[Abdelhak M. Zoubir, Visa Koivunen, Yacine Chakhchoukh, and Michael Muma]

Robust Estimation in Signal Processing

[A tutorial-style treatment of fundamental concepts]

The word *robust* has been used in many contexts in signal processing. Our treatment concerns statistical robustness, which deals with deviations from the distributional assumptions. Many problems encountered in engineering practice rely on the Gaussian distribution of the data, which in many situations is well justified. This enables a simple derivation of optimal estimators. Nominal optimality, however, is useless if the estimator was derived under distributional assumptions on the noise and the signal that do not hold in practice. Even slight deviations from the assumed distribution may cause the estimator's performance to

Digital Object Identifier 10.1109/MSP.2012.2183773
Date of publication: 12 June 2012

drastically degrade or to completely break down. The signal processing practitioner should, therefore, ask whether the performance of the derived estimator is acceptable in situations where the distributional assumptions do not hold. Isn't it robustness that is of a major concern for engineering practice? Many areas of engineering today show that the distribution of the measurements is far from Gaussian as it contains outliers, which cause the distribution to be heavy tailed. Under such scenarios, we address single and multichannel estimation problems as well as linear univariate regression for independently and identically distributed (i.i.d.) data. A rather extensive treatment of the important and challenging case of dependent data for the signal processing practitioner is also included. For these problems, a comparative analysis of the most important robust methods is carried out by evaluating their performance theoretically, using simulations as well as real-world data.

INTRODUCTION

Statistical signal processing often relies on strong assumptions, i.e., optimal estimators, detectors, and filters are derived based, e.g., on a particular parametric signal model, a probability distribution of the sensor noise (typically the normal distribution [1]), or the assumptions of stationarity, linearity, or independence and identical distribution of random variables. Optimality, however, is only achieved if the underlying assumptions hold and the performance of optimal procedures may deteriorate significantly, even for minor departures from the assumed model.

MANY AREAS OF ENGINEERING TODAY SHOW THAT THE DISTRIBUTION OF THE MEASUREMENTS IS FAR FROM GAUSSIAN AS IT CONTAINS OUTLIERS, WHICH CAUSE THE DISTRIBUTION TO BE HEAVY TAILED.

In particular, measurement campaigns [2]–[5] have confirmed the presence of impulsive (heavy-tailed) noise, which can cause optimal signal processing techniques, especially the ones derived using the nominal Gaussian probability model, to be biased or to even break down. The occurrence of impulsive noise has been reported, for example, in outdoor mobile communication channels, due to switching transients in power lines or automobile ignition [4], in radar and sonar systems as a result of natural or man-made electromagnetic and acoustic interference [3], [5], and in indoor wireless communication channels, owing, e.g., to microwave ovens and devices with electromechanical switches, such as electric motors in elevators, printers, and copying machines [2]. Moreover, biomedical sensor array measurements of the brain activity, such as in magnetic resonance imaging (MRI) were found to have non-Gaussian noise and interference in various regions of the human brain, where the complex tissue structure is known to exist [6]. In geolocation position estimation and tracking, nonline-of-sight (NLOS) signal propagation, caused by obstacles such as buildings or trees, results in outliers in the measurements, to which conventional position estimation methods are very sensitive [7]. In classical short-term load forecasting, the prediction accuracy is adversely influenced by outliers, which are caused by nonworking days or exceptional events such as strikes, soccer's World Cup, or natural disasters [8]. Moreover, on a computer platform, various components, such as the liquid crystal display (LCD) pixel clock and the peripheral component interconnect (PCI) express bus, cause impulsive interference that degrades the performance of the embedded wireless devices [9].

These situations enforce the need for robust signal processing methods, which are close to optimal in nominal conditions and highly reliable for real-life data, even if the assumptions are only approximately valid [10]. In estimation, optimality under the exact model is commonly measured by the efficiency of the estimator, while near-optimality under contamination is displayed by measures of its resistance, reliability or robustness.

