IEEE Open Journal of Signal Processing

Received 14 December 2022; revised 7 March 2023; accepted 9 March 2023. Date of publication 24 March 2023; date of current version 25 April 2023. The review of this article was arranged by Associate Editor Lisandro Lovisolo.

*Digital Object Identifier 10.1109/OJSP.2023.3261806*

# Highly Robust Complex Covariance Estimators With Applications to Sensor Array Processing

**JUSTIN A. FISHBONE (Member, IEEE), AND LAMINE MILI (Life Fellow, IEEE)**

Bradley Department of Electrical and Computer Engineering, Virginia Polytechnic Institute and State University, Falls Church, VA 22043 USA

CORRESPONDING AUTHOR: JUSTIN A. FISHBONE (e-mail: Fishbone@vt.edu).

**ABSTRACT** Many applications in signal processing require the estimation of mean and covariance matrices of multivariate complex-valued data. Often, the data are non-Gaussian and are corrupted by outliers or impulsive noise. To mitigate this, robust estimators are employed. However, existing robust estimation techniques employed in signal processing, such as $M$-estimators, provide limited robustness in the multivariate case. For this reason, this paper introduces the signal processing community to the highly robust class of multivariate estimators called multivariate $S$-estimators. The paper extends multivariate $S$-estimation theory to the complex-valued domain. The theoretical performances of $S$-estimators are explored and compared with $M$-estimators through the practical lens of the minimum variance distortionless response (MVDR) beamformer, and the empirical finite-sample performances of the estimators are explored through the practical lens of direction-of-arrival (DOA) estimation using the multiple signal classification (MUSIC) algorithm.

**INDEX TERMS** Complex elliptically symmetric distribution, complex-valued S-estimator, covariance and shape matrix estimation, robust estimation of multivariate location and scatter, Sq-estimator.

## I. INTRODUCTION

Covariance and mean matrices are used ubiquitously throughout the signal processing field, such as in linear filtering [1], array processing [2], and signal detection [3], and for most practical applications, these matrices need to be estimated. To do this, the most commonly employed estimators are the sample estimators, which happen to be the maximum likelihood estimators (MLEs) at the Gaussian distribution, but they lack statistical efficiency for many other distributions such as the Cauchy distribution. Additionally, the sample estimators are not robust against outliers or impulsive noise in the data. The ability to handle non-Gaussian data is important for many applications where the data are better modeled by other distributions, such as speech and radar clutter signals, which are commonly modeled by Laplace [4] and $K$- [5] distributions, respectively. Additionally, the ability to handle large deviations from the true data is important for applications where outliers or impulsive noise may be present. Common sources of outliers include occlusions in image processing [6]; high-power equipment in wireless communications [7]; targets, clutter discretes, other radars, and jamming in adaptive radar [8]; measurement and instrument errors, cyber attacks, and communication interference in power systems [9]; and general sensor and human errors [10].

Recently, there has been increasing interest in the robust estimation of multivariate location and scatter matrices in the signal processing literature, and much of this attention has been on $M$-estimators (for example, see [11], [12], [13], [14], [15], [16], [17], [18], [19], [20], [21], [22]). However, a major drawback of $M$-estimators is that their robustness to outliers is limited in the multivariate case [23], [24], [25], [26]. For example, the statistical asymptotic breakdown point—an important measure of outlier robustness to be defined in Section V—of multivariate $M$-estimators of dimension $p$ is at most $1/(p+1)$, which goes toward zero for large dimensions [27]. As will be demonstrated, this provides little robustness for high-dimensional estimation.

To address this lack of robustness, a high-breakdown-point class of scale-based estimators, called $S$-estimators, was introduced by Rousseeuw and Yohai [28] and Davies [23]. Unlike $M$-estimators, $S$-estimators have an asymptotic breakdown point of $1/2$, for any dimension. In fact, $S$-estimators have been compared against many other types of robust estimators, and they have been recommended in the statistics literature as

the general multivariate estimator of choice for $p \geq 10$ (for real-valued data, which corresponds to $p \geq 5$ for complex-valued data) [26], [29, Sec. 6.10.5]. To date, $S$-estimators have been employed in widely diverse fields for applications such as fault monitoring of industrial processes [30] and outlier detection of gene patterns in biological microarray data [31]. However, $S$-estimators have not been explored much by signal processing researchers. For this reason, this paper introduces $S$-estimation to the field, and it extends the theory to the complex-valued domain.

This paper is organized as follows. Elliptically symmetric distributions are summarized in Section II. Complex-valued $S$-estimators are then introduced in Section III, and common $S$-estimator $\rho$ functions are summarized and extended to the complex domain in Section IV. Properties of complex-valued $S$-estimators are provided in Section V, and these are used in Section VI to compare the theoretical performance of $S$-estimators with other estimation techniques. Finally, Section VII summarizes simulations demonstrating the practical benefits of $S$-estimators.

The following notation is used throughout this paper. The matrix operators $\boldsymbol{X}^\top$, $\boldsymbol{X}^\mathsf{H}$, $\boldsymbol{X}^*$, and $|\boldsymbol{X}|$ are the transpose, Hermitian transpose, conjugate, and determinant, respectively. The Euclidean vector norm is given by $\|\boldsymbol{x}\|$. The matrices $\boldsymbol{I}$ and $\boldsymbol{K}$ are the identity and square commutation matrices. The operator $\mathrm{vec}(\boldsymbol{X})$ stacks the columns of $\boldsymbol{X}$ into a vector. The indicator function is given by $\mathrm{I}(t)$, and $K_x(y)$ is the modified Bessel function of the second kind. The scalar squared Mahalanobis distance function is given by $d(\boldsymbol{x}, \boldsymbol{\mu}, \boldsymbol{\Sigma})$ for random vector $\boldsymbol{x}$ and location and scatter matrices $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, and $d$ is the resulting random scalar (i.e., the second-order modular variate). The matrices $\widehat{\boldsymbol{\mu}}_n$ and $\widehat{\boldsymbol{\Sigma}}_n$ are estimates of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, respectively, for a sample of size $n$. The prime operator indicates a derivative, $f'(t) = \mathrm{d}\, f(t)/\,\mathrm{d}\, t$. The relations $\overset{a}{\sim}$, $\overset{d}{\to}$, and $\overset{p}{\to}$ denote asymptotic equality in distribution, convergence of distribution, and convergence in probability, respectively. The Gaussian normal, complex normal, and generalized complex normal distributions are respectively denoted by $\mathcal{N}$, $\mathbb{C}\mathcal{N}$, and $\mathcal{GCN}$. Statistical independence of random variables $\boldsymbol{a}$ and $\boldsymbol{b}$ is signified by $\boldsymbol{a} \perp \boldsymbol{b}$. Finally, this paper sometimes considers both real- and complex-valued cases simultaneously. In this case, equations are parameterized by $\kappa$, with $\kappa = 1$ indicating the real-valued case and $\kappa = 2$ indicating the complex-valued one.

The estimators and associated properties derived in this paper, along with some of the simulation results, have also been presented in the author's dissertation [32].

## II. ELLIPTICALLY SYMMETRIC DISTRIBUTIONS

Elliptically symmetric distributions are a broad class of multivariate families that cover many common distributions such as the Gaussian, Laplace, hyperbolic, normal inverse Gaussian, variance gamma, $t$-, $K$-, and $W$-distributions. Both real elliptically symmetric (RES) and circular (about $\boldsymbol{\mu}$) complex elliptically symmetric (CES) distributions are defined as being a function of the squared Mahalanobis distance, given by

$$d(\boldsymbol{x}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = (\boldsymbol{x} - \boldsymbol{\mu})^\mathsf{H} \boldsymbol{\Sigma}^{-1} (\boldsymbol{x} - \boldsymbol{\mu}),$$

where $d \in [0, \infty)$, $\boldsymbol{x} \in \mathbb{R}^p$ (RES) or $\boldsymbol{x} \in \mathbb{C}^p$ (CES), the location $\boldsymbol{\mu} \in \mathbb{R}^p$ (RES) or $\boldsymbol{\mu} \in \mathbb{C}^p$ (CES), and $\boldsymbol{\Sigma}$ is the real or complex $p \times p$ positive semi-definite Hermitian scatter matrix. When defined, their probability density functions (PDFs) are of the form given by

$$f_X(\boldsymbol{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \phi, p, \kappa) = \kappa^p \alpha_{\kappa p} |\boldsymbol{\Sigma}|^{-\kappa/2} \phi_{\kappa p}(\kappa d(\boldsymbol{x}, \boldsymbol{\mu}, \boldsymbol{\Sigma})),$$

for some generating function $\phi_{\kappa p}(\kappa d)$, where $\kappa = 1$ corresponds to the RES PDF and $\kappa = 2$ corresponds to the CES PDF, and where the scalar constant $\alpha_{\kappa p}$ ensures $f_X(\boldsymbol{x})$ integrates to one. We denote RES generating functions with $\phi_p(d)$ and CES generating functions with $\varphi_p(d)$, and they are related as $\varphi_p(d) = \phi_{2p}(2d)$—hence the compact notation $\phi_{\kappa p}(\kappa d)$. RES generating functions are summarized for common RES families in [33, Table A1] and CES generating functions for common CES families in [34, Table 1]. Note that the RES and CES mean is equal to $\boldsymbol{\mu}$ when the distribution's mean is defined, and when the distribution's covariance matrix, $\boldsymbol{\mathcal{C}}$, is defined and full rank, then the covariance matrix is a scalar multiple of the scatter matrix, with $\boldsymbol{\mathcal{C}} = p^{-1}\,\mathrm{E}[d]\boldsymbol{\Sigma}$, for both RES [35] and CES [21] distributions.

The PDF of the corresponding RES and CES squared Mahalanobis distance is given by

$$f_D(d; \phi, p, \kappa) = \kappa^p \beta_{\kappa p}\, d^{\kappa p/2 - 1} \phi_{\kappa p}(\kappa d), \quad d \in [0, \infty), \tag{1}$$

where $\beta_{\kappa p} = \alpha_{\kappa p} \pi^{\kappa p/2}/\Gamma(\kappa p/2)$. Not all distributions have analytical expressions for $\alpha_{\kappa p}$, but $\alpha_{\kappa p}$ can easily be calculated using the scalar equation (1) to numerically solve for $\beta_{\kappa p}$ by ensuring (1) integrates to one. Hereafter, unless otherwise noted, all densities, $f(d)$, refer to the density of $d(\boldsymbol{x}, \boldsymbol{\mu}, \boldsymbol{\Sigma})$ in (1), so the subscript $D$ will be omitted.

This paper assumes complex-valued data are CES and circularly symmetric about $\boldsymbol{\mu}$. Non-circular CES data can simply be treated as $2p$-RES data.

## III. COMPLEX-VALUED $S$-ESTIMATORS

$S$-estimators were originally introduced for regression by Rousseeuw [28] in 1984 as a more robust alternative to $M$-estimators. In 1987, Davies [23] expanded the definition to the real-valued estimation of multivariate location and scatter. Using a more contemporary notation than Davies, we now extend multivariate $S$-estimators of location and scatter to the complex-valued domain with the following definition.

*Definition 1 (Complex multivariate S-estimators):* Given a sample of $n$ $p$-dimensional complex-valued observations, $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\}$, complex-valued $S$-estimates of multivariate location and scatter are defined by

$$\begin{aligned} \left(\widehat{\boldsymbol{\mu}}, \widehat{\boldsymbol{\Sigma}}\right) = \arg\min \quad & |\boldsymbol{\Sigma}| \\ \text{subject to} \quad & \frac{1}{n}\sum_{i=1}^{n} \rho\left(\frac{d(\boldsymbol{x}_i, \boldsymbol{\mu}, \boldsymbol{\Sigma})}{\sigma}\right) \leq b. \end{aligned} \tag{2}$$

The function $\rho(t)$ is a scalar, real-valued, nondecreasing, and left-continuous function in $t \geq 0$, with $\rho(0) = 0$ and $\rho(t)$ continuous at $t = 0$, and where there is a point $c$ such that $\rho(t) = \rho(\infty)$ for $t \geq c$.

It is often assumed that $\rho(t)$ is continuously differentiable, in which case the inequality can be replaced with an equality. The parameter $b$ affects the estimator's statistical robustness and efficiency, and when $b = (n - p - 1)\rho(\infty)/(2n)$, the maximum breakdown point (i.e., maximum outlier robustness) is achieved. When $b = E[\rho(d/\sigma)]$, then $\lim_{n\to\infty} \widehat{\boldsymbol{\Sigma}}_n = \boldsymbol{\Sigma}$ almost surely; otherwise, $\widehat{\boldsymbol{\Sigma}}_\infty$ is simply a real-valued scalar multiple of $\boldsymbol{\Sigma}$. The positive scalar $\sigma$ may be chosen (solved for analytically or numerically) to satisfy $E[\rho(d/\sigma)] = b = (n - p - 1)\rho(\infty)/(2n)$ in order to obtain both consistency and the maximum breakdown point. Except where otherwise noted, this paper assumes $S$-estimators where both the consistency and maximum breakdown point conditions hold.

When $\rho(t)$ is continuously differentiable, local solutions to (2) can be found by iteratively evaluating

$$\widehat{\boldsymbol{\mu}} = \frac{\sum_{i=1}^{n} w\left(\frac{d\left(\boldsymbol{x}_i, \widehat{\boldsymbol{\mu}}, \widehat{\boldsymbol{\Sigma}}\right)}{\sigma}\right) \boldsymbol{x}_i}{\sum_{j=1}^{n} w\left(\frac{d\left(\boldsymbol{x}_j, \widehat{\boldsymbol{\mu}}, \widehat{\boldsymbol{\Sigma}}\right)}{\sigma}\right)}, \tag{3}$$

$$\widehat{\boldsymbol{\Sigma}} = \frac{\sum_{i=1}^{n} w\left(\frac{d\left(\boldsymbol{x}_i, \widehat{\boldsymbol{\mu}}, \widehat{\boldsymbol{\Sigma}}\right)}{\sigma}\right) \frac{(\boldsymbol{x}_i - \widehat{\boldsymbol{\mu}})(\boldsymbol{x}_i - \widehat{\boldsymbol{\mu}})^{\mathsf{H}}}{\sigma}}{\sum_{j=1}^{n} v\left(d\left(\boldsymbol{x}_j, \widehat{\boldsymbol{\mu}}, \widehat{\boldsymbol{\Sigma}}\right)/\sigma\right)/p}, \tag{4}$$

where the weight function is the derivative of the $\rho$ function, $w(t) = \rho'(t)$, and $v(t) = tw(t) - \rho(t) + b$. Alternatively, $\widehat{\boldsymbol{\Sigma}}$ can be calculated by neglecting the denominator in (4), and then scaling its determinant, $|\widehat{\boldsymbol{\Sigma}}|$, to satisfy the equality constraint in (2). To show that local solutions of (2) satisfy (3)–(4), we can use the Lagrangian function

$$\mathcal{L} = -\ln|\boldsymbol{V}| - \lambda\left(\frac{1}{n}\sum_{i=1}^{n}\rho\left(\frac{(\boldsymbol{x}_i - \boldsymbol{\mu})^{\mathsf{H}}\boldsymbol{V}(\boldsymbol{x}_i - \boldsymbol{\mu})}{\sigma}\right) - b\right),$$

where $\boldsymbol{V} = \boldsymbol{\Sigma}^{-1}$. Taking the Wirtinger derivatives results in

$$\frac{\partial\mathcal{L}}{\partial\lambda} = \frac{1}{n}\sum_{i=1}^{n}\rho\left(\frac{d(\boldsymbol{x}_i, \boldsymbol{\mu}, \boldsymbol{\Sigma})}{\sigma}\right) - b = 0, \tag{5}$$

$$\frac{\partial\mathcal{L}}{\partial\boldsymbol{\mu}} = \frac{\lambda}{n}\sum_{i=1}^{n} w\left(\frac{d(\boldsymbol{x}_i, \boldsymbol{\mu}, \boldsymbol{\Sigma})}{\sigma}\right)\frac{(\boldsymbol{x}_i - \boldsymbol{\mu})^{\mathsf{H}}\boldsymbol{V}}{\sigma} = \boldsymbol{0}, \tag{6}$$

$$\frac{\partial\mathcal{L}}{\partial\boldsymbol{\mu}^*} = \frac{\lambda}{n}\sum_{i=1}^{n} w\left(\frac{d(\boldsymbol{x}_i, \boldsymbol{\mu}, \boldsymbol{\Sigma})}{\sigma}\right)\frac{(\boldsymbol{x}_i - \boldsymbol{\mu})^{\top}\boldsymbol{V}^{\top}}{\sigma} = \boldsymbol{0}, \tag{7}$$

$$\frac{\partial\mathcal{L}}{\partial\boldsymbol{V}} = -\frac{\lambda}{n}\sum_{i=1}^{n} w\left(\frac{d(\boldsymbol{x}_i, \boldsymbol{\mu}, \boldsymbol{\Sigma})}{\sigma}\right)\frac{(\boldsymbol{x}_i - \boldsymbol{\mu})^{*}(\boldsymbol{x}_i - \boldsymbol{\mu})^{\top}}{\sigma} - \boldsymbol{V}^{-\top} = \boldsymbol{0}, \tag{8}$$

$$\frac{\partial\mathcal{L}}{\partial\boldsymbol{V}^*} = \boldsymbol{0}.$$

Substituting the factorization $\boldsymbol{\Sigma} = \boldsymbol{A}\boldsymbol{A}^{\mathsf{H}}$ into (8) results in

$$-\frac{n}{\lambda}\boldsymbol{I} = \sum_{i=1}^{n} w\left(\frac{d(\boldsymbol{x}_i, \boldsymbol{\mu}, \boldsymbol{\Sigma})}{\sigma}\right)\frac{\boldsymbol{A}^{-1}(\boldsymbol{x}_i - \boldsymbol{\mu})(\boldsymbol{x}_i - \boldsymbol{\mu})^{\mathsf{H}}\boldsymbol{A}^{-\mathsf{H}}}{\sigma}.$$

Taking the trace of this equation gives

$$\lambda^{-1} = -\frac{1}{np}\sum_{i=1}^{n} w\left(\frac{d(\boldsymbol{x}_i, \boldsymbol{\mu}, \boldsymbol{\Sigma})}{\sigma}\right)\frac{d(\boldsymbol{x}_i, \boldsymbol{\mu}, \boldsymbol{\Sigma})}{\sigma}. \tag{9}$$

Combining (9) with (5) and (8) results in (4). Solving for $\boldsymbol{\mu}$, (6) and (7) both equal (3).

