IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received July 24, 2019, accepted August 29, 2019, date of publication September 2, 2019, date of current version September 16, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2938856

A Data-Cleaning Approach to Robust Multisensor Detection of Improper Signals

JITENDRA K. TUGNAIT, (Life Fellow, IEEE)
Department of Electrical and Computer Engineering, Auburn University, Auburn, AL 36849, USA
e-mail: tugnajk@auburn.edu

This work was supported by the National Science Foundation under Grant CCF-1617610.

ABSTRACT We consider the problem of detecting the presence of an improper complex-valued signal, common among two or more sensors (channels), in the presence of spatially independent, colored improper noise and additive outliers. A source of improper noise is in-phase/quadrature-phase imbalance during down-conversion of bandpass noise to baseband at the receiver. For clean data (i.e., known to be outlier free), a nonparametric binary hypothesis testing formulation in frequency-domain, utilizing the discrete Fourier transform of an augmented measurement sequence, has been proposed in the literature, and it results in a generalized likelihood ratio test (GLRT). In this paper we robustify this approach against additive outliers in the data, using a data-cleaning approach. Using some existing robust estimators of multivariate scatter, we first detect the outliers, and subsequently remove and replace them using vector median filtering to yield cleaned data. The existing GLRT is then applied to the cleaned data. The approach is illustrated via simulations. The considered problem has applications in diverse areas including spectrum sensing for cognitive radio.

INDEX TERMS Improper complex random signals, generalized likelihood ratio test (GLRT), robust detection, multichannel signal detection, spectral analysis, hypothesis testing.

I. INTRODUCTION

Robust detection of a complex-valued signal observed at multiple sensors in the presence of noise and additive outliers is considered in this paper. A frequency-domain nonparametric approach is followed, without requiring detailed modeling of correlation function or statistical distribution of signal or noise, except requiring that both signal and noise are stationary. It is assumed that signal is common to all sensors whereas noise and outliers are independent across sensors. Moreover, signal, noise and outliers are all allowed to be improper.

A complex-valued, stationary, p-dimensional random vector sequence $\{\boldsymbol{x}(t)\}$ is said to be proper if it is uncorrelated with its complex conjugate, i.e., $E\{\boldsymbol{x}(t+\tau)(\boldsymbol{x}^*(t))^H\} = E\{\boldsymbol{x}(t+\tau)\boldsymbol{x}^\top(t)\} = E\{\boldsymbol{x}(t+\tau)\}E\{\boldsymbol{x}^\top(t)\}$ for $-\infty < t, \tau < \infty$ [1], where superscripts $*$, $\top$ and H denote the complex conjugate, transpose and Hermitian (conjugate transpose) operations, respectively, and E denotes the expectation operation. If $\{\boldsymbol{x}(t)\}$ is not proper, it is called improper [1]. Although it has typically been assumed in complex-valued signal processing in statistical signal processing and communications applications that the complex signals and noise are proper, this is not always the case [1]–[5]. In digital communications, binary phase shift keying (BPSK), amplitude shift keying (ASK) and offset quadrature phase shift keying (QPSK) modulated signals, among others, are improper [1]. Also, impropriety in both signals and noise results when there is an in-phase/quadrature-phase (IQ) imbalance in the receiver during down-conversion of bandpass signals to baseband [3]. Similarly, an IQ imbalance at the transmitter during up-conversion of baseband signals to bandpass, can result in an improper signal [3]. If $\{\boldsymbol{x}(t)\}$ is improper, signal processing performance can be much improved if complementary correlation function $E\{\boldsymbol{x}(t+\tau)(\boldsymbol{x}^*(t))^H\}$ is also exploited in addition to the usual correlation function $E\{\boldsymbol{x}(t+\tau)\boldsymbol{x}^H(t)\}$ [1], [4], [5].

Detection of spatially correlated random signals in noise is of interest in several applications [6]–[14]. An application of this problem is in spectrum sensing for cognitive radio (CR) to decide if the received signal, in addition to noise, contains signals from a single or multiple primary users [6]–[9], [11]–[14]; for other applications, see [10]. The detection problem is formulated as a binary hypothesis testing

The associate editor coordinating the review of this article and approving it for publication was Khalid Aamir.

problem and is a well-investigated topic [9]. If the underlying communications channel is frequency-selective, the received signal will be colored. An irreducible source of noise at the receiver is thermal noise which is zero-mean white Gaussian and whose baseband-equivalent model is that of a zero-mean, white, proper complex Gaussian random process (it could be improper if there is an IQ imbalance at the receiver). Often, as a result of receive filters or oversampling at the cognitive user's receiver, the filtered thermal noise will not be white (i.e., it will be colored) [14]. The effect of noise correlation on spectrum sensing performance for a class of spectrum sensing algorithms has been analyzed in [15].

In [14] a nonparametric frequency-domain approach has been presented to derive a generalized likelihood ratio test (GLRT) for detection of a spatially correlated random vector signal, possibly improper, in the presence of spatially independent, possibly improper and colored, noise. The GLRT in [14] has been derived using asymptotic distribution of a frequency-domain sufficient statistic, based on the discrete Fourier transform (DFT) of an augmented measurement sequence. The GRLT framework is needed since the parameters of the distribution of the DFT of the augmented measurement sequence are unknown under both hypotheses of the hypothesis testing problem. Detection of improper signals in proper noise, and detection of proper signals in proper noise, are also considered in [14] in the same framework. The objective of this paper is to robustify the approach of [14] against additive outliers in the data.

While several approaches exist for detection of spatially correlated random signals in random noise, based on differing signal and noise models [6]–[14], only [11]–[14] consider improper signals. Detection of improper signals are investigated in [11], which is limited to temporally white but spatially correlated improper complex Gaussian signals in temporally and spatially uncorrelated proper complex Gaussian noise. The model of [12] allows temporal correlation for both improper signal and proper noise. Both show improved performance compared to the case where improper signals are treated as proper. The model of [12] is limited to improper signals in spatially independent proper noise. In [13], [14] noise is allowed to be improper also; [13] is a preliminary conference version of [14]. In this paper we robustify the approach of [14] against additive outliers in the data.

The main contribution of this paper is robustification of the GLRT of [14] against additive outliers in the data. None of [6]–[14] consider outliers. It is well documented (see e.g., [16]–[20]) that even a small fraction of outliers (i.e., gross errors in a small fraction of measurements), when unmodeled or unaccounted for, can significantly affect the result in any statistical inference (detection or parameter estimation) problem. In the context of multisensor detection of improper signals formulated as a binary hypothesis testing problem, outliers can have a deleterious effect on the size of the test (probability of false alarm) and on the power of the test (probability of detection) for a given test size. The robustification approach discussed in this paper is based on robust estimation of multivariate covariance (scatter) of data, and it follows a similar approach used in [21] for robust comparison of multivariate proper complex-valued random signals. The basic idea is to use some robust methods for identifying outliers in the data, and then to clean the data by removing the outliers. The existing multisensor detection approach of [14] is then applied to the cleaned data. In this paper, we follow the approach of [21] with some enhancements such as using vector median filtering for outlier removal compared to component-wise (scalar) median filtering used in [21], and investigating a more (statistically) efficient MM-estimator for multivariate scatter estimation (see Sec. IV-A.2) in addition to the S-estimator (Sec. IV-A.1) used in [21]. Also, while [21] deals only with proper complex-valued signals in proper outliers, in this paper we additionally consider improper outliers also.

The paper is organized as follows. In Sec. II, we introduce notation, state the binary hypothesis testing problem, and discuss the additive outlier model of interest in this paper. In Sec. III, we review the frequency-domain GLRT solution of [14] for the binary hypothesis testing problem when clean (outlier-free) data is available. This test forms the basis for our proposed robust detector presented in Sec. IV. Simulation examples are presented in Sec. V in support of the proposed approach.

II. PRELIMINARIES AND SYSTEM MODEL

Here we first introduce notation in Sec. II-A, and describe the system model in Sec. II-B where we state the binary hypothesis testing problem of interest. In Sec. II-B.1 we introduce the widely used additive outlier model where both proper and improper outliers are considered.

A. NOTATION

The superscripts $*$, $\top$ and H denote the complex conjugate, transpose and Hermitian (conjugate transpose) operations, respectively, and E denotes the expectation operation. We use $\boldsymbol{R} \succeq 0$ and $\boldsymbol{R} \succ 0$ to denote that Hermitian $\boldsymbol{R}$ is positive semi-definite and positive definite, respectively. For a square matrix $\boldsymbol{B}$, $|\boldsymbol{B}|$ and etr($\boldsymbol{B}$) denote the determinant and the exponential of the trace of $\boldsymbol{B}$, respectively, i.e., etr($\boldsymbol{B}$) = exp(tr($\boldsymbol{B}$)), $[\boldsymbol{A}_k]_{i:l,j:m}$ denotes the submatrix of the matrix $\boldsymbol{A}_k$ comprising its rows i through l and columns j through m, $[\boldsymbol{A}_k]_{ij}$ is its ijth element, A_{ij} is the ijth element of $\boldsymbol{A}$, and $\boldsymbol{I}_p$ is the $p \times p$ identity matrix. The set of integers 1 through n is denoted by $[1, n]$. The sets of real and complex numbers are denoted by $\mathbb{R}$ and $\mathbb{C}$, respectively. Given $\boldsymbol{x} \in \mathbb{C}^p$, Re$\{\boldsymbol{x}\} \in \mathbb{R}^p$ and Im$\{\boldsymbol{x}\} \in \mathbb{R}^p$ denote its real and imaginary parts, respectively. The notation $y = \mathcal{O}(g(x))$ means that there exists some finite real number $b > 0$ such that $\lim_{x\to\infty} |y/g(x)| \leq b$. Given square matrices $\boldsymbol{B}_i$, $i = 1, 2, \cdots n$, block-diag$\{\boldsymbol{B}_1, \boldsymbol{B}_2, \cdots, \boldsymbol{B}_n\}$ denotes a square matrix with $\boldsymbol{B}_i$s along its main block-diagonal and zeros everywhere else. The notation $\delta(k - k_0)$ denotes a Kronecker delta at $k = k_0$, i.e., $\delta(k - k_0) = 1$ for $k = k_0$, and is zero otherwise.