Early contributions to robust estimation can be traced back to the 1800s. They include Laplace's work on linear functions of order statistics, Legendre's comments on the necessity of rejecting outliers to provide stability to the least-squares method, and Newcomb's proposal at the end of the 19th century, to model heavy-tailed distributions as mixtures of normal densities. All of these examples provide evidence that there was awareness about what we today call the "robustness" of an estimator for at least two centuries [11]. The first formal theory of robustness was proposed in 1964 by Huber in response to an article by Tukey [10]. A second theory of robustness was later developed by Hampel in 1968, who introduced important concepts, such as the influence function (IF) and the breakdown point (BP) [12].

In engineering, robust estimators and detectors have been of interest since the early days of digital signal processing; see the review paper on robust methods published by Kassam and Poor in 1985 [13] and references therein. There are also excellent textbooks that consider robust estimation for signal processing [14], [15]. There has been much research activity on robust methods in image processing and computer vision; see [16] and [17] and references therein. Robustness has been an important topic in control engineering as well, see, e.g., [18] for further references.

The complexity of new engineering problems suggest the urgent need to revisit robust estimation techniques in an accessible manner for the signal processing practitioner with ample real-life examples. Further, one of our aims is to illustrate recent developments in robust statistics for dependent data [19], which have not been considered much in the engineering literature.

We emphasize that the robustness we treat in this article is that of analyzing the impact on statistical methods caused by a discrepancy between the (statistical) modeling assumptions and reality. It provides methods that tradeoff some efficiency at the nominal model to gain resistance against the effects of deviations.

The nonrobustness of the maximum likelihood estimator (MLE) under the Gaussian noise assumption in the presence of non-Gaussian or heavy-tailed noise is well known and there exist different approaches to compensate for it. A frequently used and successful engineering strategy to robustify the MLE under the

Gaussian assumption consists in applying preprocessing steps, i.e., detecting and rejecting suspicious observations. Common approaches include trimming some percentage of the data or rejecting observations, which exceed a few hopefully robustly estimated standard deviations. More complex applications arise in the fields of machine learning and data mining. Here, large outliers/anomalies/deviations are sought for in often high-dimensional data sets. One can distinguish between distribution- and distance-based methods. The first approach consists of learning the probability distributions of the data, which are modeled as a mixture of outliers and clean data [20]. A second approach is to treat outlier detection as a combinatorial problem [21]. Here, a measure of similarity between all data points is calculated and outliers are those points which are least similar to their neighbors. In both cases, ideas from robust statistics can be introduced to ignore outliers. This allows for identifying simpler and more accurate patterns, which capture the majority of the data into distinguishable clusters. Based on these clusters, outliers can be singled out and, either discarded as erroneous data, or even in some cases identified as the most informative and interesting data.

Because of limited space, we do not treat outlier diagnostics here and refer the interested reader to the seminal work of Rousseeuw and Leroy [22] and some recent advances [20], [21], [23]. Neither do we treat α-stable distributions, which have been successfully applied to model impulsiveness in real-world signal processing applications [24]. They can model a wide range of signals and have the advantageous property that the output of a linear system with α-stable input is again α-stable. A drawback is that closed-form analytical expressions for the probability density function do not generally exist [24]. Bayesian methods for robust estimation are not treated. The interested reader is referred to [25] and references therein for a recent overview of Bayesian robust estimation.

BASIC CONCEPTS OF ROBUSTNESS

The signal processing practitioner interested in applying robust methods should be familiar with some concepts of robust statistics, which we revisit in this section.