The formulation used in Definition 1 is convenient for proving properties of $S$-estimators. An alternative, and perhaps more common, formulation is given by

$$\begin{aligned}\left(\widehat{\boldsymbol{\mu}}, \widehat{\boldsymbol{\Omega}}\right) = \arg\min \quad & \widehat{\sigma} \\ \text{subject to} \quad & |\boldsymbol{\Omega}| = 1, \\ & \frac{1}{n}\sum_{i=1}^{n}\rho\left(\frac{d(\boldsymbol{x}_i, \boldsymbol{\mu}, \boldsymbol{\Omega})}{\widehat{\sigma}}\right) \leq b,\end{aligned}$$

where the shape matrix, $\boldsymbol{\Omega}$, is the unit-volume scaled scatter matrix given by $\boldsymbol{\Omega} = \boldsymbol{\Sigma}/|\boldsymbol{\Sigma}|^{1/p}$, meaning $|\boldsymbol{\Omega}| = 1$. When $\rho$ is continuous, local solutions are found using the same weighted sums (3) and (4), but with $\widehat{\boldsymbol{\Omega}}$ substituted for $\widehat{\boldsymbol{\Sigma}}$, where $\widehat{\boldsymbol{\Omega}}$ is rescaled with each iteration so that $|\widehat{\boldsymbol{\Omega}}| = 1$, and with $\widehat{\sigma}$ substituted for $\sigma$, where $\widehat{\sigma}$ is the solution of (5) but using $\widehat{\sigma}$. See [29, Sec. 6.8.2 and Sec. 9.6] for more explicit details on numerical solutions of $S$-estimators. This formulation is convenient because it does not require the calculation of $\sigma$ or $v(\cdot)$, and many practical applications only require the shape matrix. The scatter matrix can be estimated by scaling $\widehat{\boldsymbol{\Omega}}$ using (5), which then results in the same estimate as Definition 1. An alternative and more robust estimate to scale $\widehat{\boldsymbol{\Omega}}$ is discussed in [33] and given by

$$\widehat{\boldsymbol{\Sigma}} = \frac{\text{Median}\left\{d\left(\boldsymbol{x}_1, \widehat{\boldsymbol{\mu}}, \widehat{\boldsymbol{\Omega}}\right), \ldots, d\left(\boldsymbol{x}_n, \widehat{\boldsymbol{\mu}}, \widehat{\boldsymbol{\Omega}}\right)\right\}}{F^{-1}(0.5)}\widehat{\boldsymbol{\Omega}},$$

where $F(d)$ is the cumulative distribution function (CDF) corresponding to (1), and $F^{-1}(0.5)$ is the distribution's median.

For the initial estimates in the iterative approach of (3)–(4), it is generally recommended to use a highly robust, albeit less statistically efficient, estimator such as the minimum volume ellipsoid estimate obtained through subsampling (see [29, Sec. 6.8]).

The result of Definition 1 and (3)–(4) is that the definition and calculation of complex-valued $S$-estimators are a straightforward extension of real-valued $S$-estimators. However, some of the $\rho$ functions and properties need to be extended to the complex-valued domain. These are addressed in the next two sections, respectively.

## IV. ESTIMATOR $\rho$ FUNCTIONS

There are four notable $\rho$ functions used for location and scatter $S$-estimators. Regression $S$-estimators were first introduced with the minimum volume ellipsoid (MVE) $\rho$ function [28]. Although highly robust, the function is discontinuous, resulting in poor statistical efficiency. The smooth bisquare (alternatively known as biweight) function quickly gained popularity for its improved efficiency, and it was the standard for multivariate $S$-estimators when they were introduced [23], [24]. However, for location and scatter estimation, the bisquare is not tunable, and therefore its effective robustness diminishes with large dimension, $p$. Rocke [25] addressed this problem by introducing a tunable $\rho$ function, the biflat, which was later extended by Maronna et al. [36, Sec. 6.4.4] and termed the Rocke function. The Rocke $S$-estimator, however, can have poor efficiency—particularly for low dimensions or non-Gaussian distributions. To improve the efficiency of $S$-estimators, Fishbone and Mili [33] introduced the tunable $S_q$ $\rho$ function for RES distributions.

These four $\rho$ functions are now provided. The Rocke and $S_q$ functions assume real-valued data, so they are extended to the complex domain. Later, in Section VI and Fig. 1, we will explore examples of the corresponding weight functions and compare them to weight functions for other estimation techniques.

The minimum volume ellipsoid $\rho$ function is given by $\rho_{\text{MVE}}(t) = \text{I}(t > 1)$. This is discontinuous, so (3)–(4) do not apply, and it has a slow asymptotic convergence rate of $n^{-1/3}$, unlike smooth $\rho$ functions, which have convergence rates of $n^{-1/2}$. The minimum volume ellipsoid is often implemented with subsampling (see [29, Sec. 6.8.4]), and it is applicable for both real- and complex-valued data.

The $\rho$ function for the bisquare is given by $\rho_{\text{bisq}}(t) = \min\{1, 1-(1-t)^3\}$, and the corresponding weight function is given by $w_{\text{bisq}}(t) = 3(1-t)^2\,\text{I}(t \le 1)$. This is applicable for both real and complex data.

The Rocke $\rho$ and weight functions are given by

$$\rho_\gamma(t) = \begin{cases} 0 & \text{if } 0 \le t \le 1-\gamma \\ \frac{t-1}{4\gamma}\left[3 - \left(\frac{t-1}{\gamma}\right)^2\right] + \frac{1}{2} & \text{if } 1-\gamma < t < 1+\gamma \\ 1 & \text{if } 1+\gamma \le t \end{cases},$$

$$w_\gamma(t) = \frac{3}{4\gamma}\left[1 - \left(\frac{t-1}{\gamma}\right)^2\right]\text{I}(1-\gamma \le t \le 1+\gamma),$$

where the tuning parameter $\gamma \in (0, 1]$ increases the estimator's robustness, and generally decreases the estimator's efficiency, as its value is decreased. While these equations are the same for both real and complex data, Maronna et al. [36, eq. (6.40)] suggest tuning this using $\gamma = \min\{\chi^2_p(1-\alpha)/p - 1, 1\}$, where $\chi^2_p(1-\alpha)$ is the $(1-\alpha)^{\text{th}}$ percentile of the chi-squared distribution. This assumes real-valued Gaussian data and is designed to reject observations with an approximate asymptotic rejection probability of $\alpha$. For CES Gaussian data, we note that $d \sim \text{Gamma}(p, 1)$, and proceeding as in [36, Sec. 6.4.4], we obtain $\gamma = \min\{\Gamma^{-1}_{p,1}(1-\alpha)/p - 1, 1\}$, where $\Gamma^{-1}_{p,1}(1-\alpha)$ is the inverse CDF of the Gamma$(p, 1)$ distribution. Noting that $\Gamma^{-1}_{p,1}(1-\alpha)/p = \chi^2_{2p}(1-\alpha)/(2p)$, we can compactly write

$$\gamma = \min\left\{\frac{\chi^2_{\kappa p}(1-\alpha)}{\kappa p} - 1, 1\right\},$$

for RES ($\kappa = 1$) and CES ($\kappa = 2$) Gaussian data.

The $S_q$ function is derived from a density-weighted maximum likelihood estimator for the scale of $d(\boldsymbol{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$, and it is defined for most common continuous RES distributions (such as those in [33, Table A1]). The $S_q$ function is tuned with parameter $q \le 1$ and is given by

$$\rho_q(t) = \begin{cases} 0 & \text{if } q < 1 \text{ and } t \le a \\ s_1\left(\tilde{\rho}_q(t) - \tilde{\rho}_q(a)\right) & \text{if } q < 1 \text{ and } a < t < c \\ 1 & \text{if } q < 1 \text{ and } t \ge c \\ \tilde{\rho}_q(t) & \text{if } q = 1 \end{cases},$$

where $s_1 = (\tilde{\rho}_q(b) - \tilde{\rho}_q(a))^{-1}$, and $a$ is the minimum and $c$ the maximum of $\tilde{\rho}_q(t)$, which is given by $\tilde{\rho}_q(t) = -tf'(t)/f^q(t)$. Using (1) and proceeding as in [33], we obtain for both RES and CES distributions,

$$\tilde{\rho}_q(t) = -f(t)^{s_q}\left(\kappa t\frac{\phi'_{\kappa p}(\kappa t)}{\phi_{\kappa p}(\kappa t)} + s_{\kappa p}\right),$$

where $s_{\kappa p} = \kappa p/2 - 1$, and $s_q = 1 - q$, and the values of $a$ and $c$ are summarized in Table 1 for general elliptically symmetric subclasses. The values in Table 1 correspond to the RES $\phi$ functions as defined in [33, Table A1]. The corresponding $S_q$ weight function is given by

$$w_q(t) = \begin{cases} 0 & \text{if } q < 1 \text{ and } t \le a \\ s_1\tilde{w}_q(t) & \text{if } q < 1 \text{ and } a < t < c \\ 0 & \text{if } q < 1 \text{ and } t \ge c \\ \tilde{w}_q(t) & \text{if } q = 1 \end{cases},$$

with

$$\tilde{w}_q(t) = -\kappa f(t)^{s_q}\left(\frac{s_q s^2_{\kappa p}}{\kappa t} + \left(2s_q s_{\kappa p} + 1\right)\frac{\phi'_{\kappa p}(\kappa t)}{\phi_{\kappa p}(\kappa t)} - q\kappa t\left(\frac{\phi'_{\kappa p}(\kappa t)}{\phi_{\kappa p}(\kappa t)}\right)^2 + \kappa t\frac{\phi''_{\kappa p}(\kappa t)}{\phi_{\kappa p}(\kappa t)}\right).$$

There are a few important remarks about the $S_q$ function. Although the $S_q$ is indexed to an assumed PDF, it is distributionally robust to model mismatch [33], which will be illustrated in simulations later in this paper. When $q = 1$, $\rho_q(t)$ is not generally a proper (bounded) $S$-estimator $\rho$ function, but it does correspond to the maximum likelihood estimator for the scale of $d$. Using the Gaussian phi function with $q = 1$ results in the sample estimators. When $q < 1$, $\beta_{\kappa p}$ can be dropped from the calculations of $\rho_q(t)$ and $w_q(t)$ for simplicity. Finally, in some cases, there are a few restrictions

**TABLE 1.** $S_q$ **Inlier and Outlier Rejection Points for Common RES/CES Families**

| Distribution | Inlier Rejection Point $a$ and Outlier Rejection Point $c$ |
|---|---|
| Kotz type | $a, c = \left(\frac{s+2s_q N+2s_{\kappa p}s_q \mp\sqrt{s^2+4ss_q N+4ss_{\kappa p}s_q}}{2\kappa^s s_q rs}\right)^{1/s}$ |
| Gaussian | $a, c = \frac{1+2s_{\kappa p}s_q \mp\sqrt{1+4s_{\kappa p}s_q}}{s_q\kappa}$ |
| Pearson type II | $a, c = \frac{2s_q s_{\kappa p}^2+m(2s_q s_{\kappa p}+1)\mp\sqrt{m^2(4s_q s_{\kappa p}+1)+4ms_q s_{\kappa p}^2}}{2\kappa\left(s_q s_{\kappa p}^2+m(2s_q s_{\kappa p}+ms_q)\right)}$ |
| Pearson type VII | $a, c = s\frac{2Ns_q s_{\kappa p}+N-2s_q s_{\kappa p}^2 \mp\sqrt{4N^2 s_q s_{\kappa p}-4Ns_q s_{\kappa p}^2+N^2}}{2\kappa s_q\left(s_{\kappa p}^2-2Ns_{\kappa p}+N^2\right)}$ |
| Generalized hyperbolic | $a = \begin{cases} 0 & when\ \chi = 0\ and\ \lambda = 1 \\ \{t\|\tilde{w}_q(t) = 0\ and\ t \in (0,c)\} & otherwise \end{cases}$ |
| | $c = \{t\|\tilde{w}_q(t) = 0\ and\ t \in (a,\infty)\}$ |

**TABLE 2.** **Restrictions on Parameter** $q$ **for Common RES/CES Families**

| Distribution | Valid Range of $q$ |
|---|---|
| Kotz type | $q \le 1$ unless $-1 - s_{\kappa p} < N < -s_{\kappa p}$, then $1 + 2^{-2}s\,(s_{\kappa p}+N)^{-1} < q \le 1$ |
| Gaussian | $q \le 1$ |
| Pearson type II | $q = 1$ or $q < 1 - m^{-1}$ |
| Pearson type VII | $q \le 1$ |
| Generalized hyperbolic* | $q \le 1$ unless $\chi = 0$ and $\lambda < 1$, then *unknown* $< q \le 1$ |

**Empirically inferred. Computational precision restricts* $q \notin (0.998, 1)$, *approximately.*

on $q$ beyond $q \le 1$. These follow directly from the real-valued restrictions in [33] and are summarized in Table 2.

## V. PROPERTIES OF COMPLEX-VALUED *S*-ESTIMATORS

Properties of complex-valued $S$-estimators will now be provided. Some of the properties follow directly from their real-valued counterparts provided by Davies [23] and Lopuhaä [24], [37]. Note that Davies defines $S$-estimators using the function $\kappa(t)$ and the robustness value $\epsilon$, where $\kappa(t) = 1 - \rho(t)/\rho(\infty)$, and $\epsilon = b/\rho(\infty)$. Lopuhaä defines $\rho_{\mathrm{L}}$ as a function of the Mahalanobis distance, not the squared Mahalanobis distance (i.e., $\rho_{\mathrm{L}}(\sqrt{t}) = \rho(t)$). Both authors absorb the constant $\sigma$ into their definition of $\rho(t)$.

In this section, we define three sets of assumptions. Assumptions (A0) are the basic assumptions of the distribution in order to establish the $S$-estimator properties. These assumptions are

(A0) $b = \mathrm{E}[\rho(d/\sigma)]$,
$\boldsymbol{x}_i \sim \mathrm{CES}\,(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \varphi, p)$, where $\boldsymbol{x}_{\mathrm{i}}$ are i.i.d.,
$\varphi_p(d)$ is non-increasing,
$\varphi_p(d), -\rho(d/\sigma)$ have common point(s) of decrease.

These follow straightforwardly from the RES case. Most common continuous CES distributions satisfy these assumptions, including those in [33, Table A1].

Assumptions (A1) are assumptions on the derivatives of the $\rho$ function that enable the derivation of the asymptotic distribution and influence functions of $S$-estimators. Assumptions (A1) depend on whether the data come from RES or CES distributions, so we parameterize them with $\kappa$. These assumptions are

(A1) $w(t)$ and $w'(t)$ are bounded and continuous,
$$\mathrm{E}\left[w(d/\sigma) + (3-\kappa)p^{-1}(d/\sigma)w'(d/\sigma)\right] > 0,$$
$$\mathrm{E}\left[(p+3-\kappa)(d/\sigma)w(d/\sigma) + (3-\kappa)(d/\sigma)^2 w'(d/\sigma)\right] > 0.$$

Finally, Assumptions (A2) provide an alternate set of assumptions to Assumptions (A1) for the asymptotic distribution. These assumptions are

(A2) $w(t)$ and $tw(t)$ are of bounded variation,
$\phi_p'(t)$ is continuous and decreasing with $\phi_p'(d) < 0$.

The proofs in this paper utilize Assumptions (A0) and (A1), but we also provide results under Assumptions (A2) and leave it to future work to derive the proofs under those assumptions.

### A. BASIC PROPERTIES

The affine equivariance, uniqueness, existence, consistency, and breakdown point of complex-valued $S$-estimators follows directly from their real-valued counterparts. For completeness, we briefly summarize those properties here.

Complex-valued $S$-estimators are affine equivariant, which follows directly from their definition and is discussed for the real-valued case in [23, p. 1271].

*Theorem 1 (Uniqueness):* Assuming (A0), the solution of minimizing $|\widehat{\boldsymbol{\Sigma}}|$ subject to

$$\int_{\mathbb{C}^p} \rho\left(\frac{d\left(\boldsymbol{x}, \widehat{\boldsymbol{\mu}}, \widehat{\boldsymbol{\Sigma}}\right)}{\sigma}\right) f_X\left(d\left(\boldsymbol{x}, \boldsymbol{\mu}, \boldsymbol{\Sigma}\right)\right) \mathrm{d}\boldsymbol{x} = b$$

has a unique solution $(\widehat{\boldsymbol{\mu}}, \widehat{\boldsymbol{\Sigma}}) = (\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

*Proof:* The proof follows the same as for [23, Th. 1]. ■

*Remark 1:* While this establishes the asymptotic uniqueness for most elliptical distributions, note that $S$-estimators generally have many finite-sample solutions (i.e., (2) is non-convex for the finite-sample case).

*Theorem 2 (Existence):* Assuming (A0), if $n \geq (p+1)/(1-b/\rho(\infty))$, then (2) has at least one solution with probability one.

*Proof:* The proof follows the same as for [23, Th. 2]. ■

*Theorem 3 (Consistency):* Assuming (A0), given $b = \mathrm{E}[\rho(d/\sigma)]$, and assuming that $p+1 \leq n(1-b/\rho(\infty))$, then, almost surely,

$$\lim_{n\to\infty} \left(\widehat{\boldsymbol{\mu}}_n, \widehat{\boldsymbol{\Sigma}}_n\right) = \left(\boldsymbol{\mu}, \boldsymbol{\Sigma}\right).$$

*Proof:* The proof follows the same as for [23, Th. 3]. ■

The finite-sample *breakdown point* of a multivariate estimator of location and scatter is an important measure of robustness and is defined to be the proportion of the sample that can be set such that either $\|\widehat{\boldsymbol{\mu}}\| = \infty$ or that an eigenvalue of $\widehat{\boldsymbol{\Sigma}}$ can be driven to either zero or infinity [38]. In the following, the meaning of a sample in *general position* is that at most $p$ observations lie in any hyperplane of dimension less than $p$.

*Theorem 4 (Breakdown point):* Assuming (A0), when $n$ observations are in general position, with $n(1-2b/\rho(\infty)) \geq p+1$, the breakdown point of complex-valued $S$-estimators is $(\lfloor nb/\rho(\infty)\rfloor + 1)/n$.

*Proof:* The proof follows the same as for [23, Th. 5], which assumes that $\rho(\delta) = 0$, for some $\delta > 0$. Lopuhaä and Rousseeuw [38, Th. 3.2] provide an alternative proof that does not require $\delta > 0$ but that assumes $\rho(t)$ is twice continuously differentiable. ■

*Corollary 1:* The maximum breakdown point is $\lfloor (n-p+1)/2\rfloor/n$, which is achieved when $b/\rho(\infty) = 1/2 - (p+1)/(2n)$.

*Remark 2:* $S$-estimators, therefore, obtain the maximum theoretical breakdown point that any affine equivariant estimator may achieve [23, Th. 6].

### B. ASYMPTOTIC NORMALITY

To derive the asymptotic normal distribution of complex-valued multivariate $S$-estimators for CES distributions, we can utilize the local solutions (3) and (4). We first recall the RES asymptotic distribution and then state the CES asymptotic distribution. To prove the CES asymptotic distribution, we map complex-valued $S$-estimates to real-valued ones, derive those real-valued distributions, then map the result back to the complex-valued domain.