The notation χ^2_n represents central chi-square distribution with n degrees of freedom, whereas $\chi^2_n(\lambda)$ denotes the non-central chi-square distribution with n degrees of freedom and non-centrality parameter λ. The notation $\boldsymbol{y} \sim \mathcal{N}_c(\boldsymbol{m}, \boldsymbol{\Sigma})$ denotes a random vector $\boldsymbol{y}$ that is proper (circularly symmetric), complex Gaussian with mean $\boldsymbol{m}$ and covariance $\boldsymbol{\Sigma}$ ($= E\{[\boldsymbol{y}-\boldsymbol{m}][\boldsymbol{y}-\boldsymbol{m}]^H\}$). Also, for real $\boldsymbol{y}$, $\boldsymbol{y} \sim \mathcal{N}_r(\boldsymbol{m}, \boldsymbol{\Sigma})$ denotes a real random vector $\boldsymbol{y}$ that is Gaussian with mean $\boldsymbol{m}$ and covariance $\boldsymbol{\Sigma}$. The abbreviations w.p.1. and i.i.d. stand for with probability one, and independent and identically distributed, respectively.

B. SYSTEM MODEL

A noisy signal, uniformly sampled in time, is measured at p sensors. Let $\boldsymbol{v}(t) \in \mathbb{C}^p$ denote a zero-mean, spatially independent, complex-valued, stationary, random noise sequence for $t = 0, \pm 1, \cdots$. By spatial independence we mean that the ith and kth components $v_i(t)$ and $v_k(t)$ of $\boldsymbol{v}(t)$ are independent for $i \neq k$. The sequence $\{\boldsymbol{v}(t)\}$ is allowed to be improper. Recall that zero-mean $\{\boldsymbol{v}(t)\}$ is proper if $E\{\boldsymbol{v}(t+\tau)(\boldsymbol{v}^*(t))^H\} = E\{\boldsymbol{v}(t+\tau)\boldsymbol{v}^\top(t)\} = \boldsymbol{0}$ for $\tau = 0, \pm 1, \pm 2, \cdots$, otherwise it is improper [1]. Let $\boldsymbol{s}(t) \in \mathbb{C}^p$ denote a zero-mean, stationary, complex-valued, random signal sequence which is independent of $\{\boldsymbol{v}(t)\}$. The sequence $\{\boldsymbol{s}(t)\}$ also could be improper. Both noise and signal may be non-Gaussian.

The measurement vector sequence at p sensors is denoted by $\boldsymbol{y}(t) \in \mathbb{C}^p$. The problem considered in this paper is to decide if $\boldsymbol{y}(t) = \boldsymbol{v}(t)$ (noise only) or $\boldsymbol{y}(t) = \boldsymbol{s}(t) + \boldsymbol{v}(t)$ (signal and noise), given n samples $\boldsymbol{y}(t), t = 0, 1, \cdots, n$. This problem has previously been considered in [14]. In this paper we revisit this problem by allowing a class of outliers in the measurement sequence, unlike [14].

Let $\mathcal{H}_0$ denote the null hypothesis that the sensors are receiving just noise, and $\mathcal{H}_1$ is the alternative that signal common to all sensors is also present. The following binary hypothesis testing problem for the measurement sequence $\boldsymbol{y}(t), t = 0, 1, \cdots, n-1$, is addressed in [14]:

$$\begin{aligned} \mathcal{H}_0 &: \boldsymbol{y}(t) = \boldsymbol{v}(t), \quad \text{noise only} \\ \mathcal{H}_1 &: \boldsymbol{y}(t) = \boldsymbol{s}(t) + \boldsymbol{v}(t), \quad \text{signal and noise.} \end{aligned} \tag{1}$$

In [14] no assumption is made regarding the distribution of $\boldsymbol{s}(t)$ or $\boldsymbol{v}(t)$ (they may be Gaussian or non-Gaussian), or regarding detailed correlation structure of $\boldsymbol{s}(t)$ or $\boldsymbol{v}(t)$.

1) ADDITIVE OUTLIERS

In this paper, unlike [14], we assume that data may be corrupted with outliers. In particular, we consider the widely used additive outlier model for measurements. This model is not essential to robust processing, but is one of the useful models frequently used in the literature [16, p. 253], [18, p.74], and is used in our simulations. Let $\{\boldsymbol{y}(t)\}$ be corrupted by additive outliers $\boldsymbol{v}_y(t)$ to yield corrupted measurement $\{\tilde{\boldsymbol{y}}(t)\}$, given by

$$\tilde{\boldsymbol{y}}(t) = \boldsymbol{y}(t) + \boldsymbol{v}_y(t), \tag{2}$$

where, with a "small" probability p_o, we have either

$$\boldsymbol{v}_y(t) \sim \mathcal{N}_c(0, \sigma^2_{vy}\boldsymbol{I}_p), \quad \sigma^2_{vy} \gg \mathbb{E}\{\|\boldsymbol{y}(t)\|^2\}, \tag{3}$$

or

$$\boldsymbol{v}_y(t) = \frac{1}{\sqrt{2}}\boldsymbol{v}_c(t) + \frac{1}{\sqrt{2}}\boldsymbol{v}^*_c(t), \quad \boldsymbol{v}_c(t) \sim \mathcal{N}_c(0, \sigma^2_{vy}\boldsymbol{I}_p), \tag{4}$$

with $\sigma^2_{vy} \gg \mathbb{E}\{\|\boldsymbol{y}(t)\|^2\}$, the outliers are i.i.d. and independent of the clean signals, and with probability $1 - p_o$,

$$\tilde{\boldsymbol{y}}(t) = \boldsymbol{y}(t). \tag{5}$$

In outlier model (3), we have

$$E\{\boldsymbol{v}_y(t+\tau)\boldsymbol{v}^H_y(t)\} = \sigma^2_{vy}\boldsymbol{I}_p\delta(\tau), \tag{6}$$

$$E\{\boldsymbol{v}_y(t+\tau)\boldsymbol{v}^\top_y(t)\} = \boldsymbol{0}\ \forall\tau. \tag{7}$$

In outlier model (4), we have

$$E\{\boldsymbol{v}_y(t+\tau)\boldsymbol{v}^H_y(t)\} = \sigma^2_{vy}\boldsymbol{I}_p\delta(\tau), \tag{8}$$

$$E\{\boldsymbol{v}_y(t+\tau)\boldsymbol{v}^\top_y(t)\} = j\sigma^2_{vy}\boldsymbol{I}_p\delta(\tau). \tag{9}$$

Thus, in outlier model (3), the outliers $\boldsymbol{v}_y(t)$ are proper, whereas in outlier model (4), the outliers are improper.

III. GENERALIZED LIKELIHOOD RATIO TEST FOR OUTLIER-FREE DATA

In this section we review the frequency-domain GLRT of [14] for binary hypothesis testing problem (1) using clean (outlier-free) data. This test forms the basis for our proposed robust detector (Sec. IV) after detection of outliers in the time-domain data and subsequent removal (cleaning) of the outliers. The approach of [14] is then applied to the cleaned data.

Define the augmented complex process $\{\boldsymbol{z}(t)\}$ and the real-valued process $\{\boldsymbol{x}(t)\}$ as

$$\boldsymbol{z}(t) = \begin{bmatrix} \boldsymbol{y}(t) \\ \boldsymbol{y}^*(t) \end{bmatrix}, \quad \boldsymbol{x}(t) = \begin{bmatrix} \boldsymbol{y}_r(t) \\ \boldsymbol{y}_i(t) \end{bmatrix} \tag{10}$$

where $\boldsymbol{y}(t) = \boldsymbol{y}_r(t) + j\boldsymbol{y}_i(t)$, with $\boldsymbol{y}_r(t)$ and $\boldsymbol{y}_i(t)$ denoting its real and imaginary components, respectively. Consider the (normalized) DFT $\boldsymbol{d}_z(f_n)$ of complex-valued $\boldsymbol{z}(t)$, $t = 0, 1, \cdots, n-1$, given by

$$\boldsymbol{d}_z(f_k) := \frac{1}{\sqrt{n}}\sum_{t=0}^{n-1}\boldsymbol{z}(t)e^{-j2\pi f_k t} \tag{11}$$

where $f_k = k/n$, $k = 0, 1, \cdots, n-1$. Using the results of [22, p. 280, Sec. 6.2] and symmetry properties of $\boldsymbol{d}_z(f_k)$, it is shown in [14] that the set of complex-valued random vectors $\{\boldsymbol{d}_z(f_k)\}_{k=0}^{n/2}$ is a sufficient statistic for the binary hypothesis testing problem (1).

The following assumptions on $\boldsymbol{z}(t)$ and $\boldsymbol{x}(t)$ are imposed in [14].

(A1) *Assumption 2.6.1* [23]. Random sequence $\{\boldsymbol{x}(t)\}$, $\boldsymbol{x}(t) \in \mathbb{R}^{2p}$, is stationary with components $\boldsymbol{x}_a(t)$, $a = 1, 2, \cdots, 2p$, such that $E\{|\boldsymbol{x}_a(t)|^m\} < \infty$, satisfying

$$\sum_{\substack{\tau_1,\tau_2,\cdots,\\ \tau_{m-1}=-\infty}}^{\infty} |c_{\boldsymbol{x}:a_1,a_2,\cdots,a_m}(\tau_1, \tau_2, \cdots, \tau_{m-1})| < \infty \tag{12}$$

for $a_1, a_2, \cdots, a_m = 1, 2, \cdots, 2p$ and $m = 2, 3, \cdots$, where $c_{\boldsymbol{x}:a_1,a_2,\cdots,a_m}(\tau_1, \tau_2, \cdots, \tau_{m-1})$ is the joint cumulant function of order m of stationary $\{\boldsymbol{x}(t)\}$.

(A2) The power spectral density (PSD) matrix of $\boldsymbol{z}(t)$, $\boldsymbol{S}_z(f) \succ 0$ for any $0 \le f \le 1$.

(A3) The PSD matrix $\boldsymbol{S}_z(f_k)$ is locally smooth such that it is constant over K ($\ge 2p$) consecutive frequency points, where $f_k = k/n$, $k \in [0, n-1]$.