THE BREAKDOWN POINT AND QUANTITATIVE ROBUSTNESS

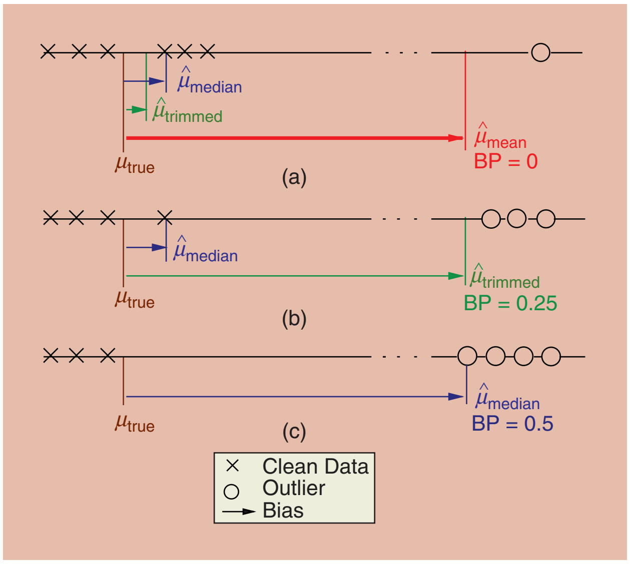
The BP is used to characterize quantitative robustness of an estimator. It indicates the maximal fraction of outliers (highly deviating samples) in the observations, which an estimator can handle without breaking down. The BP takes values between 0 and 50%, where a higher BP value corresponds to larger quantitative robustness. The BP of the sample mean is zero, which means that a single outlier may throw the estimator completely off, while for the sample median it is 50%. Beyond 50%, one cannot distinguish between the nominal and contaminating distributions. Figure 1 illustrates the concept of the BP for the sample mean, the α-trimmed mean and the sample median. In the sections "Robust Estimators for Multichannel Data" and "Robust Estimators for Dependent Data," we give extensions and examples of the BP for multichannel and dependent data, respectively.

THE INFLUENCE FUNCTION AND QUALITATIVE ROBUSTNESS

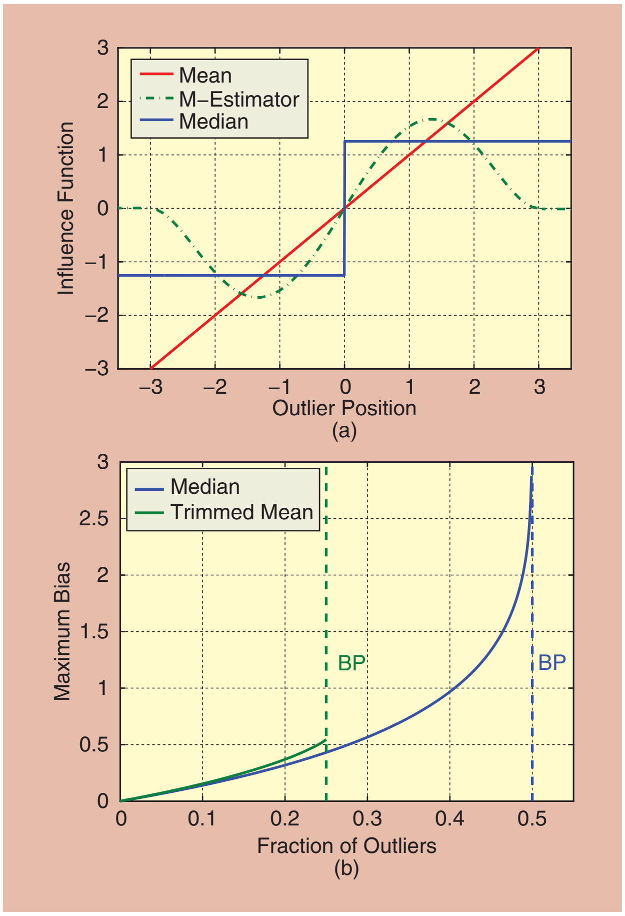
The IF describes the bias impact of infinitesimal contamination at an arbitrary point on the estimator, standardized by the fraction of contamination. When the limit exists, the asymptotic IF, which is basically the first derivative of the functional version of an estimator $\hat{\theta}$ at a nominal distribution F_θ, is defined by