Using the real-valued equivalents of (3) and (4), Lopuhaä [24] derived the asymptotic distribution of real-valued $S$-estimates of location and scatter using Assumptions (A1). Lopuhaä [37] then extended his previous results using Assumptions (A2) instead of Assumptions (A1). When both (A1) and (A2) are true, both approaches lead to different, but equal, formulations, which are now summarized. The asymptotic distribution of real-valued multivariate $S$-estimates is given by

$$\sqrt{n}\left(\widehat{\boldsymbol{\mu}}_n - \boldsymbol{\mu}, \operatorname{vec}\left(\widehat{\boldsymbol{\Sigma}}_n - \boldsymbol{\Sigma}\right)\right) \xrightarrow{d} (\boldsymbol{a}, \boldsymbol{b}), \tag{10}$$

with $\boldsymbol{a} \perp \boldsymbol{b}$, where $\boldsymbol{a} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Gamma}_{\boldsymbol{\mu},\mathbb{R}} = \frac{\omega_{1,\mathbb{R}}}{\omega_{2,\mathbb{R}}^2}\boldsymbol{\Sigma})$ with

$$\omega_{1,\mathbb{R}} = 4p^{-1}\,\mathrm{E}\left[dw^2(d/\sigma)\right],$$

$$\omega_{2,\mathbb{R}} = \begin{cases} 2\,\mathrm{E}\left[w(d/\sigma) + 2p^{-1}(d/\sigma)w'(d/\sigma)\right] & \text{(A1)} \\ -2\beta_p \int_0^\infty 2p^{-1}d^{p/2}w(d/\sigma)\phi_p'(d)\ \mathrm{d}d & \text{(A2)}, \end{cases}$$

and where $\boldsymbol{b} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Gamma}_{\boldsymbol{\Sigma},\mathbb{R}})$, with

$$\boldsymbol{\Gamma}_{\boldsymbol{\Sigma},\mathbb{R}} = \zeta_{1,\mathbb{R}}\left(\boldsymbol{I}+\boldsymbol{K}\right)\left(\boldsymbol{\Sigma}\otimes\boldsymbol{\Sigma}\right) + \zeta_{2,\mathbb{R}}\operatorname{vec}\left(\boldsymbol{\Sigma}\right)\operatorname{vec}\left(\boldsymbol{\Sigma}\right)^\top,$$

$$\zeta_{1,\mathbb{R}} = \lambda_{1,\mathbb{R}}^{-2}p(p+2)\,\mathrm{E}\left[(d/\sigma)^2\,w^2(d/\sigma)\right],$$

$$\zeta_{2,\mathbb{R}} = \lambda_{2,\mathbb{R}}^{-2}\,\mathrm{E}\left[(\rho(d/\sigma)-b)^2\right] - 2p^{-1}\zeta_1,$$

$$\lambda_{1,\mathbb{R}} = \begin{cases} \mathrm{E}\left[(p+2)(d/\sigma)w(d/\sigma)+2(d/\sigma)^2\ w'(d/\sigma)\right] & \text{(A1)} \\ -2\beta_p\int_0^\infty \sigma^{-1}d^{p/2+1}w(d/\sigma)\phi_p'(d)\ \mathrm{d}d & \text{(A2)}, \end{cases}$$

$$\lambda_{2,\mathbb{R}} = \begin{cases} \mathrm{E}\left[(d/\sigma)w(d/\sigma)\right] & \text{(A1)} \\ -\beta_p\int_0^\infty d^{p/2}\left(\rho(d/\sigma)-b\right)\phi_p'(d)\ \mathrm{d}d & \text{(A2)}. \end{cases}$$

We now extend this result to complex-valued $S$-estimates.

*Theorem 5 (Asymptotic normality):* Assuming (A0) and (A1), the asymptotic distribution of $(\widehat{\boldsymbol{\mu}}_n, \widehat{\boldsymbol{\Sigma}}_n)$ is given by

$$\sqrt{n}\left(\widehat{\boldsymbol{\mu}}_n - \boldsymbol{\mu}, \operatorname{vec}\left(\widehat{\boldsymbol{\Sigma}}_n - \boldsymbol{\Sigma}\right)\right) \xrightarrow{d} (\boldsymbol{a}, \boldsymbol{b}),$$

with $\boldsymbol{a} \perp \boldsymbol{b}$, where $\boldsymbol{a} \sim \mathbb{C}\mathcal{N}(\boldsymbol{0}, \boldsymbol{\Gamma}_{\boldsymbol{\mu}} = \frac{\omega_1}{\omega_2^2}\boldsymbol{\Sigma})$ with

$$\omega_1 = p^{-1}\,\mathrm{E}\left[dw^2(d/\sigma)\right],$$

$$\omega_2 = \mathrm{E}\left[w(d/\sigma) + p^{-1}(d/\sigma)w'(d/\sigma)\right], \tag{11}$$

and where $\boldsymbol{b} \sim \mathcal{GCN}(\mathbf{0}, \boldsymbol{\Gamma}_{\boldsymbol{\Sigma}}, \boldsymbol{C}_{\boldsymbol{\Sigma}} = \boldsymbol{\Gamma}_{\boldsymbol{\Sigma}} \boldsymbol{K})$ with

$$\begin{aligned}
\boldsymbol{\Gamma}_{\boldsymbol{\Sigma}} &= \zeta_1 \boldsymbol{\Sigma}^{\top} \otimes \boldsymbol{\Sigma} + \zeta_2 \operatorname{vec}(\boldsymbol{\Sigma}) \operatorname{vec}(\boldsymbol{\Sigma})^{\mathsf{H}}, \\
\zeta_1 &= \lambda_1^{-2} 4p(p+1) \operatorname{E}\left[(d/\sigma)^2 w^2(d/\sigma)\right], \\
\zeta_2 &= \lambda_2^{-2} \operatorname{E}\left[(\rho(d/\sigma) - b)^2\right] - p^{-1}\zeta_1, \\
\lambda_1 &= 2 \operatorname{E}\left[(p+1)(d/\sigma) w(d/\sigma) + (d/\sigma)^2 w'(d/\sigma)\right], \quad (12) \\
\lambda_2 &= \operatorname{E}\left[(d/\sigma) w(d/\sigma)\right]. \quad (13)
\end{aligned}$$

### C. PROOF OF THEOREM 5

In order to prove Theorem 5, we can exploit the $S$-estimator solution equations given by (3) and (4). These equations can be considered to be of the form of an $M$-estimator, but require a more general multivariate $M$-estimator definition than was introduced by Maronna [27] and is usually defined in the literature (for example, see [20], [21], [29], [39]). This more prevalent $M$-estimator definition is defined with equations similar to (3) and (4), but without the function $v(t)$, which is instead replaced by the constant $p$. For this proof, we take an approach similar to [17, Th. IV.1] for $M$-estimators, but we have to extend it to allow for the definition of $v(t)$ in (4). It is $v(t)$ that incorporates the constraint (5), which is the key to the high breakdown point of $S$-estimators.

#### 1) DEFINITIONS

Let us define the functions

$$\begin{aligned}
h(\boldsymbol{a}) &= \begin{bmatrix} \operatorname{Re}(\boldsymbol{a})^{\top} & \operatorname{Im}(\boldsymbol{a})^{\top} \end{bmatrix}^{\top}, \\
g(\boldsymbol{A}) &= \frac{1}{2} \begin{bmatrix} \operatorname{Re}(\boldsymbol{A}) & -\operatorname{Im}(\boldsymbol{A}) \\ \operatorname{Im}(\boldsymbol{A}) & \operatorname{Re}(\boldsymbol{A}) \end{bmatrix},
\end{aligned}$$

the skew-symmetric orthogonal matrix

$$\boldsymbol{P} = \begin{bmatrix} \mathbf{0}_{p\times p} & -\boldsymbol{I}_{p\times p} \\ \boldsymbol{I}_{p\times p} & \mathbf{0}_{p\times p} \end{bmatrix},$$

and the vectors and matrices $\boldsymbol{\mu}_{\mathbb{R}} = h(\boldsymbol{\mu})$, $\widehat{\boldsymbol{\mu}}_n^{\mathbb{R}} = h(\widehat{\boldsymbol{\mu}}_n)$, $\boldsymbol{\Sigma}_{\mathbb{R}} = g(\boldsymbol{\Sigma})$, $\widehat{\boldsymbol{\Sigma}}_n^{\mathbb{R}} = g(\widehat{\boldsymbol{\Sigma}}_n)$, $\boldsymbol{u} = h(\boldsymbol{x})$, $\boldsymbol{v} = \boldsymbol{P}\boldsymbol{u}$, $\boldsymbol{\mu}_u = h(\boldsymbol{\mu})$, and $\boldsymbol{\mu}_v = \boldsymbol{P}\boldsymbol{\mu}_u$. The vectors $\boldsymbol{u} \sim \operatorname{RES}(\boldsymbol{\mu}_u, \boldsymbol{\Sigma}_{\mathbb{R}}, \phi, 2p)$ and $\boldsymbol{v} \sim \operatorname{RES}(\boldsymbol{\mu}_v, \boldsymbol{\Sigma}_{\mathbb{R}}, \phi, 2p)$.

Note that $g(\boldsymbol{x}\boldsymbol{x}^{\mathsf{H}}) = (\boldsymbol{u}\boldsymbol{u}^{\top} + \boldsymbol{v}\boldsymbol{v}^{\top})/2$, and for any $p \times p$ Hermitian matrix $\boldsymbol{A}$, $2\boldsymbol{x}^{\mathsf{H}}\boldsymbol{A}^{-1}\boldsymbol{x} = \boldsymbol{u}^{\top} g(\boldsymbol{A})^{-1}\boldsymbol{u} = \boldsymbol{v}^{\top} g(\boldsymbol{A})^{-1}\boldsymbol{v}$ [20]. Therefore, we have $2d(\boldsymbol{x}, \widehat{\boldsymbol{\mu}}_n, \widehat{\boldsymbol{\Sigma}}_n) = d(\boldsymbol{u}, \widehat{\boldsymbol{\mu}}_n^{\mathbb{R}}, \widehat{\boldsymbol{\Sigma}}_n^{\mathbb{R}}) = d(\boldsymbol{v}, \boldsymbol{P}\widehat{\boldsymbol{\mu}}_n^{\mathbb{R}}, \widehat{\boldsymbol{\Sigma}}_n^{\mathbb{R}})$. Defining $\rho_{\mathbb{R}}(t) = \rho(t/2)$ gives $w_{\mathbb{R}}(t) = w(t/2)/2$ and $v_{\mathbb{R}}(t) = v(t/2)$. Applying $h(\cdot)$ to (3) gives

$$\widehat{\boldsymbol{\mu}}_n^{\mathbb{R}} = \frac{\sum_{i=1}^n w_{\mathbb{R}}\left(\frac{d(\boldsymbol{u}_i, \widehat{\boldsymbol{\mu}}_n^{\mathbb{R}}, \widehat{\boldsymbol{\Sigma}}_n^{\mathbb{R}})}{\sigma}\right) \boldsymbol{u}_i}{\sum_{j=1}^n w_{\mathbb{R}}\left(\frac{d(\boldsymbol{u}_j, \widehat{\boldsymbol{\mu}}_n^{\mathbb{R}}, \widehat{\boldsymbol{\Sigma}}_n^{\mathbb{R}})}{\sigma}\right)}, \quad (14)$$

and applying $g(\cdot)$ to (4) gives

$$\begin{aligned}
\widehat{\boldsymbol{\Sigma}}_n^{\mathbb{R}} = \frac{1}{2} &\frac{\sum_{i=1}^n w_{\mathbb{R}}\left(\frac{d(\boldsymbol{u}_i, \widehat{\boldsymbol{\mu}}_n^{\mathbb{R}}, \widehat{\boldsymbol{\Sigma}}_n^{\mathbb{R}})}{\sigma}\right) \frac{(\boldsymbol{u}_i - \widehat{\boldsymbol{\mu}}_n^{\mathbb{R}})(\boldsymbol{u}_i - \widehat{\boldsymbol{\mu}}_n^{\mathbb{R}})^{\top}}{\sigma}}{\sum_{j=1}^n v_{\mathbb{R}}\left(d\left(\boldsymbol{u}_j, \widehat{\boldsymbol{\mu}}_n^{\mathbb{R}}, \widehat{\boldsymbol{\Sigma}}_n^{\mathbb{R}}\right)/\sigma\right)/(2p)} \\
&+ \frac{1}{2} \frac{\sum_{i=1}^n w_{\mathbb{R}}\left(\frac{d(\boldsymbol{v}_i, \boldsymbol{P}\widehat{\boldsymbol{\mu}}_n^{\mathbb{R}}, \widehat{\boldsymbol{\Sigma}}_n^{\mathbb{R}})}{\sigma}\right) \frac{(\boldsymbol{v}_i - \boldsymbol{P}\widehat{\boldsymbol{\mu}}_n^{\mathbb{R}})(\boldsymbol{v}_i - \boldsymbol{P}\widehat{\boldsymbol{\mu}}_n^{\mathbb{R}})^{\top}}{\sigma}}{\sum_{j=1}^n v_{\mathbb{R}}\left(d\left(\boldsymbol{v}_j, \boldsymbol{P}\widehat{\boldsymbol{\mu}}_n^{\mathbb{R}}, \widehat{\boldsymbol{\Sigma}}_n^{\mathbb{R}}\right)/\sigma\right)/(2p)}. \quad (15)
\end{aligned}$$

Now applying (3) and (4) for real-valued $S$-estimates of location and scatter for $\boldsymbol{u}$, and rearranging for later simplicity, gives

$$\mathbf{0} = \frac{1}{n}\sum_{i=1}^n w_{\mathbb{R}}\left(\frac{d\left(\boldsymbol{u}_i, \widehat{\boldsymbol{\mu}}_n^u, \widehat{\boldsymbol{\Sigma}}_n^u\right)}{\sigma}\right)\left(\boldsymbol{u}_i - \widehat{\boldsymbol{\mu}}_n^u\right), \quad (16)$$

$$\begin{aligned}
\mathbf{0} = \frac{1}{n}\sum_{i=1}^n &w_{\mathbb{R}}\left(\frac{d\left(\boldsymbol{u}_i, \widehat{\boldsymbol{\mu}}_n^u, \widehat{\boldsymbol{\Sigma}}_n^u\right)}{\sigma}\right) \frac{\left(\boldsymbol{u}_i - \widehat{\boldsymbol{\mu}}_n^u\right)\left(\boldsymbol{u}_i - \widehat{\boldsymbol{\mu}}_n^u\right)^{\top}}{\sigma} \\
&- \frac{1}{2p} v_{\mathbb{R}}\left(\frac{d\left(\boldsymbol{u}_i, \widehat{\boldsymbol{\mu}}_n^u, \widehat{\boldsymbol{\Sigma}}_n^u\right)}{\sigma}\right) \widehat{\boldsymbol{\Sigma}}_n^u. \quad (17)
\end{aligned}$$

Equivalently for $\boldsymbol{v}$, we have

$$\mathbf{0} = \frac{1}{n}\sum_{i=1}^n w_{\mathbb{R}}\left(\frac{d\left(\boldsymbol{v}_i, \widehat{\boldsymbol{\mu}}_n^v, \widehat{\boldsymbol{\Sigma}}_n^v\right)}{\sigma}\right)\left(\boldsymbol{v}_i - \widehat{\boldsymbol{\mu}}_n^v\right), \quad (18)$$

$$\widehat{\boldsymbol{\Sigma}}_n^v = \frac{\sum_{i=1}^n w_{\mathbb{R}}\left(\frac{d\left(\boldsymbol{v}_i, \widehat{\boldsymbol{\mu}}_n^v, \widehat{\boldsymbol{\Sigma}}_n^v\right)}{\sigma}\right) \frac{\left(\boldsymbol{v}_i - \widehat{\boldsymbol{\mu}}_n^v\right)\left(\boldsymbol{v}_i - \widehat{\boldsymbol{\mu}}_n^v\right)^{\top}}{\sigma}}{\sum_{j=1}^n v_{\mathbb{R}}\left(d\left(\boldsymbol{v}_j, \widehat{\boldsymbol{\mu}}_n^v, \widehat{\boldsymbol{\Sigma}}_n^v\right)/\sigma\right)/(2p)}. \quad (19)$$

#### 2) ASYMPTOTIC RELATION BETWEEN REAL ESTIMATES

*Lemma 1 (Consistency of $\widehat{\boldsymbol{\Sigma}}_n^u$ and $\widehat{\boldsymbol{\Sigma}}_n^v$):*

$$\widehat{\boldsymbol{\Sigma}}_{\infty}^u = \boldsymbol{\Sigma}_u = \boldsymbol{\Sigma}_{\mathbb{R}} = \boldsymbol{\Sigma}_v = \widehat{\boldsymbol{\Sigma}}_{\infty}^v$$

*Proof:* By [23, Th. 3], (14) and (15) are consistent estimates. The $\boldsymbol{u}$ and $\boldsymbol{v}$ estimator equations (16)–(19) match the form of the $S$-estimator equations (3)–(4). Applying (1) to the consistency constraint that $b = \operatorname{E}[\rho(d/\sigma)]$, we have

$$\begin{aligned}
\operatorname{E}\left[\rho_{\mathbb{R}}\left(\frac{d\left(\boldsymbol{u}; \boldsymbol{\mu}_u, \boldsymbol{\Sigma}_{\mathbb{R}}\right)}{\sigma}\right)\right] &= \operatorname{E}\left[\rho_{\mathbb{R}}\left(\frac{d\left(\boldsymbol{v}; \boldsymbol{\mu}_v, \boldsymbol{\Sigma}_{\mathbb{R}}\right)}{\sigma}\right)\right] \\
&= \int_0^{\infty} \rho_{\mathbb{R}}\left(\frac{d_u}{\sigma}\right) f(d_u; \phi, 2p, 1)\, \mathrm{d}d_u \\
&= \int_0^{\infty} \rho\left(\frac{d_x}{\sigma}\right) f(d_x; \phi, p, 2)\, \mathrm{d}d_x \\
&= \operatorname{E}\left[\rho\left(\frac{d\left(\boldsymbol{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}\right)}{\sigma}\right)\right] \\
&= b.
\end{aligned}$$

Therefore, by Theorem 3 and [23, Th. 3], we have consistency. ■

*Lemma 2:* $\widehat{\boldsymbol{\mu}}_n^{\mathbb{R}} \overset{a}{\sim} \widehat{\boldsymbol{\mu}}_n^{u}$ and $\widehat{\boldsymbol{\Sigma}}_n^{\mathbb{R}} \overset{a}{\sim} \frac{1}{2}(\widehat{\boldsymbol{\Sigma}}_n^{u} + \widehat{\boldsymbol{\Sigma}}_n^{v})$, and asymptotically, $\widehat{\boldsymbol{\mu}}_n^{\mathbb{R}} \perp \widehat{\boldsymbol{\Sigma}}_n^{\mathbb{R}}$ and $\widehat{\boldsymbol{\mu}}_n^{u} \perp \widehat{\boldsymbol{\Sigma}}_n^{u}$.

*Proof:* See Appendix A. ■

#### 3) ASYMPTOTIC DISTRIBUTION

The final part of this proof directly follows parts three of the proofs of [20, Th. IV.1] and [17, Th. IV.1], but using the asymptotic distribution of real-valued $S$-estimators (10). The inverse transformation matrix $\boldsymbol{g} = [\boldsymbol{I}_p, -j\boldsymbol{I}_p]^\top$ gives, for any $p \times p$ Hermitian matrix $\boldsymbol{A}$, $\boldsymbol{A} = \boldsymbol{g}^{\mathsf{H}} g(\boldsymbol{A}) \boldsymbol{g}$, and it gives $\boldsymbol{x} = \boldsymbol{g}^{\mathsf{H}} h(\boldsymbol{x})$.