See [24, Appendix A] (also [25, Appendix A]) for more details regarding satisfaction of Assumption 2.6.1 of [23] for a class of stationary improper sequences. If $\boldsymbol{y}(t)$ is Gaussian, then it is sufficient to verify (12) only for $m = 2$. Regarding assumption (A2), one can always add artificial proper white Gaussian noise to $\boldsymbol{y}(t)$ to achieve $\boldsymbol{S}_z(f) \succ 0$. Assumption (A3) is a standard assumption in PSD estimation literature [23].

Following [23, Theorem 4.4.1], it is shown in [14] that under (A1), asymptotically (as $n \to \infty$), $\boldsymbol{d}_z(f_k)$, $k = 1, 2, \cdots, (n/2) - 1$, ($n$ even), are asymptotically independent proper complex Gaussian $\mathcal{N}_c(\boldsymbol{0}, \boldsymbol{S}_z(f_k))$ random vectors, respectively, where

$$\boldsymbol{S}_z(f) = \begin{bmatrix} \boldsymbol{S}_y(f) & \tilde{\boldsymbol{S}}_y(f) \\ \tilde{\boldsymbol{S}}_y^*(-f) & \boldsymbol{S}_y^*(-f) \end{bmatrix}, \tag{13}$$

$\boldsymbol{S}_y(f) = \sum_{\tau=-\infty}^{\infty} \boldsymbol{R}_y(\tau) e^{-j2\pi f\tau}$ denotes the PSD of $\{\boldsymbol{y}(t)\}$, $\boldsymbol{R}_y(\tau) = E\{\boldsymbol{y}(t+\tau)\boldsymbol{y}^H(t)\}$, $\tilde{\boldsymbol{S}}_y(f) = \sum_{\tau=-\infty}^{\infty} \tilde{\boldsymbol{R}}_y(\tau) e^{-j2\pi f\tau}$ is the complementary PSD (C-PSD) of $\{\boldsymbol{y}(t)\}$, and $\tilde{\boldsymbol{R}}_y(\tau) = E\{\boldsymbol{y}(t+\tau)\boldsymbol{y}^\top(t)\}$. Also, while $\boldsymbol{d}_z(f_0)$ and $\boldsymbol{d}_z(f_{n/2})$, ($n$ even), are independent Gaussian random vectors, also independent of $\boldsymbol{d}_z(f_k)$ for $k = 1, 2, \cdots, (n/2) - 1$, they are, in general, improper whereas $\boldsymbol{d}_z(f_k)$ for $k = 1, 2, \cdots, (n/2) - 1$ is proper.

In [14], the GLRT for test (1) is derived using the frequency-domain sufficient statistic $\{\boldsymbol{d}_y(f_k)\}_{k=0}^{n/2}$ after discarding the end frequency points f_0 and $f_{n/2}$, as the asymptotic distribution of the latter is improper Gaussian, whereas the DFT at other frequencies is proper complex Gaussian. Define

$$\boldsymbol{D} = [\boldsymbol{d}_z(f_1)\, \boldsymbol{d}_z(f_2)\, \cdots\, \boldsymbol{d}_z(f_{(n/2)-1})]^H \in \mathbb{C}^{((n/2)-1)\times(2p)}. \tag{14}$$

The asymptotic joint probability density function (pdf) of $\boldsymbol{D}$ is given by

$$f_{\boldsymbol{D}}(\boldsymbol{D}) = \prod_{k=1}^{(n/2)-1} \frac{\exp\left(-\boldsymbol{d}_z^H(f_k)\boldsymbol{S}_z^{-1}(f_k)\boldsymbol{d}_z(f_k)\right)}{\pi^{2p}\,|\boldsymbol{S}_z(f_k)|}. \tag{15}$$

The PSD matrix $\boldsymbol{S}_z(f_k)$ is unknown. Under $\mathcal{H}_0$, the ℓth component $y_\ell(t)$ of $\boldsymbol{y}(t)$ is independent of $y_m(t)$ for $\ell \ne m$. Let ($\ell = 1, 2, \cdots, p$)

$$\boldsymbol{S}_z^{(\ell)}(f) := \begin{bmatrix} [\boldsymbol{S}_z(f)]_{\ell\ell} & [\boldsymbol{S}_z(f)]_{\ell(\ell+p)} \\ [\boldsymbol{S}_z(f)]_{(\ell+p)\ell} & [\boldsymbol{S}_z(f)]_{(\ell+p)(\ell+p)} \end{bmatrix}. \tag{16}$$

Then in terms of $\boldsymbol{S}_y$ and $\tilde{\boldsymbol{S}}_y$,

$$\boldsymbol{S}_z^{(\ell)}(f) = \begin{bmatrix} [\boldsymbol{S}_y(f)]_{\ell\ell} & [\tilde{\boldsymbol{S}}_y(f)]_{\ell\ell} \\ [\tilde{\boldsymbol{S}}_y^*(-f)]_{\ell\ell} & [\boldsymbol{S}_y^*(-f)]_{\ell\ell} \end{bmatrix}. \tag{17}$$

Under $\mathcal{H}_0$, all entries in $\boldsymbol{S}_z(f)$ are zeros except for those in $\boldsymbol{S}_z^{(\ell)}(f)$, $\ell = 1, 2, \cdots, p$. Under $\mathcal{H}_1$, $\boldsymbol{y}(t)$ is improper with $\boldsymbol{S}_z(f) \succ 0$ with no specific structure. Testing for the presence of an improper common signal in spatially independent improper noise is then reformulated as the problem

$$\begin{aligned} \mathcal{H}_0 &: \boldsymbol{S}_z(f_k) \text{ is such that } \boldsymbol{S}_y(\pm f_k) \text{ and } \tilde{\boldsymbol{S}}_y(\pm f_k) \\ &\quad \text{are diagonal } \forall k \in [1, (n/2)-1] \\ \mathcal{H}_1 &: \boldsymbol{S}_z(f_k) \succ \boldsymbol{0} \text{ with no specific structure} \\ &\quad \forall k \in [1, (n/2)-1]. \end{aligned} \tag{18}$$

By (13), (16) and (17), the above constraint under $\mathcal{H}_0$ is equivalent to $[\boldsymbol{S}_z(f_k)]_{\ell m} = 0\ \forall l, m$, except for the entries in $\boldsymbol{S}_z^{(q)}(f_k)$, $q = 1, 2, \cdots, p$, $\forall k \in [1, (n/2)-1]$.

Under assumption (A1), $\boldsymbol{S}_y(f_k)$ is constant over $K = 2m_t + 1 \ge 2p$ consecutive frequency points, where we define $m_t = (K-1)/2$ with K as odd. Define

$$\tilde{f}_m = \frac{(m-1)K + m_t + 1}{n}, \qquad m = 1, 2, \cdots, M, \tag{19}$$

$$M = \left\lfloor \frac{\frac{n}{2} - m_t - 1}{K} \right\rfloor, \tag{20}$$

leading to M equally spaced frequencies $\tilde{f}_m$ in the interval $(0, 0.5)$, at intervals of K/n. Using assumptions (A1)-(A3), the GLRT for (18) is derived by [14] as

$$\mathcal{L} := \frac{\sup_{\substack{S_z(\tilde{f}_k) \\ \text{under } \mathcal{H}_1}} f_{\boldsymbol{D}|\mathcal{H}_1}(D|\mathcal{H}_1)}{\sup_{\substack{S_z(\tilde{f}_k) \\ \text{under } \mathcal{H}_0}} f_{\boldsymbol{D}|\mathcal{H}_0}(D|\mathcal{H}_0)} \tag{21}$$

$$= \prod_{m=1}^{M} \frac{\prod_{q=1}^{p} |\hat{\boldsymbol{S}}_z^{(q)}(\tilde{f}_m)|^K}{|\hat{\boldsymbol{S}}_z(\tilde{f}_m)|^K} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \tau_1 \tag{22}$$

where the threshold τ_1 is picked to achieve a pre-specified probability of false alarm $P_{fa} = P\{\mathcal{L} \ge \tau_1 \,|\, \mathcal{H}_0\}$,

$$\hat{\boldsymbol{S}}_z(\tilde{f}_m) = \frac{1}{K} \sum_{l=-m_t}^{m_t} \boldsymbol{d}_z(\tilde{f}_{m,\ell}) \boldsymbol{d}_z^H(\tilde{f}_{m,\ell}), \tag{23}$$

$$\hat{\boldsymbol{S}}_z^{(q)}(\tilde{f}_m) = \frac{1}{K} \tilde{\boldsymbol{D}}^{(q)}(\tilde{f}_m) \in \mathbb{C}^{2\times 2}, \tag{24}$$

the (i_1, i_2)th component of $\tilde{\boldsymbol{D}}^{(q)}(\tilde{f}_m)$ is given by

$$\tilde{D}_{i_1 i_2}^{(q)}(\tilde{f}_m) = \sum_{l=-m_t}^{m_t} [\boldsymbol{d}_z(\tilde{f}_{m,\ell})]_{i_1} [\boldsymbol{d}_z(\tilde{f}_{m,\ell})]_{i_2}^* \tag{25}$$

and

$$\tilde{f}_{m,\ell} = \frac{(m-1)K + m_t + 1 + \ell}{n}. \tag{26}$$

It is shown in [14, Appendix A] that $\mathcal{L} \ge 1$.