$$\mathrm{IF}(z;\hat{\theta},F_\theta)=\lim_{\epsilon\to 0}\frac{\hat{\theta}_\infty(G)-\hat{\theta}_\infty(F_\theta)}{\epsilon}=\left[\frac{\partial\hat{\theta}_\infty(G)}{\partial\epsilon}\right]_{\epsilon\to 0},$$

where $\hat{\theta}_\infty(F_\theta)$ and $\hat{\theta}_\infty(G)$ are the asymptotic values of the estimator when the data is distributed, following, respectively, F_θ and the contaminated distribution $G=(1-\epsilon)F_\theta+\epsilon\Delta_z$ with Δ_z being the point-mass probability on z and ϵ the fraction of contamination. The IF is depicted with respect to z, the position of the infinitesimal contamination. Desirable properties of the IF are boundedness and continuity. Boundedness ensures that a small fraction of contamination or outliers can have only a limited effect on the estimate, whereas continuity means that small changes in the data lead to small changes in the estimate. If both properties are satisfied, an estimator is considered to be qualitatively robust against infinitesimal contamination. Figure 2(a) illustrates the IFs of three estimators of location for the standard normal distribution $\mathcal{N}(0,1)$. In the sections "Robust Estimators for Multichannel Data" and "Robust Estimators for



[FIG1] The bias and BP of three traditional estimators of location μ, i.e., (a) the sample mean (red), (b) the α-trimmed mean (green), and (c) the sample median (blue). "Clean" observations are depicted as crosses and outliers as circles. The BP of the sample mean is zero, which means that a single outlier has an unbounded effect on its bias. The α-trimmed mean, where α = 0.25 for this example, resists one outlier by ignoring the largest and smallest 25% of the data. The BP of the sample median equals the highest possible value of 50%, which means that its bias remains bounded even in situations when up to half of the observations, in this case three, are replaced by arbitrarily large values.



[FIG2] The (a) IFs of three estimators of location for the standard normal distribution $\mathcal{N}(0, 1)$. The IF of the sample mean is unbounded. The IF of the sample median is bounded but not continuous at the origin. The IF of the robust M-estimator (to be introduced in the section "Robust Estimators for Single-Channel Data") is bounded, continuous, and redescending, which means that the estimator is qualitatively robust. For this estimator, the impact of infinitesimal large-valued outliers on the estimate is suppressed. (b) The MBC for the sample median and the α-trimmed mean, with $\alpha = 0.25$, at $\mathcal{N}(0, 1)$. The maximum bias is infinite beyond the BP.

Dependent Data," we give extensions and examples of the IF for multichannel and dependent data, respectively.

THE MAXIMUM BIAS CURVE

The maximum bias curve (MBC) gives information on the bias affected by a specific amount of contamination. It plots the absolute value of the maximum possible asymptotic bias $b_{\hat{\theta}}(F,\theta) = \hat{\theta}_\infty(F_\theta) - \theta$ of an estimator $\hat{\theta}$ with respect to the fraction of contamination ϵ, whereby the most robust one minimizes the MBC. Formally, it is defined as

$$\mathrm{MBC}(\epsilon, \theta) = \max\{|b_{\hat{\theta}}(F, \theta)| : F \in \mathcal{F}_{\epsilon,\theta}\},$$

where $\mathcal{F}_{\epsilon,\theta} = \{(1-\epsilon)F_\theta + \epsilon G\}$ is an ϵ-neighborhood of distributions around the nominal distribution F_θ with G being an arbitrary contaminating distribution. Figure 2(b) plots the MBCs for the sample median and the α-trimmed mean at $\mathcal{N}(0, 1)$. In the section "Robust Estimators for Dependent Data," we discuss the MBC for dependent data.