Using $\boldsymbol{\Sigma}_{\mathbb{R}} = \boldsymbol{\Sigma}_u$ from Lemma 1 and $\widehat{\boldsymbol{\mu}}_n^{\mathbb{R}} \overset{a}{\sim} \widehat{\boldsymbol{\mu}}_n^{u}$ from Lemma 2, and applying $\widehat{\boldsymbol{\mu}}_n = \boldsymbol{g}^{\mathsf{H}} \widehat{\boldsymbol{\mu}}_n^{\mathbb{R}}$, it is found that

$$\begin{aligned}\boldsymbol{\Gamma}_{\boldsymbol{\mu}} &= \boldsymbol{g}^{\mathsf{H}} n \,\mathrm{E}\left[ \left(\widehat{\boldsymbol{\mu}}_n^{u} - \boldsymbol{\mu}_u\right) \left(\widehat{\boldsymbol{\mu}}_n^{u} - \boldsymbol{\mu}_u\right)^\top \right] \boldsymbol{g} \\ &= \boldsymbol{g}^{\mathsf{H}} \left[ \frac{\omega_{1,\mathbb{R}}}{\omega_{2,\mathbb{R}}^2} \boldsymbol{\Sigma}_{\mathbb{R}} \right] \boldsymbol{g} = \frac{\omega_{1,\mathbb{R}}}{\omega_{2,\mathbb{R}}^2} \boldsymbol{\Sigma},\end{aligned}$$

where the scalars $\omega_{1,\mathbb{R}}$ and $\omega_{2,\mathbb{R}}$ are given in (10) using $2p$, $w_{\mathbb{R}}(d(\boldsymbol{u}; \boldsymbol{\mu}_u, \boldsymbol{\Sigma}_u)/\sigma)$, and $w'_{\mathbb{R}}(d(\boldsymbol{u}; \boldsymbol{\mu}_u, \boldsymbol{\Sigma}_u)/\sigma)$. Substituting in $\boldsymbol{x}$, $w(\cdot)$, and $w'(\cdot)$, results in the expressions stated in the theorem. Using the same approach, it is found that

$$\boldsymbol{C}_{\boldsymbol{\mu}} = \boldsymbol{g}^{\mathsf{H}} n \,\mathrm{E}\left[ \left(\widehat{\boldsymbol{\mu}}_n^{u} - \boldsymbol{\mu}_u\right) \left(\widehat{\boldsymbol{\mu}}_n^{u} - \boldsymbol{\mu}_u\right)^\top \right] \boldsymbol{g}^* = \boldsymbol{0}.$$

Similarly, applying $\widehat{\boldsymbol{\Sigma}}_n = \boldsymbol{g}^{\mathsf{H}} \widehat{\boldsymbol{\Sigma}}_n^{\mathbb{R}} \boldsymbol{g}$ and Lemmas 1 and 2, it is found that [17, e.q. (20)–(23)]

$$\begin{aligned}\boldsymbol{\Gamma}_{\boldsymbol{\Sigma}} &= n \,\mathrm{E}\left[ \mathrm{vec}\left(\widehat{\boldsymbol{\Sigma}}_n - \boldsymbol{\Sigma}\right) \mathrm{vec}\left(\widehat{\boldsymbol{\Sigma}}_n - \boldsymbol{\Sigma}\right)^{\mathsf{H}} \right] \\ &= \left(\boldsymbol{g}^\top \otimes \boldsymbol{g}^{\mathsf{H}}\right) \boldsymbol{\Gamma}_{\boldsymbol{\Sigma}_u,\mathbb{R}} \left(\boldsymbol{g}^\top \otimes \boldsymbol{g}^{\mathsf{H}}\right)^{\mathsf{H}},\end{aligned}$$

where $\boldsymbol{\Gamma}_{\boldsymbol{\Sigma}_u,\mathbb{R}}$ is given in (10) using $2p$, $\rho_{\mathbb{R}}(d(\boldsymbol{u}; \boldsymbol{\mu}_u, \boldsymbol{\Sigma}_u)/\sigma)$, $w_{\mathbb{R}}(d(\boldsymbol{u}; \boldsymbol{\mu}_u, \boldsymbol{\Sigma}_u)/\sigma)$, and $w'_{\mathbb{R}}(d(\boldsymbol{u}; \boldsymbol{\mu}_u, \boldsymbol{\Sigma}_u)/\sigma)$. Substituting in $\boldsymbol{\Sigma} = \boldsymbol{g}^{\mathsf{H}} \boldsymbol{\Sigma}_u \boldsymbol{g}$, $\boldsymbol{x}$, $\rho(\cdot)$, $w(\cdot)$, and $w'(\cdot)$, results in the expressions stated in the theorem. We also have [17, e.q. (28)–(30)]

$$\begin{aligned}\boldsymbol{C}_{\boldsymbol{\mu}} &= n \,\mathrm{E}\left[ \mathrm{vec}\left(\widehat{\boldsymbol{\Sigma}}_n - \boldsymbol{\Sigma}\right) \mathrm{vec}\left(\widehat{\boldsymbol{\Sigma}}_n - \boldsymbol{\Sigma}\right)^\top \right] \\ &= n \,\mathrm{E}\left[ \mathrm{vec}\left(\widehat{\boldsymbol{\Sigma}}_n - \boldsymbol{\Sigma}\right) \mathrm{vec}\left(\widehat{\boldsymbol{\Sigma}}_n - \boldsymbol{\Sigma}\right)^{\mathsf{H}} \boldsymbol{K} \right] = \boldsymbol{\Gamma}_{\boldsymbol{\Sigma}} \boldsymbol{K}.\end{aligned}$$

#### 4) COMMENTS

Assuming Lemma 2 holds under Assumptions (A2), then Theorem 5 also holds under (A2) without needing (A1). In this case, we have

$$\omega_2 = -2^{p+1} \beta_{2p} \int_0^\infty p^{-1} d^p w(d/\sigma) \phi'_{2p}(2d) \,\mathrm{d}d, \tag{20}$$

$$\lambda_1 = -2^{p+2} \beta_{2p} \int_0^\infty \sigma^{-1} d^{p+1} w(d/\sigma) \phi'_{2p}(2d) \,\mathrm{d}d, \tag{21}$$

$$\lambda_2 = -2^{p+1} \beta_{2p} \int_0^\infty d^p \left(\rho(d/\sigma) - b\right) \phi'_{2d}(2\,d) \,\mathrm{d}d. \tag{22}$$

An alternative approach for calculating the asymptotic variance is to use the influence function, with [40, Sec. IV.B]

$$\begin{aligned}\boldsymbol{\Gamma}_{\boldsymbol{T}} &= \mathrm{E}\left[ \mathrm{vec}\left(\mathbf{IF}\left(z; \boldsymbol{T}\right)\right) \mathrm{vec}\left(\mathbf{IF}\left(z; \boldsymbol{T}\right)\right)^{\mathsf{H}} \right], \\ \boldsymbol{C}_{\boldsymbol{T}} &= \mathrm{E}\left[ \mathrm{vec}\left(\mathbf{IF}\left(z; \boldsymbol{T}\right)\right) \mathrm{vec}\left(\mathbf{IF}\left(z; \boldsymbol{T}\right)\right)^\top \right],\end{aligned}$$

where the influence function, $\mathbf{IF}(\cdot)$, for the estimate $\boldsymbol{T}$ will now be defined and derived.

### D. INFLUENCE FUNCTION

The influence function (IF) is another important measure of asymptotic robustness [41]. The influence function can be seen as a function that characterizes the sensitivity to an infinitesimal point contamination at location $\boldsymbol{z} \in \mathbb{C}^p$, standardized by the probability (mass) of the contamination, $\epsilon$. Boundedness and continuity of the influence function are critical for robust estimators since these characteristics imply limited effects from small amounts of contamination. Additionally, from the influence function, other important robustness measures can be calculated, such as the asymptotic rejection point, gross-error sensitivity, and local-shift sensitivity. For an estimator $\boldsymbol{T}$ at nominal distribution $F$, the influence function is defined as

$$\begin{aligned}\mathbf{IF}\left(z; \boldsymbol{T}, F\right) &= \lim_{\epsilon \to 0^+} \frac{\boldsymbol{T}((1-\epsilon)F + \epsilon \Delta_z) - \boldsymbol{T}(F)}{\epsilon} \\ &= \frac{\partial}{\partial \epsilon} \boldsymbol{T}((1-\epsilon)F + \epsilon \Delta_z)|_{\epsilon=0},\end{aligned}$$

where the $\epsilon$ proportion of the sample is a point-mass, $\Delta_z$, at location $z$.

*Theorem 6 (Influence function):* Assuming (A0) and (A1), the influence functions of $\widehat{\boldsymbol{\mu}}$ and $\widehat{\boldsymbol{\Sigma}}$ are given by

$$\begin{aligned}\mathbf{IF}\left(z; \boldsymbol{\mu}, F\right) &= \frac{2\sqrt{d_z} w(d_z/\sigma)}{\omega_2} \frac{z_c}{\sqrt{d_z}}, \\ \mathbf{IF}\left(z; \boldsymbol{\Sigma}, F\right) &= \frac{\rho\left(d_z/\sigma\right) - b}{\lambda_2} \boldsymbol{\Sigma} \\ &\quad + \frac{2p(p+1) d_z w\left(d_z/\sigma\right)}{\sigma \lambda_1} \left( \frac{z_c z_c^{\mathsf{H}}}{d_z} - \frac{1}{p} \boldsymbol{\Sigma} \right),\end{aligned} \tag{23}$$

where $z_c = z - \boldsymbol{\mu}$, $d_z = z_c^{\mathsf{H}} \boldsymbol{\Sigma}^{-1} z_c$, $\omega_2$ is given by (11), $\lambda_1$ is given by (12), and $\lambda_2$ is given by (13).

*Proof:* See Appendix B. ■

Assuming Lemma 2 holds under Assumptions (A2), then Theorem 6 also holds under (A2) without needing (A1). In this case, $\omega_2$, $\lambda_1$, and $\lambda_2$ are respectively given by (20)–(22).

## VI. ASYMPTOTIC COMPARISON WITH *M*-ESTIMATORS

The asymptotic statistical performances of $S$-estimators are now compared with other common estimation techniques, specifically, the traditional sample covariance matrix (SCM) estimator and the common Huber $M$-estimator. The SCM is an $M$-estimator with $w(t) = 1$ and is the maximum likelihood estimator for mean and covariance at the Gaussian

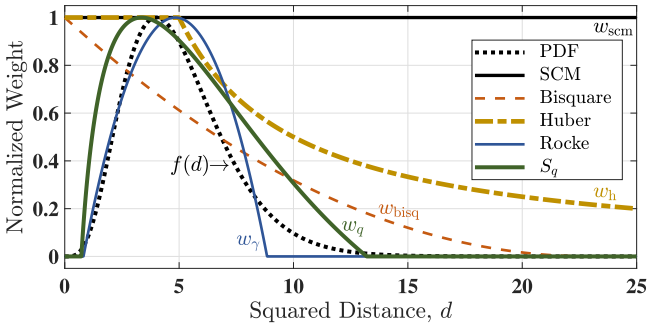



**FIGURE 1. Normalized asymptotic weight functions at the five-dimensional CES Gaussian distribution. The reference PDF is also plotted.**

distribution. The Huber weight function is given by $w_h(t) = \zeta \cdot \min\{1, h/t\}$, where the parameter $h$ controls the estimator's robustness and the parameter $\zeta$ affects the estimator's consistency.

As mentioned above, multivariate $M$-estimators of location and scatter are commonly defined with equations similar to (3) and (4), but with $v(t) = p$. For comprehensive details on the common $M$-estimator assumptions, see [27] and [21]. Most notably, however, their weight functions are usually assumed to be continuous and nonincreasing, and $tw(t)$ is usually assumed to be nondecreasing [42]. The implication of this is that $M$-estimators are what Maronna et al. [29] call *monotone* estimators. Monotone $M$-estimators provide the benefit of a unique solution, but as previously mentioned, have limited outlier robustness. As discussed above, $S$-estimators solved with (3) and (4) only have unique solutions asymptotically (under Assumptions (A0)), but they are highly robust.

To explore the asymptotic performance of $S$-estimators, and for a fair comparison with $M$-estimators, we follow an approach similar to that used by Ollila and Koivunen [22] to introduce and explore complex-valued $M$-estimators. We start by comparing the weight functions of the different estimation methods. The influence functions and asymptotic efficiencies are then compared through the lens of a common signal processing application, adaptive beamforming. Although the results here are through the lens of beamforming, the relative and qualitative results generally apply across most applications. Here, the focus is on the estimation of scatter matrices since they generally have significantly larger errors than location estimates, and they tend to drive overall estimator performance [26], [33].

The array processing technique employed here is Capon's minimum variance distortionless response (MVDR) beamformer [43]. The adaptive MVDR array weights are given by $\mathbf{w} = (\boldsymbol{s}^{\mathsf{H}}\widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{s})^{-1}\widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{s}$, where $\widehat{\boldsymbol{\Sigma}}$ is the estimated scatter of the received signal, and $\boldsymbol{s}$ is the array response (or steering) vector.

### A. ESTIMATOR WEIGHT FUNCTIONS

To help convey an initial understanding of the relative benefits of $S$-estimators, Fig. 1 plots normalized asymptotic estimator weight functions at the CES Gaussian distribution with $p = 5$. The bisquare, Rocke, and $S_q$ $S$-estimators are compared with the Huber $M$-estimator and the SCM. For reference, the probability density function is also plotted. By definition, the SCM weights, $w_{\text{scm}}$, are uniform, no matter how large the outlier. The bisquare $S$-estimator, $w_{\text{bisq}}$, is also not tunable. For the other estimators, their tuning parameters affect their cutoff points.

For the Huber weight function, $w_h$, we chose $h = p$ and $\zeta$ such that $\sigma = 1$. The Huber function provides uniform weighting for observations of squared distance less than $h$. Therefore, at the limit as $h \to \infty$, the SCM is obtained. As $h \to 0$, the Huber $M$-estimator becomes more robust. For squared distances slightly above $h$, the weight function first drops off rapidly, but for larger $d$, the function only slowly approaches zero. In fact, multivariate $M$-estimator weight functions only asymptotically approach zero [42], which means outliers will always receive positive weight, no matter how large.

The benefit of $S$-estimators is apparent by the fact that $S$-estimators give zero weight to large outliers. The Rocke and $S_q$ $S$-estimators also give zero weight to small inliers. For the plot in Fig. 1, we used the moderate tuning values $\alpha = 0.05$ for the Rocke $S$-estimator weights, $w_\gamma$, and $q = 0.9$ for the $S_q$-estimator weights, $w_q$. A benefit of the $S_q$-estimator is that its weight function follows the general shape of the assumed PDF, providing more weight to more probable observations and less weight to less probable ones.

### B. MVDR INFLUENCE FUNCTION

The influence function of the MVDR beamformer is now used to better understand the benefits of $S$-estimators for practical applications. Ollila and Koivunen [22, Th. 1] derived the influence function of the centered adaptive MVDR beamformer for CES distributions, giving

$$\mathbf{IF}\,(z; \mathbf{w}_{\boldsymbol{\Sigma}}, F) = \alpha_{\boldsymbol{\Sigma}}(d_z)\left(\mathbf{w}\boldsymbol{s}^{\mathsf{H}} - \boldsymbol{I}\right)\frac{\boldsymbol{\Sigma}^{-1}z_c z_c^{\mathsf{H}}\mathbf{w}}{d_z},$$

where the function $\alpha_{\boldsymbol{\Sigma}}(d_z)$ depends on the estimator, and the other variables are defined as before. For $M$-estimators of scatter, they gave

$$\alpha_{\boldsymbol{\Sigma},M}(t) = \frac{(t/\sigma)\,w\,(t/\sigma)}{1 + \left(p^2 + p\right)^{-1}\mathrm{E}\left[w'\,(d/\sigma)\,(d/\sigma)^2\right]}.$$

For $S$-estimators of scatter, by applying [22, Lemma 1] to (23), it is seen that

$$\alpha_{\boldsymbol{\Sigma},S}(t) = 2p(p+1)\lambda_1^{-1}\,(t/\sigma)\,w\,(t/\sigma).$$

Fig. 2 plots $\alpha_{\boldsymbol{\Sigma}}(t)$ for the same estimators as depicted in Fig. 1. Here, however, the Huber $M$-estimator was changed to have $h = 1$ (and $\zeta$ such that $\sigma = 1$) to emphasize the fact that inliers influence $M$-estimates [42]. Additionally, although $M$-estimators can have bounded influence functions, exceedingly large outliers are still able to influence their results, resulting in large biases. $S$-estimators, however, provide the benefit of nulling the influence of large outliers. Additionally, some $S$-estimators, such as the Rocke and $S_q$, can also null the influence of inliers.

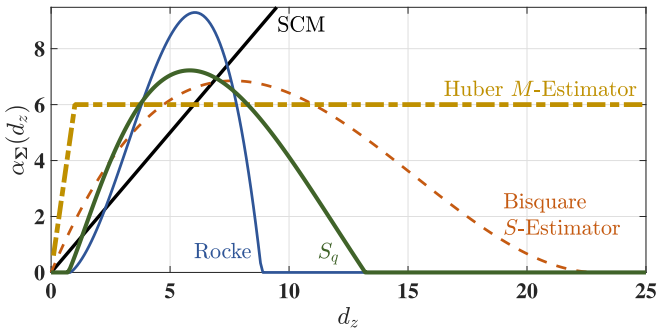



**FIGURE 2.** $\alpha_\Sigma(d_z)$ of the MVDR influence functions at the five-dimensional CES Gaussian distribution.

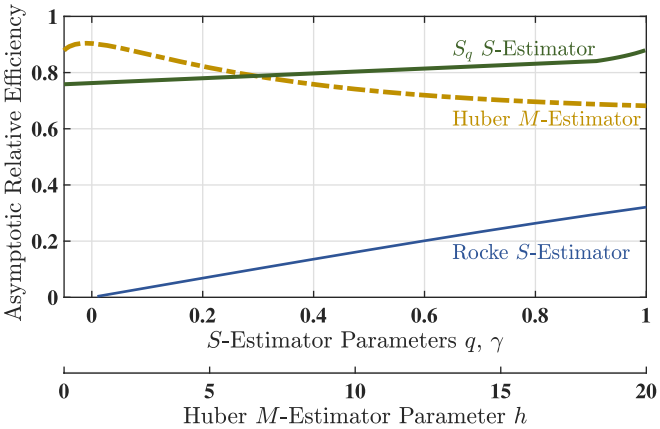



**FIGURE 3.** MVDR asymptotic relative efficiencies versus tuning parameter at the two-dimensional CES $K$-distribution with $\nu = 1$.