A. THRESHOLD SELECTION

Theorem 1 of [14] allows one to calculate the test threshold analytically. In the following, χ_n^2 denotes a random variable with central chi-square distribution with n degrees of freedom as well as the distribution itself. Also, $B_r(x)$ denotes the

Bernoulli polynomial of degree r and order unity. The first five Bernoulli polynomials are ($B_0(x) = 1$):

$$B_1(x) = x - (1/2), \quad B_2(x) = x^2 - x + (1/6),$$
$$B_3(x) = x^3 - (3/2)x^2 + (1/2)x,$$
$$B_4(x) = x^4 - 2\,x^3 + x^2 - (1/30),$$
$$B_5(x) = x^5 - (5/2)x^4 + (5/3)x^3 - (1/6)x.$$

Theorem 1: The GLRT for (18) is given by

$$2\rho\ \ln(\mathcal{L}) \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \tau_2$$

where

$$\rho = 1 - \frac{2(p+1)}{3K}, \tag{27}$$

$$\ln(\mathcal{L}) = K \sum_{m=1}^{M} \left(\left[\sum_{q=1}^{p} \ln |\hat{\boldsymbol{S}}_z^{(q)}(\tilde{f}_m)| \right] - \ln(|\hat{\boldsymbol{S}}_z(\tilde{f}_m)|) \right). \tag{28}$$

For any $x > 0$, the probability $P\{2\rho\ \ln(\mathcal{L}) \le x \,|\, \mathcal{H}_0\}$ is given by

$$\begin{aligned} &P\{2\rho\ \ln(\mathcal{L}) \le x \,|\, \mathcal{H}_0\} \\ &= P\{\chi^2_\nu \le x\} + \omega_2 \Big[P\{\chi^2_{\nu+4} \le x\} \\ &\quad - P\{\chi^2_\nu \le x\} \Big] + \omega_3 \Big[P\{\chi^2_{\nu+6} \le x\} - P\{\chi^2_\nu \le x\} \Big] \\ &\quad + \Big\{ \omega_4 \Big[P\{\chi^2_{\nu+8} \le x\} - P\{\chi^2_\nu \le x\} \Big] + \frac{1}{2}\omega_2^2 \Big[P\{\chi^2_{\nu+8} \le x\} \\ &\quad - 2P\{\chi^2_{\nu+4} \le x\} + P\{\chi^2_\nu \le x\} \Big] \Big\} + \mathcal{O}(M/K^5) \end{aligned} \tag{29}$$

where

$$\nu = 4Mp(p-1) \tag{30}$$

and

$$\omega_r = \frac{(-1)^{r+1} M}{r(r+1)(\rho K)^r} \left\{ \left(\sum_{l=1}^{2p} B_{r+1}((1-\rho)K + 1 - l) \right) - p \left(\sum_{l=1}^{2} B_{r+1}((1-\rho)K + 1 - l) \right) \right\}. \tag{31}$$

The threshold τ_2 is picked to achieve a pre-specified probability of false alarm $P_{fa} = 1 - P\{2\rho\ \ln(\mathcal{L}) \le \tau_2 \,|\, \mathcal{H}_0\}$ •

For further details, one is referred to [14]

IV. ROBUSTIFICATION VIA DATA CLEANING

The GLRT of Sec. III is based on clean data (i.e., no outliers). We now turn to robustifying this GLRT against additive outliers described in Sec. II-B.1.

The literature on robust signal processing and robust statistics is quite extensive ([16]–[18], [20], [26] and references therein). The robustification approach discussed in this section is based on robust estimation of multivariate covariance (scatter) of data, and it follows a similar approach used in [21] for robust comparison of multivariate proper complex-valued random signals. The basic idea is to use some robust methods for identifying outliers in the data, and then to clean the data by removing the outliers. Data cleaning via dynamic modeling of the underlying signals (time series) has a long history; see, e.g., [16], [26, Chapter 8], and references therein. If one has a robust, dynamic model of the multivariate time series, one can use robust Kalman filtering methods for identification of outliers based on the residuals (innovations or prediction errors), and their substitution with the predicted values [18], [27]. An advantage of such methods (compared to the approach followed in this paper) is that in addition to identifying large outliers, "small" outliers inconsistent with the prediction errors can also be identified. Robust dynamic modeling (autoregressive, autoregressive moving average, or state-space models) of vector time series is significantly more complicated and computationally demanding than the relatively simple robust estimation of multivariate scatter. Auxiliary considerations such as model order determination arise. Our attempt to use [26] to robustly fit vector autoregressive models and to clean contaminated data based on the fitted model (as noted on [26, p. 72]) has been unsuccessful for the simulation examples presented later in Sec. V. Therefore, we use robust scatter estimator based approach to outlier identification, which can identify only "large" outliers.

In this paper, we follow the approach of [21] with some enhancements such as using vector median filtering for outlier removal compared to component-wise (scalar) median filtering used in [21], and investigating a more (statistically) efficient MM-estimator for multivariate scatter estimation (see Sec. IV-A.2) in addition to the S-estimator (Sec. IV-A.1) used in [21]. Also, while [21] deals only with proper complex-valued signals in proper outliers, in this paper we additionally consider improper outliers also.

A. ROBUST ESTIMATION OF MULTIVARIATE SCATTER

There exist quite a few methods for robust estimation of multivariate scatter [16]–[18], [20], [26]. We will use the computationally efficient algorithms DetS for S-estimator, and DetMM for MM-estimator, of [20]. These approaches are described below.

1) S-ESTIMATOR

Given a sample $\boldsymbol{x}_0,\ \boldsymbol{x}_2,\ \cdots,\ \boldsymbol{x}_{n-1}$ of size n, of a real-valued random vector $\boldsymbol{x} \in \mathbb{R}^q$, a multivariate S-estimator of location (mean) $\boldsymbol{m}$ and scatter (covariance matrix) $\boldsymbol{S}$ is defined as the couple $(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}})$ which solves the following problem:

$$(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}}) = \arg \left\{ \min_{\boldsymbol{m}, \boldsymbol{S}} |\boldsymbol{S}| \right\} \tag{32}$$

subject to

$$\frac{1}{n} \sum_{t=0}^{n-1} \rho_c \left(d(\boldsymbol{x}_t, \boldsymbol{m}, \boldsymbol{S}) \right) = b \tag{33}$$

where the constant $b \in (0, 1)$ influences the "breakdown" value of the estimator under the nominal (Gaussian)

model [16], [20],

$$d(\boldsymbol{x}_t, \boldsymbol{m}, \boldsymbol{S}) = \sqrt{(\boldsymbol{x}_t - \boldsymbol{m})^T \boldsymbol{S}^{-1} (\boldsymbol{x}_t - \boldsymbol{m})} \tag{34}$$

is the Mahalanobis distance, $\boldsymbol{m} \in \mathbb{R}^q$, $\boldsymbol{S}$ is a $q \times q$ symmetric positive definite matrix, and ρ_c is a smooth bounded loss function (''$\rho-$function'') with a tuning parameter $c > 0$ that satisfies the following two conditions:

(C1) $\rho_c(d)$ is real, symmetric and twice continuously differentiable function, with $\rho_c(0) = 0$.

(C2) $\rho_c(d)$ is strictly increasing on an interval $d \in [0, c]$ and constant on $d \in [c, \infty)$.

Thus, an S-estimator of multivariate location and scale minimizes the determinant of the covariance matrix, subject to a constraint on the magnitudes of the corresponding Mahalanobis distances. The loss function is typically picked to be Tukey's bisquare $\rho-$function given by

$$\rho_c(d) = \begin{cases} \dfrac{d^2}{2} - \dfrac{d^4}{2c^2} + \dfrac{d^6}{4c^4} & \text{for } |d| \le c \\ \dfrac{c^2}{6} & \text{for } |d| > c \end{cases} \tag{35}$$

where c is a tuning constant that, together with b in (33), influences the breakdown point (BP). The BP of an estimator characterizes its quantitative robustness. It is the maximal fraction of outliers in the observations which an estimate can handle without breaking down [18]. The BP takes values between 0 and 50%, with higher value indicating larger quantitative robustness. The BP of a multivariate S-estimator is $b/\rho_c(c)$ [20]. If the nominal (i.e., outlier free) model for the data generates Gaussian vectors, then [28]

$$b = \frac{q}{2}\bar{\chi}^2_{q+2}(c^2) - \frac{q(q+2)}{2c^2}\bar{\chi}^2_{q+4}(c^2) + \frac{q(q+2)(q+4)}{6c^4}\bar{\chi}^2_{q+6}(c^2) + \frac{c^2}{6}(1 - \bar{\chi}^2_q(c^2)) \tag{36}$$

where $\bar{\chi}^2_\nu$ denotes the cumulative distribution function (CDF) of a χ^2_ν random variable. For a given BP between 0 and 50%, one can derive the value of the corresponding tuning parameter c in (35) [20].

Tukey's bisquare $\rho-$function is a non-convex function, hence, to obtain a global minimum that satisfies (33), random subsampling methods are used to obtain a good initial guess [20]. The end result is a robust scatter estimate $\hat{\boldsymbol{\Sigma}}$ of $\boldsymbol{\Sigma}$ that is ''close'' to the true value under no outliers, but is ''robust'' to outliers. Software in MATLAB for computationally efficient algorithm DetS of [20] is available from the author's website (`http://wis.kuleuven.be/stat/robust/LIBRA`), and was used in our simulations reported in Sec. V with the default BP $b/\rho_c(c) = 0.5$.

2) MM-ESTIMATOR

In addition to the BP value (a measure of robustness), a measure of performance of an estimator is its asymptotic (statistical) efficiency (efficiency, in short) [16]. For an estimator of a scalar parameter, efficiency equals the asymptotic (as sample size $n \to \infty$) variance of the maximum likelihood estimator divided by the asymptotic variance of the estimator under consideration [16, Sec. 3.4]. For multidimensional parameters, the definition of efficiency is more involved [16, Sec. 3.6]. The S-estimators have good robustness properties but are not very efficient [20], [29]. For instance, for the nominal Gaussian model (no outliers and $\boldsymbol{x} \sim \mathcal{N}_q(\boldsymbol{m}, \sigma^2 \boldsymbol{I}_q)$), the diagonal element of the S-estimator of the scatter matrix with BP of 50% has an efficiency of 50.2% for $q = 2$ and 92% for $q = 10$ [19]. The multivariate MM-estimator, introduced by [30], yields a more efficient estimator of multivariate scatter while retaining the BP value of the S-estimator.

The multivariate MM-estimator $(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}})$ of $(\boldsymbol{m}, \boldsymbol{S})$, given sample $\boldsymbol{x}_0, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_{n-1}$ of size n, is computed in two steps using two different loss functions $\rho_{c0}(d)$ and $\rho_{c1}(d)$ [19], [20]:

(S1) As in (32)-(34), compute the multivariate S-estimator $(\tilde{\boldsymbol{\mu}}, \tilde{\boldsymbol{\Sigma}})$ for some $\rho-$function $\rho_{c0}(d)$. Set $\tilde{\sigma} = |\tilde{\boldsymbol{\Sigma}}|^{1/(2q)}$.