ROBUST ESTIMATORS FOR SINGLE-CHANNEL DATA

Consider a situation, where we have measurements y_n, $n = 1, \ldots, N$, which describe a direct current (DC) voltage in noise. Let us assume that the noise process consists of independent random variates of an identical and unknown distribution. The model for the observations can be written as follows:

$$Y_n = \mu + V_n \quad n = 1, \ldots, N, \tag{1}$$

where Y_n, V_n are random variables and y_n, v_n are their respective realizations for $n = 1, \ldots, N$. The goal is to estimate the value μ from observations y_n, $n = 1, \ldots, N$. Here, the sample mean

$$\hat{\mu} = \frac{1}{N}\sum_{n=1}^{N} y_n, \tag{2}$$

which is the LSE, provides an intuitive solution to the problem, especially when the noise distribution is symmetric. Under the Gaussian assumption, which in some situations is justified, (2) is the MLE and is optimal in Fisher's sense.

The above problem is not limited to DC value estimation. Many engineering problems can be formulated as in (1) and the problem is known as estimation of location [13]. The LSE of location, i.e., the sample mean as in (2), gives all observations the same weight. Provided the Gaussian assumption is fulfilled, the approach is best in statistical performance. However, it is common in practice that the noise process is non-Gaussian [1] and may contain outliers [2]–[6]. Then, our intuition dictates that we weigh the observations y_n, $n = 1, \ldots, N$, in (2) such that we give more weight to data that is close to the measurement model as compared to the one that is unlikely to occur. This is what robust location estimation is about. In terms of the IF, one can state that the MLE of location under the Gaussian noise assumption results in a linear IF because the influence of a measurement on the bias of the estimate is proportional to its value as is shown in Figure 2(a). The M-estimators discussed below result in a nonlinear IF; see Figure 2(a).

M-ESTIMATOR

An important class of robust estimators are M-estimators [10]. They are well understood in terms of their statistical properties and can resist outliers without preprocessing the data. M-estimation is easily accessible in the single-channel context, but we will later show how this concept is applied to multichannel data and linear regression. M-estimators are a generalization of MLE. In the location case, they solve

$$\sum_{n=1}^{N} \psi(Y_n - \hat{\mu}) = 0,$$

where ψ is the derivative of the loss function ρ, which is to be minimized. The MLE under the Gaussian noise assumption corresponds to an M-estimator if $\rho(X) = X^2$, which results in a linear ψ-function. By selection of an appropriate ψ-function, M-estimators achieve high efficiency under the nominal model as well as qualitative robustness and a high BP. There exists a variety of ψ-functions which are grouped into the monotone and redescending class. Redescending M-Estimators are useful when the data contains extreme outliers and the asymptotic variance decreases at the cost of an increase in minimax risk [10]. For M-estimators, the IF at a nominal Gaussian, is proportional to their ψ-function, which allows studying many robustness properties, simply based on their ψ-function. For a detailed discussion of different ψ-functions, the interested reader is referred to [10, Ch. 4].

A way to intuitively understand M-estimates of location is to interpret them as a weighted average with weights given by

$$W(x) = \begin{cases} \psi(x)/x, & \text{if } x \neq 0 \\ \psi'(0), & \text{if } x = 0. \end{cases}$$

Usually, for robust M-estimators, the weights are chosen close to one for the bulk of the data, while outliers are increasingly downweighted. M-estimators of location can be conveniently computed via an iterative reweighting algorithm, with a previously computed robust scale estimate $\hat{\sigma}$, for example, (3), as follows:

Algorithm: M-estimator

Step 1) Estimate $\hat{\sigma}$ and initial $\hat{\mu}_o$.