### C. MVDR ASYMPTOTIC EFFICIENCY

Estimator efficiency is an important performance metric that conveys the rate of the estimator's convergence with respect to sample size. Ollila and Koivunen [22, Th. 2] derived the asymptotic covariance matrix of the centered adaptive MVDR beamformer for CES distributions, giving

$$\mathrm{ASC}\left(\widehat{\mathbf{w}}_{\boldsymbol{\Sigma}}; F\right) = \xi\left(\frac{\boldsymbol{\Sigma}^{-1}}{\boldsymbol{s}^{\mathsf{H}}\boldsymbol{\Sigma}^{-1}\boldsymbol{s}} - \mathbf{w}\mathbf{w}^{\mathsf{H}}\right),$$

where $\xi = \mathrm{E}[\alpha^2_{\boldsymbol{\Sigma}}(d)]/(p^2+p)$. They also showed that the asymptotic MVDR pseudo-covariance matrix is zero. The MVDR asymptotic relative efficiency (ARE) is then defined as the ratio of $\xi$ for the maximum likelihood estimator to $\xi$ for the given estimator, that is,

$$\mathrm{ARE}\left(\widehat{\mathbf{w}}_{\boldsymbol{\Sigma}}; F\right) = \frac{\mathrm{Tr}\left[\mathrm{ASC}\left(\widehat{\mathbf{w}}_{\mathrm{mle}}; F\right)\right]}{\mathrm{Tr}\left[\mathrm{ASC}\left(\widehat{\mathbf{w}}_{\boldsymbol{\Sigma}}; F\right)\right]} = \frac{\xi_{\mathrm{mle}}}{\xi_{\widehat{\boldsymbol{\Sigma}}}}.$$

Historically, a major drawback of $S$-estimators has been poor efficiency at many non-Gaussian distributions. With the introduction of the $S_q$-estimator, however, efficiency was greatly improved. Fig. 3 plots the MVDR asymptotic relative efficiency as a function of estimator tuning parameter for the Huber $M$-estimator, and Rocke and $S_q$ $S$-estimators at the two-dimensional CES $K$-distribution with $\nu = 1$. The $K$-distribution is defined by $\varphi_p(t) = \sqrt{t}^{\nu-p}K_{\nu-p}(2\sqrt{\nu t})$. For the Huber $M$-estimator, $\zeta$ was chosen such that $\sigma = 1$ for consistency at the assumed $K$-distribution. As seen in the plot, the $S_q$-estimator provides superior efficiency as compared to the Rocke $S$-estimator. The $M$-estimator class includes maximum likelihood estimators, so whereas $M$-estimators are generally able to achieve higher efficiencies than $S$-estimators, the $S_q$-estimator is able to achieve an efficiency nearly as high as the Huber $M$-estimator in this example.

Although Huber's estimator is perhaps the most commonly applied tunable $M$-estimator, tuning it is complicated by the fact that it has two parameters, $h$ and $\zeta$, and its efficiency as a function of parameter $h$ can be non-convex, as seen in Fig. 3. The tunable $S$-estimators discussed here, however, have a single tuning parameter that generally controls the efficiency monotonically, which makes their tuning simpler.

## VII. FINITE-SAMPLE PERFORMANCE SIMULATIONS

The finite-sample performances of $S$-estimators are now compared to $M$-estimators by way of simulations. While the focus continues to be primarily on robustness to outliers, we also include a mismatched $S_q$-estimator in order to illustrate its distributional robustness. Additionally, we also include a complex-valued $R$-estimator—a rank-based class of robust estimators introduced by Hallin, Oja, and Paindaveine in [44] for RES shape matrix estimation and recently extended to CES data by Fortunati, Renaux, and Pascal in [45]. $R$-estimators are designed to be distributionally robust and computationally efficient—unlike $M$- and $S$-estimators, they do not require iterative computation to solve [46].

The application used for demonstration is non-coherent signal direction-of-arrival (DOA) estimation using the multiple signal classification (MUSIC) algorithm [47], which provides much higher resolution than the MVDR beamformer for DOA applications [48]. The sensor model employed was a uniform linear array of $p = 10$ elements spaced at half-wavelengths, and the simulation used narrowband plane waves. The steering vector is given by

$$\boldsymbol{s}(\theta) = \left[1,\ \exp(-1\pi j\sin(\theta)),\ \ldots,\ \exp(-9\pi j\sin(\theta))\right]^{\top}.$$

The MUSIC pseudospectrum (i.e., spectral power as a function of angle) is given by

$$P(\theta) = \frac{1}{\boldsymbol{s}(\theta)^{\mathsf{H}}\widehat{\boldsymbol{E}}\widehat{\boldsymbol{E}}^{\mathsf{H}}\boldsymbol{s}(\theta)},$$

where $\widehat{\boldsymbol{E}}$ is an estimate of $\boldsymbol{E}$, the matrix whose columns are the eigenvectors that span the noise subspace. The number of signals, $n_s$, was assumed known, and $\widehat{\boldsymbol{E}}$ was obtained by horizontally stacking the $p - n_s$ eigenvectors corresponding to the smallest eigenvalues of the estimated shape matrices. The direction-of-arrival estimates were then determined as the $n_s$ largest peaks of $P(\theta)$.

We first simulated a quadrature phase shift keying (QPSK) signal at $+20^\circ$ off the array normal. The noise was complex $K$-distributed with $\nu = 10$, zero mean, and scatter matrix $\boldsymbol{I}$.

The QPSK signal was scaled to achieve 3 dB signal-to-noise ratio (SNR) incident on the array.

To estimate location and shape, we used a small sample of 30 snapshots ($n = 3p$). The estimators employed were the SCM, minimum volume ellipsoid (MVE), matched and mismatched $S_q$-, Tyler $M$-, and matched $R$-estimators. We implemented the MVE estimator using the subsampling technique as recommended by Maronna et al. [29, Sec. 6.8.4] with subsample size of 30. The matched $S_q$-estimator was configured for the $\nu = 10$ $K$-distribution. One can use an auxiliary estimator of $\nu$ for the $S_q$ model, but for these simulations, the mismatched $S_q$-estimator was configured for the $K$-distribution with a constant model error with $\nu = 100$. Both $S_q$-estimators were initialized using the MVE estimate, and they were both set to their maximum breakdown point with $b = (n - p - 1)/(2n)$. The Tyler estimator [49], $M_{\text{Ty}}$, is very popular in the signal processing literature [39, Sec. 4.4.3]. It is obtained by taking Huber's $M$-estimator to its robust limit at $h \to 0$, and the resulting scatter weight function is simply $w_T(t) = 1/t$—note that this is unbounded at $t = 0$ and may result in estimator instability [29, Sec. 6.3.3]. Finally, we also included the $R$-estimator as described in [50] by using software provided at [46]. For the $R$-estimator, we used the score function given by [45, eq. (30)] assuming the same model as for the matched $S_q$-estimator—a $\nu = 10$ $K$-distribution.

A finite-sample efficiency of a MUSIC direction-of-arrival estimator can be defined as the mean squared error of the estimate in the absence of contamination. For a fair comparison, and using this definition, the matched $S_q$-estimator was tuned to obtain the same efficiency as the $M_{\text{Ty}}$-estimator, with $q = 0.865$. The mismatched $S_q$-estimator used this same value of $q$.

Samples were contaminated using replacement, where the elements of each contaminated observation, $\boldsymbol{x}$, were set to $x_i = k$. Using a large outlier value of $k = 100$ to highlight the effects of outliers, the top of Fig. 4 plots the root mean squared error (RMSE) of the direction estimates versus the number of contaminated observations. The SCM was breaking down with a single outlier. Tyler's $M$-estimator was breaking down with three outliers. Because the $R$-estimator implementation described in [50] uses the $M_{\text{Ty}}$-estimator for initialization, it too performed poorly with three outliers. The MVE and $S_q$-estimators, however, did not break down. Although the MVE estimator is highly robust, it lacks efficiency, as seen with the higher RMSE as compared to the $S_q$-estimator. The distributional robustness of the $S_q$-estimator means that it maintains high efficiency, even under model mismatch. This is evident with the similar RMSEs of the mismatched and matched $S_q$-estimates. To improve the performance of the $R$-estimator, we also tried initializing it with the MVE estimate, as we did for the $S_q$-estimators. The result was a more robust $R$-estimator, but with only slight improvement over the MVE estimator alone.

For scatter and other normal matrices, the condition number is the ratio of the largest to the smallest eigenvalue, and large condition numbers indicate numerically singular matrices. To

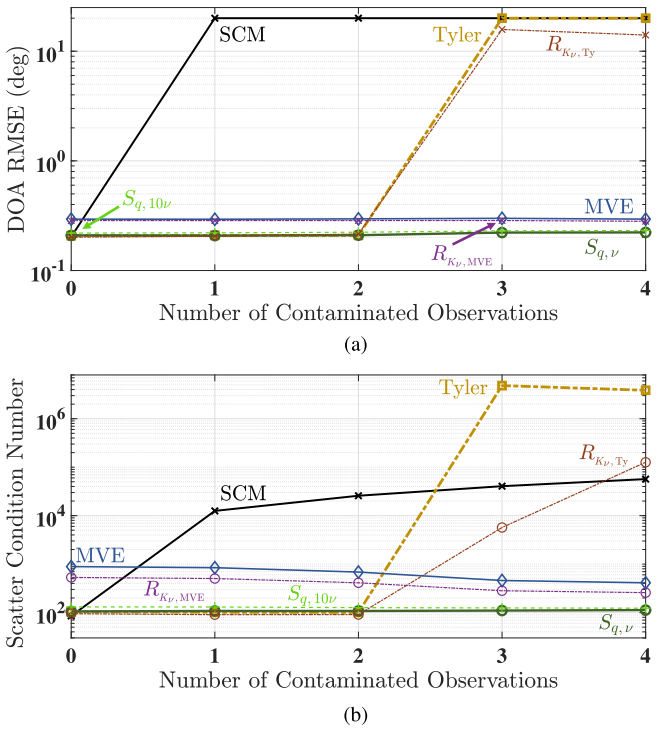



**FIGURE 4. Root mean squared errors of direction-of-arrival estimates versus the number of contaminated observations (a), and corresponding scatter matrix condition number (b).**

further illustrate the breakdown of the estimators as a function of the number of outliers, the bottom of Fig. 4 plots the average matrix condition number. Due to the breakdown of the $M_{\text{Ty}}$-estimator and its non-uniform weighting, it becomes substantially more ill-conditioned than even the SCM for just a few outliers.

Next, we briefly diverge from the focus on robustness to outliers and explore the distributional robustness of the estimators. Fig. 5 illustrates the uncontaminated performance of the estimators as a function of the $K$-distribution parameter $\nu$. The $S_q$-estimator was fixed with parameter $\nu = 10$. Fig. 5(a) depicts the DOA RMSE for the estimators. Fig. 5(b) plots the Frobenius norm of the mean squared error of the noise-only shape matrix estimates [16, eq. (69)], and these are compared to the complex constrained semiparametric Cramér-Rao bound (CCSCRB) normalized by $n$ [16, eq. (32)]. In both of these, the $M_{\text{Ty}}$-estimator performs well due to the lack of contamination, and the Tyler-initialized $R$-estimator generally provides slight improvement over the estimator used to initialize it. For the mismatched $S_q$-estimator where the underlying distribution had a larger $\nu$, the estimator actually performed better than the $M_{\text{Ty}}$-estimator, and almost as well as the Tyler-initialized $R$-estimator. However, its performance did fall off some for distributions with smaller $\nu$. Although setting a smaller value of $q$ results in an $S_q$-estimate that is more robust to outliers, setting a larger value for $q$ generally results in an estimator that is more robust to perturbations in the assumed model [33]. This is evident in Fig. 5

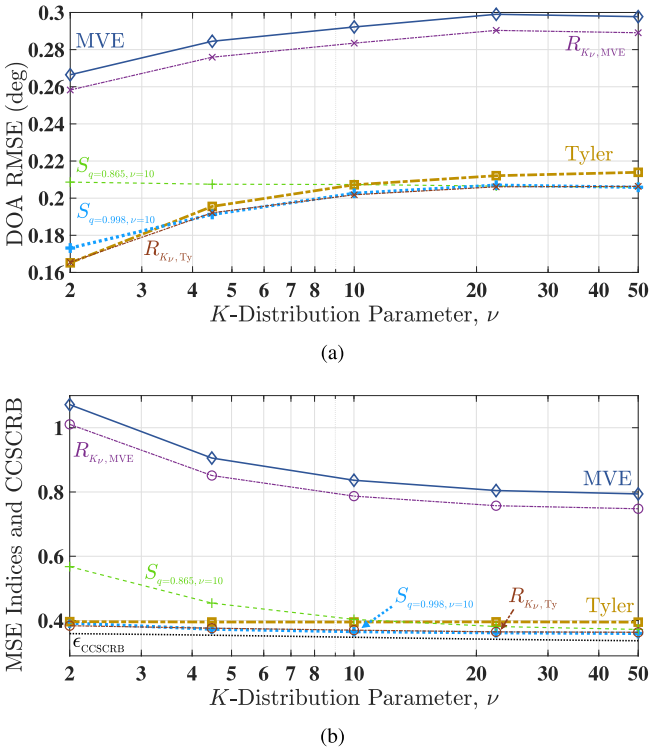



**FIGURE 5.** Root mean squared errors of uncontaminated direction-of-arrival estimates versus distribution parameter $\nu$ (a), and RMSE and CCSCRB metrics for noise-only scatter matrix estimates (b).

where we also included an $S_q$-estimator where we set $q = 0.998$. This estimator performed very well against the model mismatch.

Next, we added additional QPSK signals at $-37°$, $-30°$, $10°$, $30°$, and $40°$, each of equal power to the $20°$ signal. Using three contaminated observations, Fig. 6 depicts the normalized average achieved pseudospectrum of each estimator as well as the ideal MUSIC pseudospectrum obtained from the true shape matrix. Only the MVE and $S_q$-estimators were able to resolve all of the signals. The higher efficiency of the $S_q$-estimator resulted in taller and sharper peaks than the MVE estimator, and this corresponds to a higher probability of detection and higher DOA accuracy, respectively. The pseudospectrums from the $M_{\text{Ty}}$-estimator and the Tyler-initialized $R$-estimator did not resolve the two signals at $-37°$ and $-30°$. They also only exhibited three peaks for the four signals between $10°$ and $40°$, and none of the peaks coincided with any of the true signals. Finally, both the $M_{\text{Ty}}$-estimator and the Tyler-initialized $R$-estimator resulted in a false detection at $0°$, which was driven by the contaminated observations.

To characterize the variability of the pseudospectrum estimates, the plot in Fig. 6(c) shows the corresponding standard deviation of the achieved pseudospectrums in dB. For practically all angles, the $S_q$-estimators exhibited the lowest variability, and the MVE estimator generally produced the second-lowest variability. Tyler's $M$-estimator and the Tyler-initialized $R$-estimator exhibited the highest variability.

As observed above, when the $R$-estimator is initialized with a poor estimate of location and scatter, the resulting $R$-estimate is likewise of poor accuracy. Therefore, we tried initializing the $R$-estimator with the $S$-estimates. Fig. 6(b) is the same as Fig. 6(a) except that we include two $R$-estimators, one initialized with the MVE estimator and one with the $S_q$-estimator. For clarity, the figure is zoomed into the four signals between $10°$ and $40°$. We have also included the mismatched $S_q$-estimator. Then, to see the effects of sample size, Fig. 6(d) depicts the same simulation as in Fig. 6(b) but with $10\times$ the number of samples ($n = 30p$), and correspondingly, $10\times$ the number of outliers. For both the small- and large-sample cases, the relative results between the estimators is the same, and the larger sample size provides a better estimate of the shape matrix, resulting in taller and sharper pseudospectrum peaks. Both Fig. 6(b) and 6(d) illustrate the mismatched $S_q$-estimator performing approximately as well as the matched one—illustrating distributional robustness—with both $S_q$-estimators performing the best of the lot. Both the MVE-initialized $R$-estimator and the $S_q$-initialized $R$-estimator perform much better than the Tyler-initialized one, and the $S_q$-initialized $R$-estimator performed the best of the three. However, both the MVE- and $S_q$-initialized $R$-estimators demonstrate clear bias (of approximately $0.5° - 1°$) in the mean signal locations.

Finally, we test the estimators' ability to achieve high-resolution DOA estimates with the MUSIC algorithm. The angular resolution of this algorithm is driven by the efficiency of the shape matrix estimators as well as by the SNR. Here, we used the same estimators and configuration as we used for Fig. 6(b), except with only two signals, closely spaced at $3°$ apart ($10°$ and $13°$). Fig. 7(a) depicts the empirical probability of detecting both signals as a function of the SNR incident on the array. A detection was counted if the MUSIC algorithm resulted in both signal direction estimates being within $3°$ of their true direction. The relative estimator performance results of this simulation generally match those above when the SNR is below about 20 dB—the matched and mismatched $S_q$-estimators performing the best, and the $R$-estimator performing approximately as well as the estimator used to initialize it. However, around 30 dB, the $R$-estimator exhibits a strong instability that drives down its ability to resolve the two signals. Fig. 7(b) then depicts the results of the same simulation but without outliers. As expected, the $S$-estimators perform slightly better without outliers, and the $M_{\text{Ty}}$-estimator now provides valid estimates. Recalling that the $S_q$-estimator was tuned to match the $M_{\text{Ty}}$-estimator efficiency with a single signal at 3 dB SNR, we can see that their relative performance continues to track for varying SNR and numbers of signals. This indicates that the tuned estimator performance is largely invariant to different scenarios, meaning that the $S_q$-estimator can be tuned once and employed in a dynamic environment or for various scenarios.

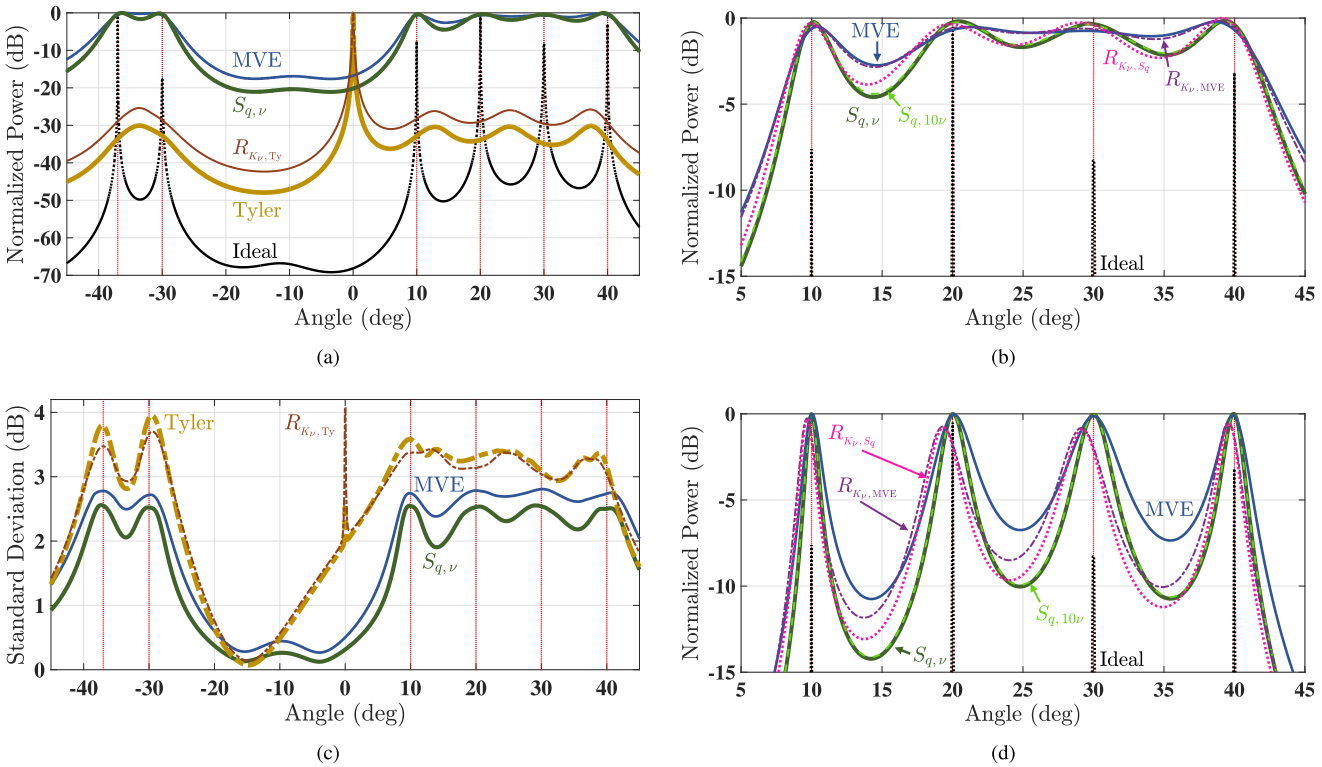



**FIGURE 6.** Small-sample normalized averages of the achieved MUSIC pseudospectrums (a), plot (a) but zoomed in with additional estimators (b), the corresponding standard deviation of the achieved pseudospectrums in dB (c), and large-sample normalized averages of the achieved MUSIC pseudospectrums (d). True signal locations are indicated with vertical lines.