(S2) The multivariate MM-estimator of location (mean) $\boldsymbol{m}$ and shape matrix $\boldsymbol{G}$ is defined as the couple $(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Gamma}})$ which solves the following problem:

$$(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Gamma}}) = \arg \left\{ \min_{\boldsymbol{m}, \boldsymbol{G}} \ \beta(\boldsymbol{x}, \boldsymbol{m}, \boldsymbol{G}) \right\} \tag{37}$$

where

$$\beta(\boldsymbol{x}, \boldsymbol{m}, \boldsymbol{G}) = \frac{1}{n}\sum_{t=0}^{n-1} \rho_{c1}\left(\frac{\sqrt{(\boldsymbol{x}_t - \boldsymbol{m})^T \boldsymbol{G}^{-1} (\boldsymbol{x}_t - \boldsymbol{m})}}{\tilde{\sigma}} \right), \tag{38}$$

$\boldsymbol{m} \in \mathbb{R}^q$, and $\boldsymbol{G}$ is a $q \times q$ symmetric positive definite shape matrix with $|\boldsymbol{G}| = 1$. The MM-estimator of location is $\hat{\boldsymbol{\mu}}$ and the MM-estimator of scatter matrix is given by $\hat{\boldsymbol{\Sigma}} = \tilde{\sigma}^2 \hat{\boldsymbol{\Gamma}}$.

The robust S-estimator is used to estimate the scale parameter $\tilde{\sigma}$ where the loss function $\rho_{c0}(d)$ is chosen to yield a high BP. Then location and shape $\boldsymbol{G}$ parameters are estimated using a different loss function $\rho_{c1}(d)$ designed to yield a higher efficiency. If $\rho_{c0}(d) = \rho_{c1}(d)$, then the MM-estimator equals the initial S-estimator. We used the Tukey's bisquare $\rho-$function given by (35), and the two functions $\rho_{c0}(d)$ and $\rho_{c1}(d)$ differ in the choice of the tuning parameter c.

The minimization in (37) is done by iteratively reweighted least squares steps starting from the initialization $(\tilde{\boldsymbol{\mu}}, \tilde{\sigma}^{-2}\tilde{\boldsymbol{\Sigma}})$. Software in MATLAB for computationally efficient algorithm DetMM of [20] is available from the author's website (`http://wis.kuleuven.be/stat/robust/CODE`), and was used with the default desired efficiency of 0.95 (for the Gaussian nominal model) in our simulations reported in Sec. V. The DetMM algorithm uses the DetS algorithm in step (S1), and it was used in our simulations reported in Sec. V with the default BP $b/\rho_c(c) = 0.5$.

B. OUTLIER DETECTION USING ROBUST SCATTER ESTIMATOR

Now we exploit the theory and results of Sec. IV-A to design an outlier detector to identify outliers in $\tilde{\boldsymbol{y}}(t)$, given by (2) and (5). Since $\tilde{\boldsymbol{y}}(t)$ is complex-valued, in order to exploit the results of Sec. IV-A, we will represent $\tilde{\boldsymbol{y}}(t)$ as two real-valued random vectors $\mathrm{Re}\{\tilde{\boldsymbol{y}}(t)\}$ and $\mathrm{Im}\{\tilde{\boldsymbol{y}}(t)\}$. We have

$$\boldsymbol{x}_t = \begin{bmatrix} \mathrm{Re}\{\tilde{\boldsymbol{y}}(t)\} \\ \mathrm{Im}\{\tilde{\boldsymbol{y}}(t)\} \end{bmatrix} \in \mathbb{R}^{2p} \tag{39}$$

and $q = 2p$. Now apply DetS or DetMM algorithm of [20] to $\boldsymbol{x}_t$, $t = 0, 1, \cdots, n-1$ to obtain $(2p) \times (2p)$ robust scatter S-estimate or MM-estimate, respectively, both denoted by $\hat{\boldsymbol{\Sigma}}$.

Having obtained outlier-resistant estimate of the covariance matrix of $\{\tilde{\boldsymbol{y}}(t)\}$, we now consider detection of outliers. Note that, by assumption, clean data are zero-mean. Hence, our nominal model for the data is $\boldsymbol{x}_t \sim \mathcal{N}_r(\boldsymbol{0}, \hat{\boldsymbol{\Sigma}})$. Define

$$w_t = \boldsymbol{x}_t^\top \hat{\boldsymbol{\Sigma}}^{-1} \boldsymbol{x}_t \,. \tag{40}$$

Under the nominal Gaussian model, $w_t \sim \chi^2_{2p}$. For a significance level (probability of false alarm) α, we set the threshold τ_3 at $\tau_3 = (\chi^2_{2p})^{-1}(1-\alpha)$, i.e., $1-\alpha = \bar{\chi}^2_{2p}(\tau_3)$, and declare $\tilde{\boldsymbol{y}}(t)$ to be an outlier if the corresponding $w_t > \tau_3$. For our simulations, we picked $\alpha = 0.025$. This method is robust since the estimate of the covariance matrix is robust.

C. DATA CLEANING VIA VECTOR MEDIAN FILTERING

Once an outlier is detected, it has to be removed and replaced. Since the outlier is a vector, we first replace it with the vector median over a 5-point window (see Sec. IV-C.1 for vector medians), centered at the location of the detected outlier unless one is at the edge of the data block, in which case the window is "unbalanced." In the unbalanced case, we still consider a 5-point window which includes the detected outlier and four additional contiguous data points. This is done for $\tilde{\boldsymbol{y}}(t)$, at $t = 0, 1, \cdots, n-1$, one sample at a time. Obviously, if at some time t_0, there is no outlier detected, we keep that data point unaltered.

1) VECTOR MEDIAN FILTERING

The need for vector median filtering to replace scalar median filtering was initially recognized in color image processing [31]–[33]. For vector-valued signals, component-by-component scalar median filtering fails to exploit "interchannel" correlation. In this paper, we follow [31] and [33, Sec. 2.1], using the Euclidean distance function. Let $\boldsymbol{x}_t \in \mathbb{R}^q$ be given over a window W of T samples, $W = \{\boldsymbol{x}_t : t = 1, 2, \cdots, T\}$. In [31], the median $\boldsymbol{x}_{\mathrm{med}}$ of the set W is defined as

$$\boldsymbol{x}_{\mathrm{med}} = \arg\left\{ \min_{\boldsymbol{x}_k \in W} \sum_{t=1}^{T} \|\boldsymbol{x}_t - \boldsymbol{x}_k\|_2 \right\} \tag{41}$$

where $\|\boldsymbol{x}_t\|_2 = \sqrt{\sum_{i=1}^{q} x_{ti}^2}$, the Euclidean norm of $\boldsymbol{x}_t$. For scalar $\boldsymbol{x}_t$s, this definition reduces to scalar median.

An efficient computational method may be found in [33]. Define the T scalars ($k = 1, 2, \cdots, T$)

$$\xi_k(\boldsymbol{x}_k) = \sum_{t=1}^{T} \|\boldsymbol{x}_t - \boldsymbol{x}_k\|_2 \,. \tag{42}$$

The scalars $\xi_1, \xi_2, \cdots, \xi_T$ are ranked in the order of their values, and the associated vectors are then corresponding ordered as follows:

$$\xi_{(1)} \le \xi_{(2)} \le \cdots \le \xi_{(T)} \,, \tag{43}$$

$$\boldsymbol{x}_{(1)}(\xi_{(1)}) \le \boldsymbol{x}_{(2)}(\xi_{(2)}) \le \cdots \le \boldsymbol{x}_{(T)}(\xi_{(T)}) \,, \tag{44}$$

where $\xi_{(i)} \in \{\xi_1, \xi_2, \cdots, \xi_T\}$ and $\boldsymbol{x}_{(i)} \in W$ for $i = 1, 2, \cdots, T$. That is, ordering of ξ_is induces similar ordering of $\boldsymbol{x}_i$s. Then

$$\boldsymbol{x}_{\mathrm{med}} = \boldsymbol{x}_{(1)} = \boldsymbol{x}_{(1)}(\xi_{(1)}) \,. \tag{45}$$

2) CLIPPING

In the final step, the selectively vector median filtered data is again checked for outliers to guard against "patchy" outliers. Any outlier detected at this stage is clipped (i.e., scaled with a positive scalar) to yield $\boldsymbol{x}_t^\top \hat{\boldsymbol{\Sigma}}^{-1} \boldsymbol{x}_t = \tau_3$, where τ_3 is the threshold used in Sec. IV-B. The sequence obtained at the end of this stage is our cleaned $\tilde{\boldsymbol{y}}(t)$, labeled $\check{\boldsymbol{y}}(t)$. While the vector median filtered sequences in the first step are well-motivated, the clipping step (if needed) is heuristic.

D. ROBUST MULTISENSOR DETECTION

Having cleaned the contaminated vector measurement sequence $\{\tilde{\boldsymbol{y}}(t)\}_{t=0}^{n-1}$ to yield $\{\check{\boldsymbol{y}}(t)\}_{t=0}^{n-1}$, we now use the latter in Theorem 1 of Sec. III for multisensor detection of improper signals.

Our proposed robust detection method is summarized below.

(i) Obtain robust scatter S-estimate or MM-estimate $\hat{\boldsymbol{\Sigma}}$ using the DetS or DetMM algorithm of [20], as discussed in Secs. IV-A.
(ii) Detect outliers in $\{\tilde{\boldsymbol{y}}(t)\}$ using $\hat{\boldsymbol{\Sigma}}$, as discussed in Sec. IV-B.
(iii) Clean $\{\tilde{\boldsymbol{y}}(t)\}$ of the detected outliers by a combination of selective vector median filtering and clipping, as discussed in Sec. IV-C. Denote the cleaned sequence as $\check{\boldsymbol{y}}(t)$.
(iv) Now perform multisensor detection of improper signals via the PSD-based approach discussed in Sec. III, operating on the cleaned sequence $\{\check{\boldsymbol{y}}(t)\}$.

V. SIMULATION EXAMPLES

We now present some computer simulation examples to illustrate the proposed data-cleaning approach to robust multisensor detection.