Step 2) While $\dfrac{|\hat{\mu}_{k+1} - \hat{\mu}_k|}{\hat{\sigma}} < \varepsilon$:

$$w_{kn} = W\left(\frac{y_n - \hat{\mu}_k}{\hat{\sigma}}\right)$$

$$\hat{\mu}_{k+1} = \frac{\sum_{n=1}^{N} w_{kn} y_n}{\sum_{n=1}^{N} w_{kn}}$$

$$k \leftarrow k+1.$$

In principle, one could choose any of the available methods for equation solving, e.g., Newton-Raphson. However, methods based on derivatives may be unsafe with robust ψ-functions, (see, e.g., [10] and [19] for a discussion on alternative algorithms). For monotone ψ-functions, the initialization only influences the number of iterations required until convergence. For redescending M-estimators, simple robust initializations of location and scale, i.e., the median and normalized median absolute deviation

$$\hat{\sigma}_{\text{mad}}(\boldsymbol{x}) = 1.483 \cdot \text{median}(|\boldsymbol{x} - \text{median}(\boldsymbol{x})|) \qquad (3)$$

are necessary and sufficient to converge to a "good" solution, i.e., a normally distributed and highly efficient estimator $\hat{\mu}$ for symmetrically distributed data [19]. The normalization factor 1.483 has been introduced in (3) to make the estimator consistent for the standard deviation at the normal distribution. A frequently used redescending estimator is Tukey's biweight, whose IF is displayed in Figure 2(a). M-estimates, when computed as described below, are location equivariant, i.e., $\hat{\mu}(x+a) = \hat{\mu}(x) + a$ and their finite sample BP is given by $0.5 \cdot ((N-1)/N)$. For example, for $N = 50$ the BP is 0.49.

M-estimation of scale is not discussed here, due to space limitations. Intuitively, M-estimates of scale are representable as a weighted root-mean-square estimate that is quadratic near the origin and then increases less rapidly, therewith downweighting unusually large-valued data (outliers). As with location estimation, the recommended computation is done by iterative reweighting [19]. For further details on scale estimation, the interested reader is referred to [10, Ch. 5] and [26].

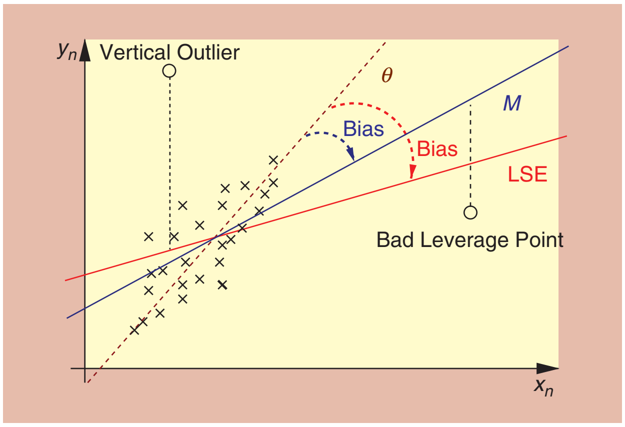
LINEAR REGRESSION MODELS

Many of today's engineering problems in areas as diverse as wireless communication [7], [27]–[30], ultrasonic systems [31], computer vision [16], [32], electric power systems [33], automated detection of defects [34], biomedical signal analysis [35], and many more, can be formulated as a linear regression for which the parameters of interest are sought. The linear regression model is given by

$$Y_n = \mathbf{X}_n^\top \boldsymbol{\theta} + V_n, \quad n = 1, \ldots, N, \qquad (4)$$

where we assume that the predictors $\mathbf{X}_n = (X_{1n}, \ldots, X_{pn})^\top$, the errors V_n and data points $(Y_n, \mathbf{X}_n^\top)$ for $n = 1, \ldots, N$, are i.i.d. random variables, $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)^\top$ are the unknown parameters of interest, and $\mathbf{X}_n$ and V_n are mutually independent. Herein, $^\top$ denotes transpose.