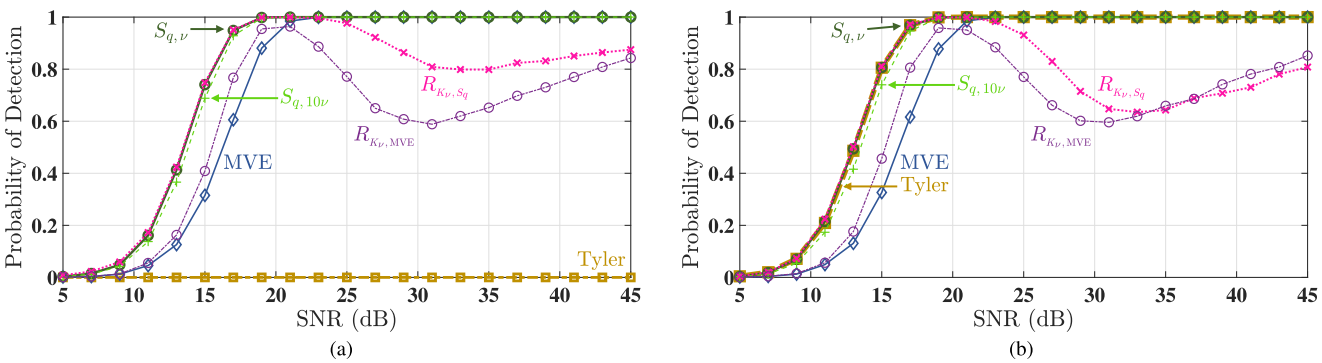



**FIGURE 7.** Probability of detecting two closely spaced signals as a function of SNR with outliers (a) and without outliers (b).

## VIII. CONCLUSION

This paper has provided an overview of multivariate *S*-estimators—a high-breakdown-point class of robust multivariate estimators of location and scatter that provide substantially higher robustness than *M*-estimators. The *S*-estimator class was extended to the complex-valued domain by updating the common $\rho$ functions and extending the class' theoretical properties. By exploring the theoretical breakdown point, and the theoretical weight and influence functions, the improved robustness of *S*-estimators over *M*-estimators was shown. Simulations demonstrated the practical benefits of *S*-estimators over *M*-estimator and *R*-estimators. While *M*-estimators—including the Huber and Tyler estimators—were demonstrated to break down with just a few outliers, *S*-estimators are able to ignore bad observations. With the recent development of the $S_q$-estimator, in many cases, *S*-estimator efficiency is now able to rival that of robust *M*-estimators for large and small dimensions, and non-Gaussian data. The $S_q$-estimator is also less sensitive to initial estimates [33] than traditional *S*-estimators, and it has been demonstrated here to be much less sensitive than the *R*-estimator. MATLAB code for the real- and complex-valued *S*-estimators discussed

in Section IV is provided on GitHub at https://github.com/JAFishbone/.

The properties presented here are also directly applicable to complex-valued multivariate *MM*-estimators. Introduced by Lopuhaä [51] and Tatsuoka and Tyler [52] for real-valued data, multivariate *MM*-estimators are an extension of multivariate *S*-estimators that provide additional flexibility by utilizing two $\rho$ functions. Extending the definition of *MM*-estimators to the complex-valued domain is trivial, and because *MM*-estimators inherit their properties from the *S*-estimator class [53], the properties derived above apply directly to complex-valued *MM*-estimators.

## APPENDIX

### A. PROOF OF LEMMA 2

#### 1) ASYMPTOTIC DISTRIBUTION OF $\widehat{\boldsymbol{\mu}}_n^u$, $\widehat{\boldsymbol{\mu}}_n^v$, $\widehat{\boldsymbol{\Sigma}}_n^u$, AND $\widehat{\boldsymbol{\Sigma}}_n^v$

Let us define the matrix $\widehat{\boldsymbol{W}}_u = \boldsymbol{\Sigma}_u^{-1/2}\widehat{\boldsymbol{\Sigma}}_n^u\boldsymbol{\Sigma}_u^{-1/2}$, and the vector $\boldsymbol{k}_i = \boldsymbol{\Sigma}_u^{-1/2}(\boldsymbol{u}_i - \boldsymbol{\mu}_u)/\sqrt{\sigma}$. Noting that $\widehat{\boldsymbol{\mu}}_n^u$ and $\widehat{\boldsymbol{W}}_u$ are consistent estimates of $\boldsymbol{\mu}_u$ and $\boldsymbol{I}$, respectively, we can write $\widehat{\boldsymbol{\mu}}_n^u = \boldsymbol{\mu}_u + \Delta\boldsymbol{\mu}_u$ and $\widehat{\boldsymbol{W}}_u = \boldsymbol{I} + \Delta\boldsymbol{\Sigma}_u$. It is easily verified that $d(\boldsymbol{u}_i, \widehat{\boldsymbol{\mu}}_n^u, \widehat{\boldsymbol{\Sigma}}_n^u)/\sigma = d(\boldsymbol{k}_i, \boldsymbol{\Sigma}_u^{-1/2}\Delta\boldsymbol{\mu}_u/\sqrt{\sigma}, \widehat{\boldsymbol{W}}_u)$. Left-multiplying (16), and left- and right-multiplying (17) by $\boldsymbol{\Sigma}_u^{-1/2}$, we obtain

$$\begin{aligned}
\boldsymbol{0} &= \frac{1}{n}\sum_{i=1}^{n} w_{\mathbb{R}}\left(d\left(\boldsymbol{k}_i, \boldsymbol{\Sigma}_u^{-1/2}\Delta\boldsymbol{\mu}_u/\sqrt{\sigma}, \widehat{\boldsymbol{W}}_u\right)\right) \\
&\qquad \times \left(\boldsymbol{k}_i - \boldsymbol{\Sigma}_u^{-1/2}\Delta\boldsymbol{\mu}_u/\sqrt{\sigma}\right), \\
\boldsymbol{0} &= \frac{1}{n}\sum_{i=1}^{n} w_{\mathbb{R}}\left(d\left(\boldsymbol{k}_i, \boldsymbol{\Sigma}_u^{-1/2}\Delta\boldsymbol{\mu}_u/\sqrt{\sigma}, \widehat{\boldsymbol{W}}_u\right)\right) \\
&\qquad \times \left(\boldsymbol{k}_i - \frac{\boldsymbol{\Sigma}_u^{-1/2}\Delta\boldsymbol{\mu}_u}{\sqrt{\sigma}}\right)\left(\boldsymbol{k}_i - \frac{\boldsymbol{\Sigma}_u^{-1/2}\Delta\boldsymbol{\mu}_u}{\sqrt{\sigma}}\right)^{\top} \\
&\qquad - \nu_{\mathbb{R}}\left(d\left(\boldsymbol{k}_i, \frac{\boldsymbol{\Sigma}_u^{-1/2}\Delta\boldsymbol{\mu}_u}{\sqrt{\sigma}}, \widehat{\boldsymbol{W}}_u\right)\right)\frac{(\boldsymbol{I} + \Delta\boldsymbol{\Sigma}_u)}{2p}.
\end{aligned}$$

The first-order Taylor expansion of these about $\Delta\boldsymbol{\mu}_u = 0$ and $\Delta\boldsymbol{\Sigma}_u = 0$ results in

$$\begin{aligned}
\boldsymbol{0} &= \frac{1}{n}\sum_{i=1}^{n} a_i\left(\boldsymbol{k}_i - \boldsymbol{\Sigma}_u^{-1/2}\Delta\boldsymbol{\mu}_u/\sqrt{\sigma}\right) \\
&\qquad + b_i\left(\boldsymbol{k}_i^{\top}\Delta\boldsymbol{\Sigma}_u\boldsymbol{k}_i + 2\boldsymbol{k}_i^{\top}\boldsymbol{\Sigma}_u^{-1/2}\Delta\boldsymbol{\mu}_u/\sqrt{\sigma}\right)\boldsymbol{k}_i, \\
\boldsymbol{0} &= \frac{1}{n}\sum_{i=1}^{n} a_i\left(\boldsymbol{k}_i\boldsymbol{k}_i^{\top} - \frac{\boldsymbol{k}_i\Delta\boldsymbol{\mu}_u^{\top}\boldsymbol{\Sigma}_u^{-1/2} + \boldsymbol{\Sigma}_u^{-1/2}\Delta\boldsymbol{\mu}_u\boldsymbol{k}_i^{\top}}{\sqrt{\sigma}}\right) \\
&\qquad + b_i\left(\boldsymbol{k}_i^{\top}\Delta\boldsymbol{\Sigma}_u\boldsymbol{k}_i + 2\boldsymbol{k}_i^{\top}\boldsymbol{\Sigma}_u^{-1/2}\Delta\boldsymbol{\mu}_u/\sqrt{\sigma}\right)\boldsymbol{k}_i\boldsymbol{k}_i^{\top} \\
&\qquad - (2p)^{-1}\nu_{\mathbb{R}}\left(\|\boldsymbol{k}_i\|^2\right)(\boldsymbol{I} + \Delta\boldsymbol{\Sigma}_u) \\
&\qquad + c_i\left(\boldsymbol{k}_i^{\top}\Delta\boldsymbol{\Sigma}_u\boldsymbol{k}_i + 2\boldsymbol{k}_i^{\top}\boldsymbol{\Sigma}_u^{-1/2}\Delta\boldsymbol{\mu}_u/\sqrt{\sigma}\right)\boldsymbol{I}/(2p),
\end{aligned}$$

where the scalars $a_i = w_{\mathbb{R}}(\|\boldsymbol{k}_i\|^2)$, $b_i = -w'_{\mathbb{R}}(\|\boldsymbol{k}_i\|^2)$, and $c_i = \nu'_{\mathbb{R}}(\|\boldsymbol{k}_i\|^2)$. Note the identities $\mathrm{vec}(\boldsymbol{k}\boldsymbol{k}^{\top}) = \boldsymbol{k}\otimes\boldsymbol{k}$, $\mathrm{vec}(\boldsymbol{X}\boldsymbol{Y}\boldsymbol{Z}) = (\boldsymbol{Z}^{\top}\otimes\boldsymbol{X})\,\mathrm{vec}(\boldsymbol{Y})$, and for any matrix $\boldsymbol{X}$ of size $p\times x$, $\boldsymbol{k}\otimes\boldsymbol{X} = (\boldsymbol{k}\otimes\boldsymbol{I}_p)\boldsymbol{X}$ for any $x$. We now take $\sqrt{n}\,\mathrm{vec}(\cdot)$ of the two above equations. Rearranging around $\Delta\boldsymbol{\mu}_u$ and $\Delta\boldsymbol{\Sigma}_u$, and accumulating the additional terms in $\boldsymbol{\chi}_n$ and $\boldsymbol{\omega}_n$, results in

$$\boldsymbol{\chi}_n = \boldsymbol{D}_n\sqrt{n}\boldsymbol{\Sigma}_u^{-1/2}\Delta\boldsymbol{\mu}_u + \boldsymbol{C}_n\sqrt{n}\,\mathrm{vec}\left(\Delta\boldsymbol{\Sigma}_u\right), \tag{24}$$

$$\boldsymbol{\omega}_n = (\boldsymbol{I}+\boldsymbol{K})\boldsymbol{B}_n\sqrt{n}\boldsymbol{\Sigma}_u^{-1/2}\Delta\boldsymbol{\mu}_u + \boldsymbol{A}_n\sqrt{n}\,\mathrm{vec}\left(\Delta\boldsymbol{\Sigma}_u\right), \tag{25}$$

where

$$\begin{aligned}
\boldsymbol{\chi}_n &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n} a_i\boldsymbol{k}_i, \qquad (26)\\
\boldsymbol{\omega}_n &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n} a_i\left(\boldsymbol{k}_i\otimes\boldsymbol{k}_i\right) - \frac{\nu_{\mathbb{R}}\left(\|\boldsymbol{k}_i\|^2\right)}{2p}\mathrm{vec}(\boldsymbol{I}), \\
\boldsymbol{D}_n &= \frac{1}{n\sqrt{\sigma}}\sum_{i=1}^{n} a_i\boldsymbol{I} - 2b_i\boldsymbol{k}_i\boldsymbol{k}_i^{\top}, \\
\boldsymbol{C}_n &= -\frac{1}{n}\sum_{i=1}^{n} b_i\boldsymbol{k}_i\left(\boldsymbol{k}_i\otimes\boldsymbol{k}_i\right)^{\top}, \\
\boldsymbol{B}_n &= \frac{1}{n\sqrt{\sigma}}\sum_{i=1}^{n}\left(\boldsymbol{k}_i\otimes\boldsymbol{I}\right)\left(a_i\boldsymbol{I} - b_i\boldsymbol{k}_i\boldsymbol{k}_i^{\top}\right) - \frac{c_i}{2p}\mathrm{vec}(\boldsymbol{I})\boldsymbol{k}_i^{\top}, \\
\boldsymbol{A}_n &= \frac{1}{n}\sum_{i=1}^{n}\frac{\nu_{\mathbb{R}}\left(\|\boldsymbol{k}_i\|^2\right)}{2p}\boldsymbol{I} - b_i\left(\boldsymbol{k}_i\otimes\boldsymbol{k}_i\right)\left(\boldsymbol{k}_i\otimes\boldsymbol{k}_i\right)^{\top} \\
&\qquad - \frac{c_i}{2p}\mathrm{vec}\left(\boldsymbol{I}\right)\left(\boldsymbol{k}_i\otimes\boldsymbol{k}_i\right)^{\top}. \qquad (27)
\end{aligned}$$

As discussed in [17], by defining $\boldsymbol{\kappa} = \boldsymbol{k}/\|\boldsymbol{k}\|$, we have $\boldsymbol{\kappa}\perp\boldsymbol{k}$, $\mathrm{E}[\boldsymbol{\kappa}] = \boldsymbol{0}$, $\mathrm{E}[\boldsymbol{\kappa}\boldsymbol{\kappa}^{\top}] = \boldsymbol{I}/(2p)$, and zero third-order moments. The only non-zero fourth-order moments are $\mathrm{E}[\kappa_i^4] = 3/(4p^2+4p)$ and $\mathrm{E}[\kappa_i^2\kappa_j^2] = 1/(4p^2+4p)$ for $i\neq j$. Therefore, the strong law of large numbers gives $\boldsymbol{A}_n \xrightarrow{p} \boldsymbol{A}$, $\boldsymbol{B}_n \xrightarrow{p} \boldsymbol{B}$, $\boldsymbol{C}_n \xrightarrow{p} \boldsymbol{C}$, and $\boldsymbol{D}_n \xrightarrow{p} \boldsymbol{D}$, with

$$\begin{aligned}
\sqrt{\sigma}\boldsymbol{D} &= \mathrm{E}\left[w_{\mathbb{R}}\left(\|\boldsymbol{k}\|^2\right)\right] + 2\,\mathrm{E}\left[\|\boldsymbol{k}\|^2 w'_{\mathbb{R}}\left(\|\boldsymbol{k}\|^2\right)\right]\mathrm{E}\left[\boldsymbol{\kappa}\boldsymbol{\kappa}^{\top}\right] \\
&= \mathrm{E}\left[w_{\mathbb{R}}\left(\|\boldsymbol{k}\|^2\right)\right] + p^{-1}\,\mathrm{E}\left[\|\boldsymbol{k}\|^2 w'_{\mathbb{R}}\left(\|\boldsymbol{k}\|^2\right)\right]\boldsymbol{I}, \\
\boldsymbol{C} &= \mathrm{E}\left[\|\boldsymbol{k}\|^3 w'_{\mathbb{R}}\left(\|\boldsymbol{k}\|^2\right)\right]\mathrm{E}\left[\boldsymbol{\kappa}\left(\boldsymbol{\kappa}\otimes\boldsymbol{\kappa}\right)^{\top}\right] \\
&= \boldsymbol{0}, \\
\sqrt{\sigma}\boldsymbol{B} &= \mathrm{E}\left[\|\boldsymbol{k}\| w_{\mathbb{R}}\left(\|\boldsymbol{k}\|^2\right)\right]\mathrm{E}\left[\boldsymbol{\kappa}\otimes\boldsymbol{I}\right] \\
&\qquad + \mathrm{E}\left[\|\boldsymbol{k}\|^3 w'_{\mathbb{R}}\left(\|\boldsymbol{k}\|^2\right)\right]\mathrm{E}\left[\left(\boldsymbol{\kappa}\otimes\boldsymbol{I}\right)\boldsymbol{\kappa}\boldsymbol{\kappa}^{\top}\right] \\
&\qquad - (2p)^{-1}\,\mathrm{E}\left[\|\boldsymbol{k}\|\nu'_{\mathbb{R}}\left(\|\boldsymbol{k}\|^2\right)\right]\mathrm{E}\left[\mathrm{vec}(\boldsymbol{I})\boldsymbol{\kappa}^{\top}\right] \\
&= \boldsymbol{0},
\end{aligned}$$

$$\begin{aligned}\boldsymbol{A} &= (2p)^{-1}\,\mathrm{E}\left[\nu_{\mathbb{R}}\left(\|\boldsymbol{k}\|^2\right)\right]\boldsymbol{I}\\ &\quad+\mathrm{E}\left[\|\boldsymbol{k}\|^4\,w'_{\mathbb{R}}\left(\|\boldsymbol{k}\|^2\right)\right]\mathrm{E}\left[(\boldsymbol{\kappa}\otimes\boldsymbol{\kappa})\,(\boldsymbol{\kappa}\otimes\boldsymbol{\kappa})^{\top}\right]\\ &\quad-(2p)^{-1}\,\mathrm{E}\left[\|\boldsymbol{k}\|^2\nu'_{\mathbb{R}}\left(\|\boldsymbol{k}\|^2\right)\right]\mathrm{E}\left[\mathrm{vec}(\boldsymbol{I})(\boldsymbol{\kappa}\otimes\boldsymbol{\kappa})^{\top}\right]\\ &=(2p)^{-1}\,\mathrm{E}\left[\nu_{\mathbb{R}}\left(\|\boldsymbol{k}\|^2\right)\right]\boldsymbol{I}\\ &\quad+\frac{\mathrm{E}\left[\|\boldsymbol{k}\|^4\,w'_{\mathbb{R}}\left(\|\boldsymbol{k}\|^2\right)\right]}{2p(2p+2)}\left(\boldsymbol{I}+\boldsymbol{K}+\mathrm{vec}(\boldsymbol{I})\,\mathrm{vec}(\boldsymbol{I})^{T}\right)\\ &\quad-(2p)^{-2}\,\mathrm{E}\left[\|\boldsymbol{k}\|^2\nu'_{\mathbb{R}}\left(\|\boldsymbol{k}\|^2\right)\right]\mathrm{E}\left[\mathrm{vec}(\boldsymbol{I})\,\mathrm{vec}(\boldsymbol{I})^{\top}\right].\end{aligned}$$