A. SIGNAL AND NOISE MODELS

These are as in [14]. We generate stationary $\boldsymbol{y}(t) \in \mathbb{C}^p$ as a noisy improper signal given by

$$\boldsymbol{y}(t) = \boldsymbol{s}(t) + \boldsymbol{v}(t) \,, \tag{46}$$

where $\{\boldsymbol{v}(t)\}$ is spatially uncorrelated, colored, improper complex Gaussian noise, and $\{\boldsymbol{s}(t)\}$ is the signal sequence. The noise sequence $\{\boldsymbol{v}(t)\} \in \mathbb{C}^p$ is generated as

$$\boldsymbol{v}(t) = \boldsymbol{v}_c(t) + \boldsymbol{v}_w(t)\,, \tag{47}$$

where $\boldsymbol{v}_w(t) \sim \mathcal{N}_c(\boldsymbol{0}, \sigma_w^2 \boldsymbol{I}_p)$ is i.i.d., and $\boldsymbol{v}_c(t)$ is generated as follows. Let $\tilde{\boldsymbol{v}}_i(t)$, $i \in [1, p]$, denote p independent zero-mean white proper Gaussian sequences. We generate $[\boldsymbol{v}_c]_i(t) = a_I h_I(t) \circledast \tilde{\boldsymbol{v}}_i(t) + a_Q h_Q(t) \circledast \tilde{\boldsymbol{v}}_i^*(t)$ where $a_I = a_Q^* = (1 + j1)/\sqrt{2}$, $\circledast$ denotes convolution, $h_I(t) = [0.3 \;\; 1 \;\; 0.3]$, and $h_Q(t) = [0.4 \;\; 1 \;\; 0.5]$. Thus noise $\boldsymbol{v}_c(t)$, hence $\boldsymbol{v}(t)$, is spatially independent, improper Gaussian. We pick $\sigma_w^2 = 0.2\mathbb{E}\{\|\boldsymbol{v}(t)\|^2\}/p$ and $\mathbb{E}\{\|\boldsymbol{v}_c(t)\|^2\}/p = 0.8\mathbb{E}\{\|\boldsymbol{v}(t)\|^2\}/p$.

The signal $\{\boldsymbol{s}(t)\}$ is a filtered digital communications signal generated by passing an information sequence through a frequency-selective Rayleigh fading channel as follows:

$$\boldsymbol{s}(t) = \sum_{l=0}^{4} \boldsymbol{h}(l) d(t-l) \tag{48}$$

where $d(t)$ is a scalar i.i.d. BPSK information sequence, with equally likely binary values $d(t) \in \{-1, +1\}$, filtered through a random time-invariant, frequency-selective Rayleigh fading channel $\boldsymbol{h}(l) \in \mathbb{C}^p$ with 5 taps, equal power delay profile, mutually independent components, which are identically distributed zero-mean proper complex Gaussian random variables. For different ls, $\boldsymbol{h}\,(l)$s are mutually independent and identically distributed as $\boldsymbol{h}(l) \sim \mathcal{N}_c(\boldsymbol{0}, \sigma_h^2 \boldsymbol{I})$. Since $d(t)$ is a BPSK signal, it is improper. Since any linear filtering preserves propriety/impropriety property [1], we have an improper $\{\boldsymbol{s}(t)\}$ for BPSK $d(t)$.

We pick σ_h^2 to achieve the desired average signal-to-noise ratio (SNR) across p components (receive antennas in the communications context), defined as ratio of the sum of signal powers at the p antennas to the sum of noise powers:

$$\text{SNR} = E\{\|\boldsymbol{s}(t)\|^2\}/E\{\|\boldsymbol{v}(t)\|^2\}\,.$$

B. ADDITIVE OUTLIERS

These are added as described in Sec. II-B.1, with $\sigma_{vy}^2 = 100\,E\{\|\boldsymbol{y}(t)\|^2\}/p$. Both proper and improper outliers were added with the probability of outlier addition $p_o = 0.1$ or 0.3 (10% or 30% outliers).

C. THRESHOLD CALCULATION

We first investigate the efficacy of data-cleaning in conjunction with Theorem 1 in computing the GLRT threshold for a given probability of false alarm P_{fa}. After cleaning the contaminated data of the outliers, is Theorem 1 still effective in yielding the design false-alarm rate P_{fa}? In order to test this, we set $\boldsymbol{y}(t) = \boldsymbol{v}(t)$, just noise, given by (47), where $\boldsymbol{v}(t)$ is improper. Both proper and improper outliers were added to the data, following Secs. II-B.1 and V-B. We consider up to two or four antennas ($p = 2$ or 4). We used unweighted smoothing in the frequency-domain to estimate PSD (see (23)). To estimate the PSD of augmented complex-valued $\boldsymbol{z}(t)$, for $n = 256$, we choose $m_t = 7$ leading to

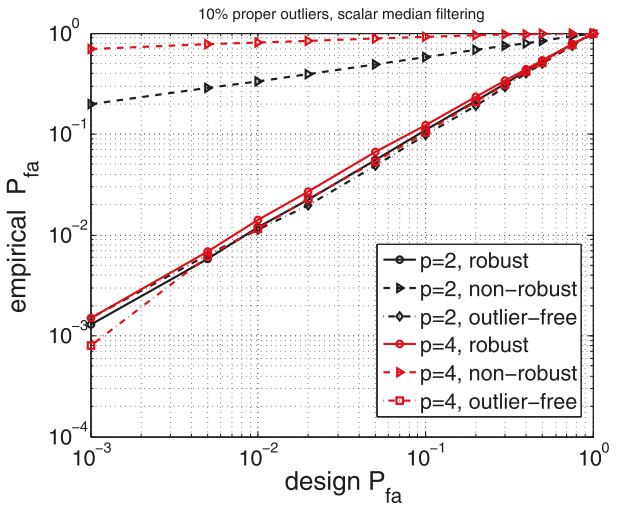

FIGURE 1. Empirical P_{fa} vs. design P_{fa} using the GLRT detector of Theorem 1, operating after data-cleaning (labeled "robust"), or on outlier-free data (labeled "outlier-free"), or on contaminated data (labeled "non-robust"): 10% proper outliers ($p_o = 0.1$), scalar median filtering used for cleaning, $n = 256$, $K = 15$, $M = 8$.

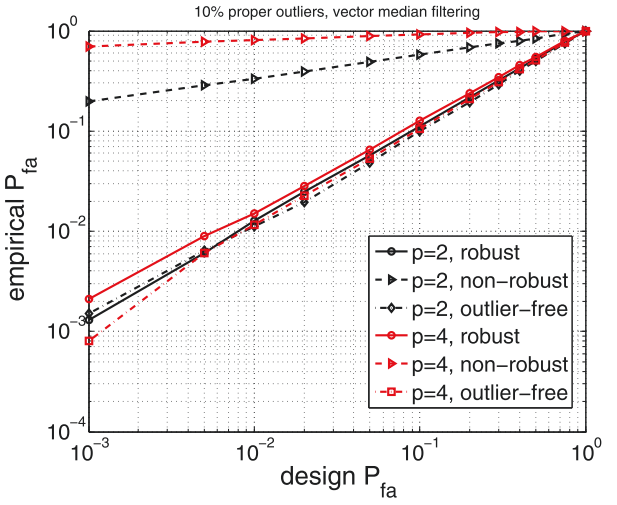

FIGURE 2. Empirical P_{fa} vs. design P_{fa} using the GLRT detector of Theorem 1, operating after data-cleaning (labeled "robust"), or on outlier-free data (labeled "outlier-free"), or on contaminated data (labeled "non-robust"): 10% proper outliers ($p_o = 0.1$), vector median filtering used for cleaning, $n = 256$, $K = 15$, $M = 8$.

$K = 15$ and $M = 8$. In Figs. 1-4, we compare the empirical P_{fa} and design P_{fa} based on 10,000 runs for $n = 256$. The threshold values were calculated based on (29). Figs. 1-2 show the results for 10% proper outliers whereas Figs. 3-4 show the results for 10% improper outliers. In Figs. 1 and 3 scalar median filtering was used over a 5-point window around a detected outlier to clean data, whereas in Figs. 2 and 4 vector median filtering was used. In these figures we also show the results for the case when there are no outliers (curves labeled ''outlier-free''), and for the case when the GLRT of Theorem 1 was applied directly to contaminated data with no data-cleaning (curves labeled ''non-robust''). Robust estimation of data scatter was done

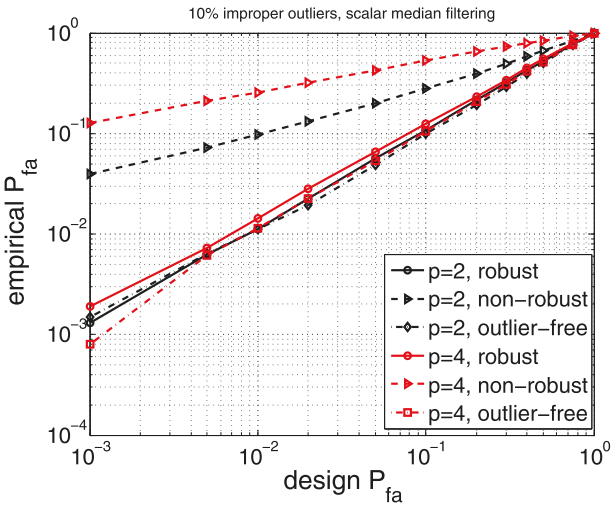

FIGURE 3. Empirical P_{fa} vs. design P_{fa} using the GLRT detector of Theorem 1, operating after data-cleaning (labeled "robust"), or on outlier-free data (labeled "outlier-free"), or on contaminated data (labeled "non-robust"): 10% proper outliers ($p_o = 0.1$), scalar median filtering used for cleaning, $n = 256$, $K = 15$, $M = 8$.

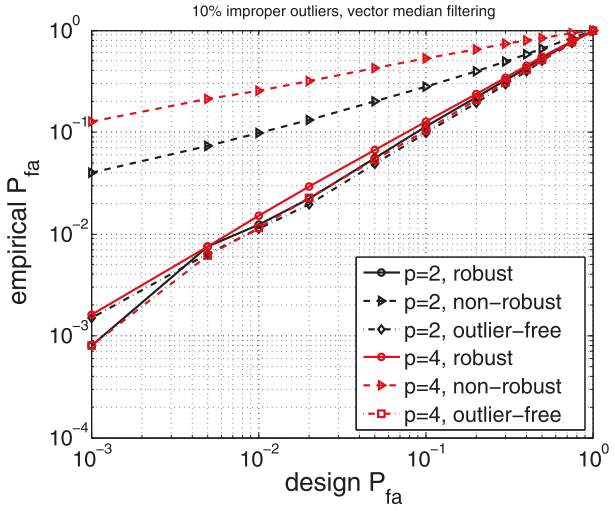

FIGURE 4. Empirical P_{fa} vs. design P_{fa} using the GLRT detector of Theorem 1, operating after data-cleaning (labeled "robust"), or on outlier-free data (labeled "outlier-free"), or on contaminated data (labeled "non-robust"): 10% proper outliers ($p_o = 0.1$), vector median filtering used for cleaning, $n = 256$, $K = 15$, $M = 8$.

using the S-estimator; the results for MM-estimator were the same as the S-estimator, and are not shown.

