n)$ are nonnegative and nondecreasing (cf. [14, Theorem 2.1]). □

ACKNOWLEDGMENT

The authors wish to thank S. Janson, P. Rousseeuw, and T.Öberg for valuable comments and suggests on the work, and the Associate Editor and referees for their careful reading of the manuscript, which considerably improved the appearance of the paper.

REFERENCES

[1] P. J. Bickel, "One-step Huber estimates in the linear model," *J. Amer. Statist. Assoc.*, vol. 70, pp. 428–434, 1975.
[2] A. H. El-Sawy and V. D. VandeLinde, "Robust detection of known signals," *IEEE Trans. Inform. Theory*, vol. IT-23, pp. 722–727, 1977.
[3] J. Hájek and Z. Šidák, *Theory of Rank Tests.* New York: Academic Press, 1967.
[4] S. Heiler and R. Willers, "Asymptotic normality of *R*-estimates in the linear model," *Math. Operationsforsch. und Stat., Series Statistics*, vol. 19, pp. 173–184, 1988.
[5] T. P. Hettmansperger, *Statistical Inference Based on Ranks.* New York: Wiley, 1984.
[6] T. P. Hettmansperger and J. W. McKean, "A geometric interpretation of inferences based on ranks in the linear model," *J. Amer. Statist. Assoc.*, vol. 78, pp. 885–893, 1983.
[7] P. J. Huber, "Robust estimation of a location parameter," *Ann. Math. Statist.*, vol. 35, pp. 73–101, 1964.
[8] ——, *Robust Statistics.* New York: Wiley, 1981.
[9] L. A. Jaeckel, "Robust estimates of location: Symmetry and asymmetric contamination," *Ann. Math. Statist.*, vol. 42, pp. 1020–1034, 1971.
[10] S. A. Kassam, "Detection of narrowband signals: Asymptotic robust ness and quantization," in *Proc. 1984 Conf. Inform. Sciences and Systems*, Princeton Univ., Princeton, NJ, 1984, pp. 297–301.
[11] S. A. Kassam and H. V. Poor, "Robust techniques for signal proces sing: A survey," *Proc. IEEE*, vol. 73, pp. 433–481, 1985.
[12] 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.
[13] R. D. Martin and S. C. Schwartz, "Robust detection of a known sig nal in nearly Gaussian noise," *IEEE Trans. Inform. Theory*, vol. IT-17, pp. 50–56, Jan. 1971.
[14] J. W. McKean and R. M. Schrader, "The geometry of robust procedures in linear models," *J. Roy. Statist. Soc. B*, vol. IT-42, pp. 366–371, 1 980.
[15] P. Rousseeuw, "Optimally robust procedures in the infinitesimal s ense," in *Proc. 42nd Session of the ISI, Manila*, 1979, pp. 467–470.
[16] ——, "Estimation and testing by means of optimally robust statistics," *Belgisch Tijdschrift voor Statistiek, Informatica en Operationeel Onderzoek*, vol. 22, no. 3, pp. 3–19, 1982.
[17] C. van Eeden, "An analogue, for signed rank statistics, of Jure čková's asymptotic linearity theorem for rank statistics," *Ann. Math. Statist.*, vol. 43, pp. 791–802, 1972.
[18] H. L. van Trees, *Detection, Estimation and Modulation Theory, 1.* New York: Wiley, 1968.