Note from (26) and (27) that $\boldsymbol{\chi}_n \xrightarrow{p} \boldsymbol{0}$ and $\boldsymbol{\omega}_n \xrightarrow{p} \boldsymbol{0}$ by the definition of the $S$-estimator solution equations in (16) and (17). Therefore, by the central limit theorem, $\boldsymbol{\chi}_n \xrightarrow{d} \boldsymbol{\chi}$ and $\boldsymbol{\omega}_n \xrightarrow{d} \boldsymbol{\omega}$, where both have zero-mean Gaussian distributions. By Slutsky's theorem and (24) and (25), we have $\sqrt{n}\Delta\boldsymbol{\mu}_u \xrightarrow{d} \boldsymbol{\Sigma}_u^{1/2}\boldsymbol{D}^{-1}\boldsymbol{\chi}$ and $\sqrt{n}\,\mathrm{vec}(\Delta\boldsymbol{\Sigma}_u) \xrightarrow{d} \boldsymbol{A}^{-1}\boldsymbol{\omega}$, which can be re-written as

$$\begin{aligned}\sqrt{n}\left(\widehat{\boldsymbol{\mu}}_n^{u}-\boldsymbol{\mu}_u\right) &\xrightarrow{d} \boldsymbol{\Sigma}_u^{1/2}\boldsymbol{D}^{-1}\boldsymbol{\chi},\\ \sqrt{n}\left(\widehat{\boldsymbol{\Sigma}}_n^{u}-\boldsymbol{\Sigma}_u\right) &\xrightarrow{d} \left(\boldsymbol{\Sigma}_u^{1/2}\otimes\boldsymbol{\Sigma}_u^{1/2}\right)\boldsymbol{A}^{-1}\boldsymbol{\omega}.\end{aligned}$$

Applying $\boldsymbol{P}$ gives

$$\begin{aligned}\sqrt{n}\left(\widehat{\boldsymbol{\mu}}_n^{v}-\boldsymbol{\mu}_v\right) &\xrightarrow{d} \boldsymbol{P}\boldsymbol{\Sigma}_u^{1/2}\boldsymbol{D}^{-1}\boldsymbol{\chi},\\ \sqrt{n}\left(\widehat{\boldsymbol{\Sigma}}_n^{v}-\boldsymbol{\Sigma}_v\right) &\xrightarrow{d} (\boldsymbol{P}\otimes\boldsymbol{P})\left(\boldsymbol{\Sigma}_u^{1/2}\otimes\boldsymbol{\Sigma}_u^{1/2}\right)\boldsymbol{A}^{-1}\boldsymbol{\omega}.\end{aligned}$$

#### 2) ASYMPTOTIC DISTRIBUTION OF $\widehat{\boldsymbol{\mu}}_n^{\mathbb{R}}$ AND $\widehat{\boldsymbol{\Sigma}}_n^{\mathbb{R}}$

From Lemma 1, $\boldsymbol{\Sigma}_{\mathbb{R}} = \boldsymbol{\Sigma}_u$ and $\boldsymbol{\mu}_{\mathbb{R}} = \boldsymbol{\mu}_u$, so we can rewrite $\boldsymbol{k}_i = \boldsymbol{\Sigma}_u^{-1/2}(\boldsymbol{u}_i - \boldsymbol{\mu}_u)/\sqrt{\sigma}$. Let us define $\widehat{\boldsymbol{W}}_{\mathbb{R}} = \boldsymbol{\Sigma}_{\mathbb{R}}^{-1/2}\widehat{\boldsymbol{\Sigma}}_n^{\mathbb{R}}\boldsymbol{\Sigma}_{\mathbb{R}}^{-1/2}$. Noting that $\widehat{\boldsymbol{\mu}}_n^{\mathbb{R}}$ and $\widehat{\boldsymbol{W}}_{\mathbb{R}}$ are consistent estimates of $\boldsymbol{\mu}_{\mathbb{R}}$ and $\boldsymbol{I}$, respectively, we can write $\widehat{\boldsymbol{\mu}}_n^{\mathbb{R}} = \boldsymbol{\mu}_{\mathbb{R}} + \Delta\boldsymbol{\mu}_{\mathbb{R}}$ and $\widehat{\boldsymbol{W}}_{\mathbb{R}} = \boldsymbol{I} + \Delta\boldsymbol{\Sigma}_{\mathbb{R}}$. Noting that $\boldsymbol{v} = \boldsymbol{P}\boldsymbol{u}$, it is easily verified that $d(\boldsymbol{u}_i, \widehat{\boldsymbol{\mu}}_n^{\mathbb{R}}, \widehat{\boldsymbol{\Sigma}}_n^{\mathbb{R}}) = d(\boldsymbol{v}_i, \boldsymbol{P}\widehat{\boldsymbol{\mu}}_n^{\mathbb{R}}, \widehat{\boldsymbol{\Sigma}}_n^{\mathbb{R}}) = \sigma\, d(\boldsymbol{k}_i, \boldsymbol{\Sigma}_{\mathbb{R}}^{-1/2}\Delta\boldsymbol{\mu}_{\mathbb{R}}/\sqrt{\sigma}, \widehat{\boldsymbol{W}}_{\mathbb{R}})$. Left-multiplying (14), and left- and right-multiplying (15) by $\boldsymbol{\Sigma}_{\mathbb{R}}^{-1/2}$, we obtain

$$\begin{aligned}\boldsymbol{0} &= \frac{1}{n}\sum_{i=1}^{n} w_{\mathbb{R}}\left(d\left(\boldsymbol{k}_i, \boldsymbol{\Sigma}_{\mathbb{R}}^{-1/2}\Delta\boldsymbol{\mu}_{\mathbb{R}}/\sqrt{\sigma}, \widehat{\boldsymbol{W}}_{\mathbb{R}}\right)\right)\\ &\qquad\times\left(\boldsymbol{k}_i - \boldsymbol{\Sigma}_{\mathbb{R}}^{-1/2}\Delta\boldsymbol{\mu}_{\mathbb{R}}/\sqrt{\sigma}\right),\\ \boldsymbol{0} &= \frac{1}{2n}\sum_{i=1}^{n} w_{\mathbb{R}}\left(d\left(\boldsymbol{k}_i, \boldsymbol{\Sigma}_{\mathbb{R}}^{-1/2}\Delta\boldsymbol{\mu}_{\mathbb{R}}/\sqrt{\sigma}, \widehat{\boldsymbol{W}}_{\mathbb{R}}\right)\right)\\ &\qquad\times\left[\left(\boldsymbol{k}_i - \frac{\boldsymbol{\Sigma}_{\mathbb{R}}^{-1/2}\Delta\boldsymbol{\mu}_{\mathbb{R}}}{\sqrt{\sigma}}\right)\left(\boldsymbol{k}_i - \frac{\boldsymbol{\Sigma}_{\mathbb{R}}^{-1/2}\Delta\boldsymbol{\mu}_{\mathbb{R}}}{\sqrt{\sigma}}\right)^{\top}\right.\\ &\qquad\left.+\,\boldsymbol{P}\left(\boldsymbol{k}_i - \frac{\boldsymbol{\Sigma}_{\mathbb{R}}^{-1/2}\Delta\boldsymbol{\mu}_{\mathbb{R}}}{\sqrt{\sigma}}\right)\left(\boldsymbol{k}_i - \frac{\boldsymbol{\Sigma}_{\mathbb{R}}^{-1/2}\Delta\boldsymbol{\mu}_{\mathbb{R}}}{\sqrt{\sigma}}\right)^{\top}\boldsymbol{P}^{\top}\right]\\ &\qquad-\nu_{\mathbb{R}}\left(d\left(\boldsymbol{k}_i, \frac{\boldsymbol{\Sigma}_{\mathbb{R}}^{-1/2}\Delta\boldsymbol{\mu}_{\mathbb{R}}}{\sqrt{\sigma}}, \widehat{\boldsymbol{W}}_{\mathbb{R}}\right)\right)\frac{(\boldsymbol{I}+\Delta\boldsymbol{\Sigma}_{\mathbb{R}})}{p}.\end{aligned}$$

Proceeding as in the previous section by taking the Taylor expansion of these about $\Delta\boldsymbol{\mu}_{\mathbb{R}} = 0$ and $\Delta\boldsymbol{\Sigma}_{\mathbb{R}} = 0$, and then rearranging the result of $\sqrt{n}\,\mathrm{vec}(\cdot)$ gives

$$\begin{aligned}\boldsymbol{\chi}_n &= \boldsymbol{D}_n\sqrt{n}\boldsymbol{\Sigma}_{\mathbb{R}}^{-1/2}\Delta\boldsymbol{\mu}_{\mathbb{R}} + \boldsymbol{C}_n\sqrt{n}\,\mathrm{vec}\left(\Delta\boldsymbol{\Sigma}_{\mathbb{R}}\right),\\ (\boldsymbol{I}+\boldsymbol{p}\otimes\boldsymbol{p})\,\boldsymbol{\omega}_n &= (\boldsymbol{I}+\boldsymbol{K})\,(\boldsymbol{I}+\boldsymbol{p}\otimes\boldsymbol{p})\,\boldsymbol{B}_n\sqrt{n}\boldsymbol{\Sigma}_{\mathbb{R}}^{-1/2}\Delta\boldsymbol{\mu}_{\mathbb{R}}\\ &\quad+\left(\boldsymbol{A}_n+(\boldsymbol{p}\otimes\boldsymbol{p})\boldsymbol{A}_n(\boldsymbol{p}\otimes\boldsymbol{p})^{\top}\right)\sqrt{n}\,\mathrm{vec}\left(\Delta\boldsymbol{\Sigma}_{\mathbb{R}}\right),\end{aligned}$$

with $\boldsymbol{\chi}_n$, $\boldsymbol{\omega}_n$, $\boldsymbol{A}_n$, $\boldsymbol{B}_n$, $\boldsymbol{C}_n$, and $\boldsymbol{D}_n$ defined as before. Noting that $(\boldsymbol{A}_n + (\boldsymbol{p}\otimes\boldsymbol{p})\boldsymbol{A}_n(\boldsymbol{p}\otimes\boldsymbol{p})^{\top}) \xrightarrow{p} 2\boldsymbol{A}$ and $\boldsymbol{\Sigma}_{\mathbb{R}} = \boldsymbol{\Sigma}_u$, and applying Slutsky's theorem, results in

$$\begin{aligned}\sqrt{n}\left(\widehat{\boldsymbol{\mu}}_n^{\mathbb{R}}-\boldsymbol{\mu}_{\mathbb{R}}\right) &\xrightarrow{d} \boldsymbol{\Sigma}_u^{1/2}\boldsymbol{D}^{-1}\boldsymbol{\chi},\\ \sqrt{n}\left(\widehat{\boldsymbol{\Sigma}}_n^{\mathbb{R}}-\boldsymbol{\Sigma}_{\mathbb{R}}\right) &\xrightarrow{d} \frac{1}{2}\,(\boldsymbol{I}+\boldsymbol{p}\otimes\boldsymbol{p})\left(\boldsymbol{\Sigma}_u^{1/2}\otimes\boldsymbol{\Sigma}_u^{1/2}\right)\boldsymbol{A}^{-1}\boldsymbol{\omega}.\end{aligned}$$

Therefore, $\widehat{\boldsymbol{\mu}}_n^{\mathbb{R}} \overset{a}{\sim} \widehat{\boldsymbol{\mu}}_n^{u}$ and $\widehat{\boldsymbol{\Sigma}}_n^{\mathbb{R}} \overset{a}{\sim} \frac{1}{2}(\widehat{\boldsymbol{\Sigma}}_n^{u} + \widehat{\boldsymbol{\Sigma}}_n^{v})$.

Finally, noting that $\boldsymbol{k}_i$ are i.i.d., using (26) and (27), and again noting that the first- and third-order moments of $\boldsymbol{\kappa}$ are zero, we have

$$\begin{aligned}\mathrm{E}[\boldsymbol{\omega}\boldsymbol{\chi}^{\top}] &= \mathrm{E}\left[\|\boldsymbol{k}\|^3 w_{\mathbb{R}}^2\left(\|\boldsymbol{k}\|^2\right)\right]\mathrm{E}\left[(\boldsymbol{\kappa}\otimes\boldsymbol{\kappa})\,\boldsymbol{\kappa}^{\top}\right]\\ &\quad-p^{-1}\,\mathrm{E}\left[\|\boldsymbol{k}\|w_{\mathbb{R}}\left(\|\boldsymbol{k}\|^2\right)\nu_{\mathbb{R}}\left(\|\boldsymbol{k}\|^2\right)\right]\mathrm{E}\left[\mathrm{vec}(\boldsymbol{I})\boldsymbol{\kappa}^{\top}\right]\\ &= \boldsymbol{0}.\end{aligned}$$

Therefore, asymptotically, $\boldsymbol{\omega} \perp \boldsymbol{\chi}$, which gives $\widehat{\boldsymbol{\mu}}_n^{u} \perp \widehat{\boldsymbol{\Sigma}}_n^{u}$ and $\widehat{\boldsymbol{\mu}}_n^{\mathbb{R}} \perp \widehat{\boldsymbol{\Sigma}}_n^{\mathbb{R}}$.

### B. PROOF OF THEOREM 6

#### 1) INFLUENCE FUNCTION OF SCATTER

We first derive the scatter influence function assuming a standard spherical distribution $F = \mathrm{CES}(\boldsymbol{0}, \boldsymbol{I}, \varphi, p)$, and then generalize it to any $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$. By the strong law of large numbers, (4) becomes

$$\boldsymbol{\Sigma} = \frac{\mathrm{E}\left[w\left((\boldsymbol{x}-\boldsymbol{\mu})^{\mathsf{H}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})/\sigma\right)\boldsymbol{x}\boldsymbol{x}^{\mathsf{H}}/\sigma\right]}{\mathrm{E}\left[\nu\left((\boldsymbol{x}-\boldsymbol{\mu})^{\mathsf{H}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})/\sigma\right)/p\right]}, \tag{28}$$

which at distribution $F$ gives

$$\boldsymbol{\Sigma}_F = \frac{\mathrm{E}\left[w\left(\boldsymbol{x}^{\mathsf{H}}\boldsymbol{x}/\sigma\right)\boldsymbol{x}\boldsymbol{x}^{\mathsf{H}}/\sigma\right]}{\mathrm{E}\left[\nu\left(\boldsymbol{x}^{\mathsf{H}}\boldsymbol{x}/\sigma\right)/p\right]} = \boldsymbol{I}. \tag{29}$$

The $\epsilon$-contaminated distribution is $F_\epsilon = (1-\epsilon)F + \epsilon\Delta_z$. Applying $F_\epsilon$ gives

$$\boldsymbol{\Sigma}_F(F_\epsilon) = \frac{(1-\epsilon)\,\mathrm{E}\left[w\left(\frac{d_x(F_\epsilon)}{\sigma}\right)\frac{\boldsymbol{x}\boldsymbol{x}^{\mathsf{H}}}{\sigma}\right] + \epsilon w\left(\frac{d_z(F_\epsilon)}{\sigma}\right)\frac{\boldsymbol{z}\boldsymbol{z}^{\mathsf{H}}}{\sigma}}{(1-\epsilon)\,\mathrm{E}\left[\nu\left(d_x(F_\epsilon)/\sigma\right)/p\right] + \epsilon\nu\left(d_z(F_\epsilon)/\sigma\right)/p},$$

with $d_x(F_\epsilon) = \boldsymbol{x}^{\mathsf{H}}\boldsymbol{\Sigma}_F(F_\epsilon)^{-1}\boldsymbol{x}$ and $d_z(F_\epsilon) = \boldsymbol{z}^{\mathsf{H}}\boldsymbol{\Sigma}_F(F_\epsilon)^{-1}\boldsymbol{z}$. We now take $\partial(\cdot)/\partial\epsilon|_{\epsilon=0}$, apply (29), and use the shorthand notation $\boldsymbol{\Sigma}'_F(z) = \mathbf{IF}(z;\boldsymbol{\Sigma}_F, F)$. This yields

$$\boldsymbol{\Sigma}'_F(z) = \frac{\frac{\partial \mathrm{E}\left[w\left(\frac{d_x(F_\epsilon)}{\sigma}\right)\frac{\boldsymbol{x}\boldsymbol{x}^{\mathsf{H}}}{\sigma}\right]}{\partial\epsilon}\Big|_{\epsilon=0} + w\left(\frac{d_z(F)}{\sigma}\right)\frac{\boldsymbol{z}\boldsymbol{z}^{\mathsf{H}}}{\sigma}}{\mathrm{E}\left[\nu\left(d_x(F)/\sigma\right)/p\right]} - \frac{\frac{\partial \mathrm{E}\left[\nu(d_x(F_\epsilon)/\sigma)/p\right]}{\partial\epsilon}\Big|_{\epsilon=0} + \nu(d_z(F)/\sigma)/p}{\mathrm{E}\left[\nu\left(d_x(F)/\sigma\right)/p\right]}\boldsymbol{I}. \quad (30)$$