It is seen from Figs. 1-4 that Theorem 1, when used with cleaned data, is effective in accurately calculating the threshold value for $n = 256$, for all values of p used, as the results with cleaned data are close to that for the outlier-free data. Both scalar and vector median filtering yield similar results. When outliers are ignored (non-robust processing), their effect on the detector is quite deleterious resulting in significantly higher empirical P_{fa} compared to the design P_{fa}; that is, contaminated noise is perceived as signal.

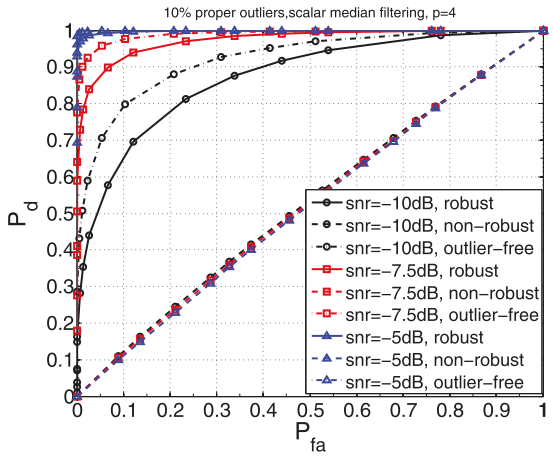

FIGURE 5. ROC curves using the GLRT detector of Theorem 1, operating after data-cleaning (labeled "robust"), or on outlier-free data (labeled "outlier-free"), or on contaminated data (labeled "non-robust"): 10% proper outliers ($p_o = 0.1$), scalar median filtering used for data-cleaning, $p = 4$, $n = 256$, $K = 15$, $M = 8$.

D. DETECTION PERFORMANCE

Next, we show the receiver operating characteristic (ROC) curves (probability of detection P_d versus P_{fa}) to illustrate the detection performance, based on 10,000 runs. In this case, we set $\boldsymbol{y}(t) = \boldsymbol{s}(t) + \boldsymbol{v}(t)$, given by (46), where $\boldsymbol{v}(t)$ and $\boldsymbol{s}(t)$ are as in (47) and (48), respectively, with improper $\boldsymbol{v}_c(t)$, $d(t)$ is BPSK, and and number of sensors (antennas) $p = 4$. Thus, both signal and noise are improper. Both proper and improper outliers were added to the data, following Secs. II-B.1 and V-B.

The empirical probability of detection P_d versus empirical false-alarm rate P_{fa} results, based on 10,000 runs, are shown in Figs. 5-9 for three different SNR values (-10, -7.5 or -5dB), $p = 4$, $N = 256$, $K = 15$ and $M = 8$, for three different implementations (''robust,'' ''non-robust,'' ''outlier-free'') of the GLRT of Theorem 1 (as for Figs. 1-4). The threshold values for the GLRT were calculated based on (29). Fig. 5 shows the results for 10% proper outliers, Figs. 6-7 display the results for 30% proper outliers, and Figs. 8-9 present the results for 30% improper outliers. In Figs. 5, 6 and 8 scalar median filtering was used over a 5-point window around a detected outlier to clean data, whereas in Figs. 7 and 9 vector median filtering was used. In these figures we also show the results for the case when there are no outliers (curves labeled ''outlier-free''), and for the case when the GLRT of Theorem 1 was applied directly to contaminated data with no data-cleaning (curves labeled ''non-robust''). Robust estimation of data scatter was done using the S-estimator; the results for MM-estimator were the same as the S-estimator, and are not shown.

Vector median filtering did not show any improvement over scalar median filtering in the 10% outlier case shown in Fig. 5, therefore, the results for vector median filtering are not shown for the 10% case. But comparing Figs. 6 and 7 for 30% proper outliers and Figs. 8 and 9 for 30% improper outliers, it is seen that vector median filtering significantly

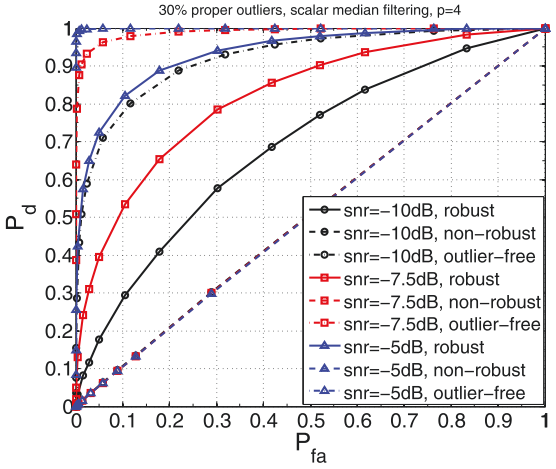

FIGURE 6. ROC curves using the GLRT detector of Theorem 1, operating after data-cleaning (labeled "robust"), or on outlier-free data (labeled "outlier-free"), or on contaminated data (labeled "non-robust"): 30% proper outliers ($p_o = 0.3$), scalar median filtering used for data-cleaning, $p = 4$, $n = 256$, $K = 15$, $M = 8$.

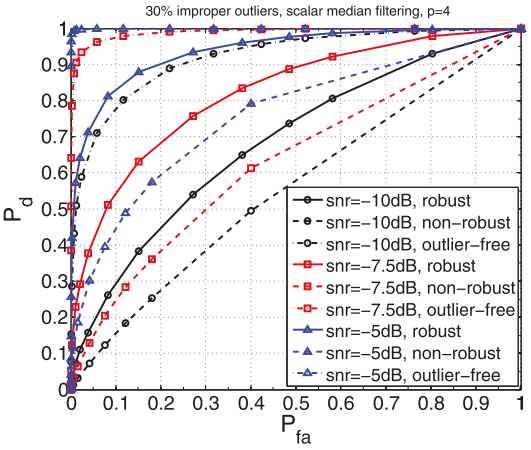

FIGURE 8. ROC curves using the GLRT detector of Theorem 1, operating after data-cleaning (labeled "robust"), or on outlier-free data (labeled "outlier-free"), or on contaminated data (labeled "non-robust"): 30% improper outliers ($p_o = 0.3$), scalar median filtering used for data-cleaning, $p = 4$, $n = 256$, $K = 15$, $M = 8$.

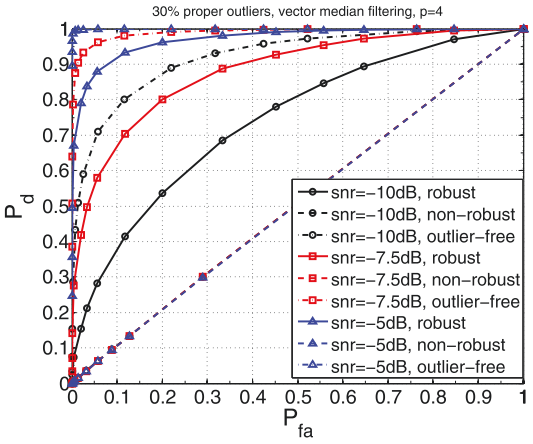

FIGURE 7. ROC curves using the GLRT detector of Theorem 1, operating after data-cleaning (labeled "robust"), or on outlier-free data (labeled "outlier-free"), or on contaminated data (labeled "non-robust"): 30% proper outliers ($p_o = 0.3$), vector median filtering used for data-cleaning, $p = 4$, $n = 256$, $K = 15$, $M = 8$.

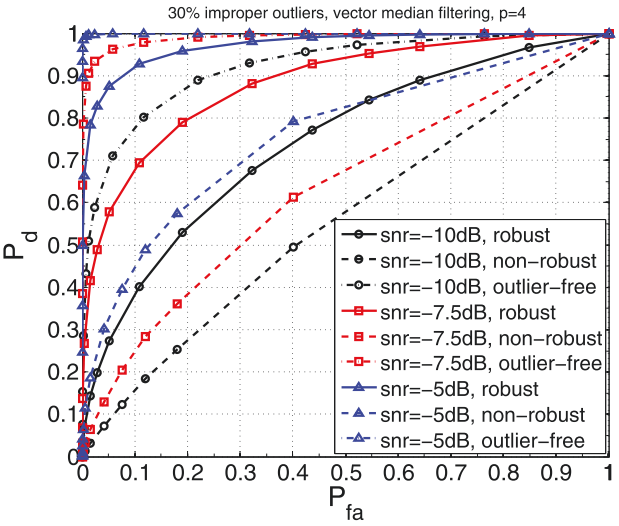

FIGURE 9. ROC curves using the GLRT detector of Theorem 1, operating after data-cleaning (labeled "robust"), or on outlier-free data (labeled "outlier-free"), or on contaminated data (labeled "non-robust"): 30% improper outliers ($p_o = 0.3$), vector median filtering used for data-cleaning, $p = 4$, $n = 256$, $K = 15$, $M = 8$.

improves the detection performance (higher P_d for a given P_{fa}) compared to scalar median filtering. For example, in Fig. 8 when using scalar median filtering in the robust detector, at SNR=-7.5 dB and $P_{fa} \approx 0.1$, $P_d \approx 0.53$, whereas in Fig. 9 when using vector median filtering in the robust detector, at SNR=-7.5 dB and $P_{fa} \approx 0.1$, $P_d \approx 0.68$. Recall that median filtering is used only in the robust detector.

It is seen from Figs. 5-9 that the robust detector significantly mitigates the effects of additive outliers (compare curves for the non-robust case with those for the robust case). In the presence of proper outliers (Figs. 5-7), the naive implementation of the GLRT of [14] (non-robust curves) results in a "total" failure of the test as the non-robust curves at all three SNR values overlay with $P_d \approx P_{fa}$. The robust detector performs much better than the non-robust detector in Figs. 5-7, although not as well as the outlier-free detector. Note that our approach can identify only "large" outliers, the "smaller" outliers remain and contribute to a deterioration of the effective SNR, hence, the robust detector performs similar to an outlier-free detector at a worse SNR. In Fig. 5 in the presence of 10% outliers, the robust detector at SNR=-5dB still outperforms the outlier-free detector at SNR=-7.5dB, and in Fig. 7 in the presence of 30% outliers, using vector median filtering the robust detector at SNR=-5dB significantly outperforms the outlier-free detector at SNR=-10dB.

In the presence of improper outliers, as shown in Figs. 8 and 9 for 30% improper outliers, the observations made regarding robust and outlier-free detectors earlier in the context of proper outliers, apply here as well. However, the non-robust detector interprets the improper outliers as improper signals and shows better performance than is the case for proper outliers. However, its performance

remains unsatisfactory. For instance, in Fig. 9 in the presence of 30% improper outliers, using vector median filtering the robust detector at SNR=-5dB significantly outperforms the outlier-free detector at SNR=-10dB, but the performance of the non-robust detector at SNR=-5dB is significantly worse than outlier-free detector at SNR=-10dB and is close to the performance of the robust detector at SNR=-10dB.

VI. CONCLUSION

We investigated the problem of detection of a common improper signal among two or more sensors, in the presence of spatially independent improper noise and additive outliers. A source of improper noise is IQ imbalance during down-conversion of bandpass noise to baseband. We exploited existing robust estimators of multivariate scatter to detect the outliers, and subsequently remove and replace them using vector median filtering to yield cleaned data. An existing GLRT of [14] was then applied to the cleaned data. The approach was illustrated via simulations involving 10% and 30% proper and improper outliers, and they show the efficacy of the proposed robustification.

REFERENCES

[1] P. J. Schreier and L. L. Scharf, *Statistical Signal Processing of Complex-Valued Data*. Cambridge, U.K.: Cambridge Univ. Press, 2010.
[2] F. D. Neeser and J. L. Massey, "Proper complex random processes with applications to information theory," *IEEE Trans. Inf. Theory*, vol. 39, no. 4, pp. 1293–1302, Jul. 1993.
[3] D. Tandur and M. Moonen, "Efficient compensation of transmitter and receiver IQ imbalance in OFDM systems," *EURASIP J. Adv. Signal Process.*, vol. 2010, Dec. 2010, Art. no. 106562.
[4] T. Adali, P. J. Schreier, and L. L. Scharf, "Complex-valued signal processing: The proper way to deal with impropriety," *IEEE Trans. Signal Process.*, vol. 59, no. 11, pp. 5101–5125, Nov. 2011.
[5] D. P. Mandic and V. S. L. Goh, *Complex Valued Nonlinear Adaptive Filters*. New York, NY, USA: Wiley, 2009.
[6] R. Zhang, T. J. Lim, Y.-C. Liang, and Y. Zeng, "Multi-antenna based spectrum sensing for cognitive radios: A GLRT approach," *IEEE Trans. Commun.*, vol. 58, no. 1, pp. 84–88, Jan. 2010.
[7] D. Ramírez, J. Vía, I. Santamaría, and L. Scharf, "Detection of spatially correlated Gaussian time series," *IEEE Trans. Signal Process.*, vol. 58, no. 10, pp. 5006–5015, Oct. 2010.
[8] J. K. Tugnait, "Multichannel spectrum sensing via multivariate power spectrum analysis," in *Proc. 12th IEEE Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, San Francisco, CA, USA, Jun. 2011, pp. 111–115.
[9] E. Axell, G. Leus, E. G. Larsson, and H. V. Poor, "Spectrum sensing for cognitive radio: State-of-the-art and recent advances," *IEEE Signal Process. Mag.*, vol. 29, no. 3, pp. 101–116, May 2012.
[10] N. Klausner, M. R. Azimi-Sadjadi, and L. L. Scharf, "Detection of spatially correlated time series from a network of sensor arrays," *IEEE Trans. Signal Process.*, vol. 62, no. 6, pp. 1396–1407, Mar. 2014.
[11] L. Huang, Y. H. Xiao, and Q. T. Zhang, "Robust spectrum sensing for non-circular signal in multiantenna cognitive receivers," *IEEE Trans. Signal Process.*, vol. 63, no. 2, pp. 498–511, Jan. 2015.
[12] J. K. Tugnait, "Multiantenna spectrum sensing for improper signals over frequency selective channels," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Shanghai, China, Mar. 2016, pp. 3781–3785.
[13] J. K. Tugnait, "Multisensor detection of improper signals in improper noise," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, New Orleans, LA, USA, Mar. 2017, pp. 3939–3943.
[14] J. K. Tugnait, "On multisensor detection of improper signals," *IEEE Trans. Signal Process.*, vol. 67, no. 4, pp. 870–884, Feb. 2019.
[15] S. K. Sharma, S. Chatzinotas, and B. Ottersten, "The effect of noise correlation on fractional sampling based spectrum sensing," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Budapest, Hungary, Jun. 2013, pp. 2589–2594.
[16] R. A. Maronna, R. D. Martin, and V. J. Yohai, *Robust Statistics: Theory and Methods*. Hoboken, NJ, USA: Wiley, 2006.
[17] M. Hubert, P. J. Rousseeuw, and S. Van Aelst, "High-breakdown robust multivariate methods," *Stat. Sci.*, vol. 23, pp. 92–119, Feb. 2008.
[18] A. M. Zoubir, V. Koivunen, Y. Chakhchoukh, and M. Muma, "Robust estimation in signal processing: A tutorial-style treatment of fundamental concepts," *IEEE Signal Process. Mag.*, vol. 29, no. 4, pp. 61–80, Jul. 2012.
[19] P. Rousseeuw and M. Hubert, "High-breakdown estimators of multivariate location and scatter," in *Robustness and Complex Data Structures: Festschrift in Honour of Ursula Gather*, C. Becker, R. Fried, and S. Kuhnt, Eds. Berlin, Germany: Springer, 2013, pp. 49–66.
[20] M. Hubert, P. Rousseeuw, D. Vanpaemel, and T. Verdonck, "The DetS and DetMM estimators for multivariate location and scatter," *Comput. Statist. Data Anal.*, vol. 81, pp. 64–75, Jan. 2015.
[21] J. K. Tugnait, "Robust spectrum-based comparison of multivariate complex random signals," *IEEE Access*, vol. 7, pp. 12521–12528, Jan. 2019.
[22] G. Casella and R. L. Berger, *Statistical Inference*, 2nd ed. Pacific Grove, CA, USA: Thomson Learning, 2002.
[23] D. R. Brillinger, *Time Series: Data Analysis and Theory*. New York, NY, USA: McGraw-Hill, 1981.
[24] J. K. Tugnait and S. A. Bhaskar, "On testing for impropriety of multivariate complex-valued random sequences," *IEEE Trans. Signal Process.*, vol. 65, no. 11, pp. 2988–3003, Jun. 2017.
[25] J. K. Tugnait, "Comparing multivariate complex random signals: Algorithm, performance analysis and application," *IEEE Trans. Signal Process.*, vol. 64, no. 4, pp. 934–947, Feb. 2016.
[26] N. Muler and V. J. Yohai, "Robust estimation for vector autoregressive models," *Comput. Statist. Data Anal.*, vol. 65, pp. 68–79, Sep. 2013.
[27] C. Masreliez and R. Martin, "Robust Bayesian estimation for the linear model and robustifying the Kalman filter," *IEEE Trans. Autom. Control*, vol. 22, no. 3, pp. 361–371, Jun. 1977.
[28] N. A. Campbell, H. P. Lopuhaä, and P. J. Rousseeuw, "On the calculation of a robust S-estimator of a covariance matrix," *Statist. Med.*, vol. 17, pp. 2685–2695, Dec. 1998.
[29] M. Salibián-Barrera, S. Van Aelst, and G. Willems, "Principal components analysis based on multivariate MM estimators with fast and robust bootstrap," *J. Amer. Stat. Assoc.*, vol. 101, no. 475, pp. 1198–1211, 2006.
[30] K. Tatsuoka and D. E. Tyler, "On the uniqueness of S-functionals and M-functionals under nonelliptical distributions," *Ann. Statist.*, vol. 28, no. 4, pp. 1219–1243, 2000.
[31] J. Astola, P. Haavisto, and Y. Neuvo, "Vector median filters," *Proc. IEEE*, vol. 78, no. 4, pp. 678–689, Apr. 1990.
[32] C. S. Regazzoni and A. Teschioni, "A new approach to vector median filtering based on space filling curves," *IEEE Trans. Image Process.*, vol. 6, no. 7, pp. 1025–1037, Jul. 1997.
[33] R. Lukac, B. Smolka, K. N. Plataniotis, and A. N. Venetsanopoulos, "Vector sigma filters for noise detection and removal in color images," *J. Vis. Commun. Image Represent.*, vol. 17, pp. 1–26, Feb. 2006.

JITENDRA K. TUGNAIT (M'79–SM'93–F'94–LF'16) received the B.Sc. degree (Hons.) in electronics and electrical communication engineering from the Punjab Engineering College, Chandigarh, India, in 1971, the M.S. and E.E. degrees from Syracuse University, Syracuse, NY, USA, and the Ph.D. degree from the University of Illinois at Urbana–Champaign, in 1973, 1974, and 1978, respectively, all in electrical engineering.

From 1978 to 1982, he was an Assistant Professor of electrical and computer engineering with the University of Iowa, Iowa City, IA, USA. He was with the Long Range Research Division, Exxon Production Research Company, Houston, TX, USA, from June 1982 to September 1989. In September 1989, he joined the Department of Electrical and Computer Engineering, Auburn University, Auburn, AL, USA, as a Professor. He currently holds the title of James B. Davis Professor. His current research interests include statistical signal processing, wireless communications, and multiple target tracking.

Dr. Tugnait was an Associate Editor of the IEEE TRANSACTIONS ON AUTOMATIC CONTROL, the IEEE TRANSACTIONS ON SIGNAL PROCESSING, the IEEE SIGNAL PROCESSING LETTERS, and the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, a Senior Area Editor of the IEEE TRANSACTIONS ON SIGNAL PROCESSING, and a Senior Editor of the IEEE WIRELESS COMMUNICATIONS LETTERS.