We now address the two remaining differential terms. Using the identity $\boldsymbol{I} = \boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}$, and taking its derivative at zero, $\partial(\boldsymbol{\Sigma}_F(F_\epsilon)^{-1}\boldsymbol{\Sigma}_F(F_\epsilon))/\partial\epsilon|_{\epsilon=0}$, it is found that $\partial(\boldsymbol{\Sigma}_F(F_\epsilon)^{-1})/\partial\epsilon|_{\epsilon=0} = -\boldsymbol{\Sigma}'_F(z)$. Therefore, $\partial d_x/\partial\epsilon|_{\epsilon=0} = -\boldsymbol{x}^{\mathsf{H}}\boldsymbol{\Sigma}'_F(z)\boldsymbol{x}$. Defining $\boldsymbol{u} = \boldsymbol{x}/\|\boldsymbol{x}\|$, we have $\boldsymbol{u} \perp \|\boldsymbol{x}\|$, so we can now write

$$\begin{aligned}
\frac{\partial \mathrm{E}\left[w\left(\frac{d_x(F_\epsilon)}{\sigma}\right)\frac{\boldsymbol{x}\boldsymbol{x}^{\mathsf{H}}}{\sigma}\right]}{\partial\epsilon}\Big|_{\epsilon=0} &= -\mathrm{E}\left[\frac{d_x(F)^2}{\sigma^2}w'\left(\frac{d_x(F)}{\sigma}\right)\right] \\
&\quad \times \mathrm{E}\left[\left(\boldsymbol{u}^{\mathsf{H}}\boldsymbol{\Sigma}'_F(z)\boldsymbol{u}\right)\boldsymbol{u}\boldsymbol{u}^{\mathsf{H}}\right], \\
\frac{\partial \mathrm{E}\left[\nu\left(d_x(F_\epsilon)/\sigma\right)\right]}{\partial\epsilon}\Big|_{\epsilon=0} &= -\mathrm{E}\left[\frac{d_x(F)}{\sigma}\nu'\left(\frac{d_x(F)}{\sigma}\right)\right] \\
&\quad \times \mathrm{E}\left[\boldsymbol{u}^{\mathsf{H}}\boldsymbol{\Sigma}'_F(z)\boldsymbol{u}\right].
\end{aligned}$$

For each of these two foregoing equations, the second expectation on the right-hand side can be replaced using [21]

$$\begin{aligned}
\mathrm{E}\left[\left(\boldsymbol{u}^{\mathsf{H}}\boldsymbol{\Sigma}'_F(z)\boldsymbol{u}\right)\boldsymbol{u}\boldsymbol{u}^{\mathsf{H}}\right] &= \frac{\boldsymbol{\Sigma}'_F(z) + \mathrm{Tr}\left(\boldsymbol{\Sigma}'_F(z)\right)\boldsymbol{I}}{p(p+1)}, \\
\mathrm{E}\left[\boldsymbol{u}^{\mathsf{H}}\boldsymbol{\Sigma}'_F(z)\boldsymbol{u}\right] &= \frac{\mathrm{Tr}\left(\boldsymbol{\Sigma}'_F(z)\right)}{p}.
\end{aligned}$$

Note that $\nu'(t) = tw'(t)$, and because $b = \mathrm{E}[\rho(d_x(F)/\sigma)]$, we also have $\mathrm{E}[\nu(d_x(F)/\sigma] = d_x(F)w(d_x(F)/\sigma)/\sigma$. Therefore, (30) becomes

$$\begin{aligned}
&\mathrm{E}\left[\frac{d_x(F)w\left(d_x(F)/\sigma\right)}{\sigma p}\right]\boldsymbol{\Sigma}'_F(z) \\
&\quad = -\mathrm{E}\left[\frac{d_x(F)^2}{\sigma^2}w'\left(\frac{d_x(F)}{\sigma}\right)\right]\frac{\boldsymbol{\Sigma}'_F(z) + \mathrm{Tr}\left(\boldsymbol{\Sigma}'_F(z)\right)\boldsymbol{I}}{p(p+1)} \\
&\qquad + w\left(\frac{d_z(F)}{\sigma}\right)\frac{\boldsymbol{z}\boldsymbol{z}^{\mathsf{H}}}{\sigma} - \frac{\nu\left(d_z(F)/\sigma\right)}{p}\boldsymbol{I} \\
&\qquad + \mathrm{E}\left[\frac{d_x(F)^2\,w'\left(d_x(F)/\sigma\right)}{\sigma^2 p}\right]\frac{\mathrm{Tr}\left(\boldsymbol{\Sigma}'_F(z)\right)}{p}\boldsymbol{I}.
\end{aligned}$$

Taking the trace and rearranging gives

$$\mathrm{Tr}\left(\boldsymbol{\Sigma}'_F(z)\right) = \frac{p\left(\rho\left(d_z(F)/\sigma\right) - b\right)}{\mathrm{E}\left[\sigma^{-1}d_x(F)w\left(d_x(F)/\sigma\right)\right]}.$$

Combining the previous two equations and simplifying yields

$$\begin{aligned}
\boldsymbol{\Sigma}'_F(z) &= \frac{\rho\left(d_z(F)/\sigma\right) - b}{\mathrm{E}\left[\sigma^{-1}d_x(F)w\left(d_x(F)/\sigma\right)\right]}\boldsymbol{I} \\
&\quad + \frac{2p(p+1)d_z(F)w\left(d_z(F)/\sigma\right)}{\sigma\lambda_1(F)}\left(\frac{\boldsymbol{z}\boldsymbol{z}^{\mathsf{H}}}{d_z(F)} - \frac{1}{p}\boldsymbol{I}\right),
\end{aligned}$$

where $\lambda_1(F)$ is given by (12). Leveraging the equivariance property of $S$-estimators and utilizing the factorization $\boldsymbol{\Sigma} = \boldsymbol{A}\boldsymbol{A}^{\mathsf{H}}$, we generalize this using $\boldsymbol{\Sigma}'_{F_{\boldsymbol{\Sigma}}}(z) = \boldsymbol{A}\boldsymbol{\Sigma}'_{F_{\boldsymbol{I}}}(\boldsymbol{A}^{-1}(z-\boldsymbol{\mu}))\boldsymbol{A}^{\mathsf{H}}$. The result is the equation stated in the theorem.

### 2) INFLUENCE FUNCTION OF LOCATION

Recall the real-valued representation of $\widehat{\boldsymbol{\mu}}$, $\widehat{\boldsymbol{\mu}}^{\mathbb{R}}$ from (14). Lopuhaä [24, eq. (5.7)] derived the influence function for real-valued $S$-estimators, giving

$$\mathbf{IF}_{\mathbb{R}}\left(z_u;\boldsymbol{\mu}_{\mathbb{R}}, F_{\mathbf{0}}^{\mathbb{R}}\right) = \frac{w_{\mathbb{R}}(\|z_u\|^2/\sigma)z_u}{\mathrm{E}\left[w_{\mathbb{R}}(d_u/\sigma) + (p\sigma)^{-1}d_u w'_{\mathbb{R}}(d_u/\sigma)\right]},$$

for $F_{\mathbf{0}}^{\mathbb{R}} = \mathrm{RES}(\mathbf{0}, \boldsymbol{I}, \phi, 2p)$, and where $z_u = h(\boldsymbol{z})$. Substituting the identities from the proof of Theorem 5, applying $\mathbf{IF} = \boldsymbol{g}^{\mathsf{H}}\mathbf{IF}_{\mathbb{R}}$ and $z = \boldsymbol{g}^{\mathsf{H}}z_u$, and applying the affine equivariance property, yields the equation stated in the theorem.

## REFERENCES


[1] S. Haykin, *Adaptive Filter Theory*, 5th ed. Upper Saddle River: Pearson Educ., 2014.

[2] H. L. *Van Trees, Optimum Array Processing*. Hoboken, NJ, USA: Wiley, 2002.

[3] A. De Maio and M. S. Greco, Eds., *Modern Radar Detection Theory*. Edison, NJ, USA: SciTech Publishing, 2016.

[4] S. Gazor and W. Zhang, "Speech probability distribution," *IEEE Signal Process. Lett.*, vol. 10, no. 7, pp. 204–207, Jul. 2003.

[5] K. D. Ward, C. J. Baker, and S. Watts, "Maritime surveillance radar. Part I: Radar scattering from the ocean surface," *IEE Proc. F. Radar Signal Process.*, vol. 137, no. 2, pp. 51–62, 1990.

[6] S. Kallummil and S. Kalyani, "Noise statistics oblivious GARD for robust regression with sparse outliers," *IEEE Trans. Signal Process.*, vol. 67, no. 2, pp. 383–398, Jan. 2019.

[7] M. G. Sánchez, L. De Haro, M. Calvo Ramón, A. Mansilla, C. Montero Ortega, and D. Oliver, "Impulsive noise measurements and characterization in a UHF digital TV channel," *IEEE Trans. Electromagn. Compat.*, vol. 41, no. 2, pp. 124–136, May 1999.

[8] M. L. Picciolo and K. Gerlach, "Reiterative median cascaded canceler for robust adaptive array processing," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 43, no. 2, pp. 428–442, Apr. 2007.

[9] J. Zhao, S. Wang, L. Mili, B. Amidan, R. Huang, and Z. Huang, "A robust state estimation framework considering measurement correlations and imperfect synchronization," *IEEE Trans. Power Syst.*, vol. 33, no. 4, pp. 4604–4613, Jul. 2018.

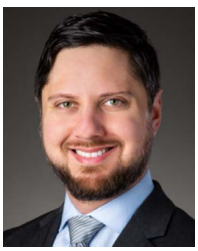
[10] C. S. Maiz, E. M. Molanes-Lopez, J. Miguez, and P. M. Djuric, "A particle filtering scheme for processing time series corrupted by outliers," *IEEE Trans. Signal Process.*, vol. 60, no. 9, pp. 4611–4627, Sep. 2012.

[11] E. Ollila, D. P. Palomar, and F. Pascal, "Shrinking the eigenvalues of M-estimators of covariance matrix," *IEEE Trans. Signal Process.*, vol. 69, pp. 256–269, 2021.

[12] B. Meriaux, C. Ren, A. Breloy, M. N. E. Korso, and P. Forster, "Matched and mismatched estimation of Kronecker product of linearly structured scatter matrices under elliptical distributions," *IEEE Trans. Signal Process.*, vol. 69, pp. 603–616, 2021.

[13] C. Schroth and M. Muma, "Robust M-estimation based Bayesian cluster enumeration for real elliptically symmetric distributions," *IEEE Trans. Signal Process.*, vol. 69, pp. 3525–3540, 2021.

[14] C. Schroth and M. Muma, "Real elliptically skewed distributions and their application to robust cluster analysis," *IEEE Trans. Signal Process.*, vol. 69, pp. 3947–3962, 2021.
[15] G. Draskovic, A. Breloy, and F. Pascal, "On the asymptotics of Maronna's robust PCA," *IEEE Trans. Signal Process.*, vol. 67, no. 19, pp. 4964–4975, Oct. 2019.
[16] S. Fortunati, F. Gini, M. S. Greco, A. M. Zoubir, and M. Rangaswamy, "Semiparametric CRB and Slepian-Bangs formulas for complex elliptically symmetric distributions," *IEEE Trans. Signal Process.*, vol. 67, no. 20, pp. 5352–5364, Oct. 2020.
[17] B. Mériaux, C. Ren, M. N. El Korso, A. Breloy, and P. Forster, "Asymptotic performance of complex M-estimators for multivariate location and scatter estimation," *IEEE Signal Process. Lett.*, vol. 26, no. 2, pp. 367–371, Feb. 2019.
[18] G. Draskovic and F. Pascal, "New insights into the statistical properties of M-estimators," *IEEE Trans. Signal Process.*, vol. 66, no. 16, pp. 4253–4263, Aug. 2018.
[19] Y. Pan, F. Duan, F. Chapeau-Blondeau, and D. Abbott, "Noise enhancement in robust estimation of location," *IEEE Trans. Signal Process.*, vol. 66, no. 8, pp. 1953–1966, Apr. 2018.
[20] M. Mahot, F. Pascal, P. Forster, and J.-P. Ovarlez, "Asymptotic properties of robust complex covariance matrix estimates," *IEEE Trans. Signal Process.*, vol. 61, no. 13, pp. 3348–3356, Jul. 2013.
[21] E. Ollila, D. E. Tyler, V. Koivunen, and H. V. Poor, "Complex elliptically symmetric distributions: Survey, new results and applications," *IEEE Trans. Signal Process.*, vol. 60, no. 11, pp. 5597–5625, Nov. 2012.
[22] E. Ollila and V. Koivunen, "Influence function and asymptotic efficiency of scatter matrix based array processors: Case MVDR beamformer," *IEEE Trans. Signal Process.*, vol. 57, no. 1, pp. 247–259, Jan. 2009.
[23] P. L. Davies, "Asymptotic behaviour of S-estimates of multivariate location parameters and dispersion matrices," *Ann. Statist.*, vol. 15, no. 3, pp. 1269–1292, 1987.
[24] H. P. Lopuhaä, "On the relation between S-estimators and M-estimators of multivariate location and covariance," *Ann. Statist.*, vol. 17, no. 4, pp. 1662–1683, 1989.
[25] D. M. Rocke, "Robustness properties of S-estimators of multivariate location and shape in high dimension," *Ann. Statist.*, vol. 24, no. 3, pp. 1327–1345, 1996.
[26] R. A. Maronna and V. J. Yohai, "Robust and efficient estimation of multivariate scatter and location," *Comput. Statist. Data Anal.*, vol. 109, pp. 64–75, 2017.
[27] R. A. Maronna, "Robust M-estimators of multivariate location and scatter," *Ann. Statist.*, vol. 4, no. 1, pp. 51–67, 1976.
[28] P. Rousseeuw and V. Yohai, "Robust regression by means of S-estimators," in *Robust and Nonlinear Time Series Analysis*. Berlin, Germany: Springer, 1984, pp. 256–272.
[29] R. A. Maronna, R. D. Martin, V. J. Yohai, and M. Salibián-Barrera, *Robust Statistics Theory and Methods (with R)*, 2nd ed. Hoboken, NJ, USA: Wiley, 2019.
[30] L. Luo, S. Bao, and X. Peng, "Robust monitoring of industrial processes using process data with outliers and missing values," *Chemometrics Intell. Lab. Syst.*, vol. 192, 2019, Art. no. 103827.
[31] A. D. Shieh and Y. S. Hung, "Detecting outlier samples in microarray data," *Stat. Appl. Genet. Mol. Biol.*, vol. 8, no. 1, 2009, Art. no. 13.
[32] J. A. Fishbone, "Highly robust and efficient estimators of multivariate location and covariance with applications to array processing and financial portfolio optimization," Ph.D. dissertation, Bradley Dept. Elect. Comput. Eng., Virginia Polytechnic Inst. State Univ., Falls Church, VA, USA, 2021.
[33] J. A. Fishbone and L. Mili, "New highly efficient high-breakdown estimator of multivariate scatter and location for elliptical distributions," *Can. J. Statist.*, early access, Apr. 16, 2023, doi: 10.1002/cjs.11770.
[34] B. Mériaux, C. Ren, M. N. El Korso, A. Breloy, and P. Forster, "Robust estimation of structured scatter matrices in (mis)matched models," *Signal Process.*, vol. 165, pp. 163–174, 2019.
[35] K.-T. Fang, S. Kotz, and K. W. Ng, *Symmetric Multivariate and Related Distributions*. London, U.K.: Chapman and Hall, 1990.
[36] R. A. Maronna, R. D. Martin, and V. J. Yohai, *Robust Statistics: Theory and Methods*, 1st ed. Hoboken, NJ, USA: Wiley, 2006.
[37] H. P. Lopuhaä, "Asymptotic expansion of S-estimators of location and covariance," *Statistica Neerlandica*, vol. 51, no. 2, pp. 220–237, 1997.
[38] H. P. Lopuhaä and P. J. Rousseeuw, "Breakdown points of affine equivariant estimators of multivariate location and covariance matrices," *Ann. Statist.*, vol. 19, no. 1, pp. 229–248, 1991.
[39] A. M. Zoubir, V. Koivunen, E. Ollila, and M. Muma, *Robust Statistics for Signal Processing*. Cambridge, U.K.: Cambridge Univ. Press, 2018.
[40] M. A. Gandhi and L. Mili, "Robust Kalman filter based on a generalized maximum-likelihood-type estimator," *IEEE Trans. Signal Process.*, vol. 58, no. 5, pp. 2509–2520, May 2010.
[41] F. R. Hampel, E. M. Ronchetti, P. J. Rousseeuw, and W. A. Stahel, *Robust Statistics: The Approach Based on Influence Function*. Hoboken, NJ, USA: Wiley, 1986.
[42] J. T. Kent and D. E. Tyler, "Redescending M-estimates of multivariate location and scatter," *Ann. Statist.*, vol. 19, no. 4, pp. 2102–2119, 1991.
[43] J. Capon, "High-resolution frequency-wavenumber spectrum analysis," *Proc. IEEE*, vol. 57, no. 8, pp. 1408–1418, 1969.
[44] M. Hallin and D. Paindaveine, "Parametric and semiparametric inference for shape: The role of the scale functional," *Statist. Decisions*, vol. 24, no. 3, pp. 327–350, 2006.
[45] S. Fortunati, A. Renaux, and F. Pascal, "Robust semiparametric efficient estimators in complex elliptically symmetric distributions," *IEEE Trans. Signal Process.*, vol. 68, pp. 5003–5015, 2020.
[46] S. Fortunati, "Computationally efficient version of the R-estimator," 2022. [Online]. Available: https://github.com/StefanoFor/
[47] R. O. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. Antennas Propag.*, vol. 34, no. 3, pp. 276–280, Mar. 1986.
[48] H. Krim and M. Viberg, "Two decades of array signal processing research: The parametric approach," *IEEE Signal Process. Mag.*, vol. 13, no. 4, pp. 67–94, Jul. 1996.
[49] D. E. Tyler, "A distribution-free M-estimator of multivariate scatter," *Ann. Statist.*, vol. 15, no. 1, pp. 234–251, 1987.
[50] S. Fortunati, A. Renaux, and F. Pascal, "Joint estimation of location and scatter in complex elliptically symmetric distributions: A robust semiparametric and computationally efficient R-estimator of the shape matrix," *J. Signal Process. Syst.*, vol. 94, no. 2, pp. 133–146, 2022.
[51] H. P. Lopuhaä, "Highly efficient estimators of multivariate location with high breakdown point," *Ann. Statist.*, vol. 20, no. 1, pp. 398–413, 1992.
[52] K. S. 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.
[53] P. Rousseeuw and M. Hubert, "High-breakdown estimators of multivariate location and scatter," in *Robustness and Complex Data Structures*, C. Becker, R. Fried, and S. Kuhnt, Eds. Berlin Germany: Springer, 2013, ch. 4, pp. 49–66.



**JUSTIN A. FISHBONE** (Member, IEEE) received the B.S. degree in mechanical engineering and the M.Eng. degree in aerospace engineering from Cornell University, Ithaca, NY, USA, in 2009 and 2010, respectively, and the Ph.D. degree in electrical engineering from Virginia Polytechnic Institute and State University, Falls Church, VA, USA, in 2021. He is currently a Senior Principal Systems Engineer with Raytheon Technologies, Chantilly, VA, USA, where he has been since 2015. From 2008 to 2015, he was with ManTech International Corporation. From 2006 to 2008, he was with MITRE Corporation.

**LAMINE MILI** (Life Fellow, IEEE) received the Electrical Engineering Diploma from the Swiss Federal Institute of Technology Lausanne, Lausanne, Switzerland, in 1976, and the Ph.D. degree from the University of Liège, Liège, Belgium, in 1987. He is currently a Professor of electrical and computer engineering with Virginia Polytechnic Institute and State University, Falls Church, VA, USA. His research interests include robust statistics, robust signal processing, radar systems, and power systems analysis and control.