Figure 3 depicts a situation for a two vector-valued parameter $\boldsymbol{\theta}$, where we observe different kinds of outlying observations, which do not follow the linear pattern of the majority of the data. They are termed vertical outliers, when their $\boldsymbol{x}_n^\top$ is not outlying and leverage point if it is. The simple illustrating example suggests that outlying observations can be detected by visual inspection. Manual screening, however becomes infeasible for larger dimensions of the parameter vector. As depicted, the LSE is not robust against any type of outliers that contribute in an



[FIG3] Different types of outliers in regression and their effect on the bias of the LSE and on an M-estimator. The LSE is not robust against any type of outliers. M-estimators are robust against vertical outliers, but they become nonrobust with a BP equal to zero in the presence of leverage points.

unbounded fashion to the bias of the estimator. To overcome this problem, several robust estimators of $\boldsymbol{\theta}$ have been proposed, see, e.g., [10], [19], [22], [36]–[38], and references therein. We shall describe the ones whose statistical properties are well understood and for which implementations are available.

ANOTHER COMMON PROBLEM IN SIGNAL PROCESSING, WHICH MAY CONTAIN VERTICAL OUTLIERS, IS THE ESTIMATION OF THE PARAMETERS OF A SINUSOID IN CONTAMINATED NOISE.

When dealing with data that contains regular observations and vertical outliers only, M-estimators are robust in the sense described in the "Introduction" section. Analogously to the location case, discussed in the section "Basic Concepts of Robustness," M-estimators gain their robustness compared to the MLE under the Gaussian noise assumption, which coincides with the LSE, by downweighting vertical outliers with a bounded $\psi = \rho'$. For the LSE $\rho(x) = x^2$, while $\rho(x) = |x|$ gives the ℓ_1-estimator. If ρ is constant outside some interval, redescending ψ-functions are obtained. The M-estimate is a solution of

$$\sum_{n=1}^{N} \psi\left(\frac{\hat{v}_n}{\hat{\sigma}_V}\right) \boldsymbol{x}_n = \mathbf{0} \tag{5}$$

with residuals $\hat{v}_n = y_n - \boldsymbol{x}_n^{\top}\hat{\boldsymbol{\theta}}$ and corresponding robust scale estimate $\hat{\sigma}_V$. The solution is computed by iteratively reweighted least-squares with similar considerations concerning monotone or redescending ψ-function as described for the location case in the section "Robust Estimators for Single-Channel Data." When M-estimates are computed in this way, they are regression equivariant, i.e., $\hat{\boldsymbol{\theta}}(\mathbf{X}, \mathbf{Y} + \mathbf{Xu}) = \hat{\boldsymbol{\theta}}(\mathbf{X}, \mathbf{Y}) + \mathbf{u}$, where $\mathbf{u}$ is any $p \times 1$ vector, scale equivariant, i.e., $\hat{\boldsymbol{\theta}}(\mathbf{X}, \sigma\mathbf{Y}) = \sigma\hat{\boldsymbol{\theta}}(\mathbf{X}, \mathbf{Y})$, where σ is any scalar, and affine equivariant, i.e., $\hat{\boldsymbol{\theta}}(\mathbf{XA}, \mathbf{Y}) = \mathbf{A}^{-1}\hat{\boldsymbol{\theta}}(\mathbf{X}, \mathbf{Y})$, where $\mathbf{A}$ is any $p \times p$ nonsingular matrix. These are important properties, since they make the analysis of the estimator independent, e.g., of the scale of the variables as well as translations or rotations of the data. M-estimation has been successfully adapted to solve many engineering problems, e.g., geolocation of user equipment (UE) for wireless networks in NLOS environments [7], [30], multiuser detection in wireless communications [27]–[29], [39], and in computer vision [16].

Another common problem in signal processing, which may contain vertical outliers, is the estimation of the parameters of a sinusoid in contaminated noise. Robust estimation of complex sinusoids with known frequencies based on sensor measurements $Y_n, n = 1, \ldots, N$, can be formulated as a linear regression problem. Indeed, the observations follow the model

$$Y_n = \sum_{k=1}^{K} \theta_k e^{j\omega_k n} + V_n \quad n = 1, 2, \ldots, N, \tag{6}$$

where θ_k is the unknown complex amplitude of the sinusoid with known frequencies ω_k, $k = 1, \ldots, K$ and V_n, $n = 1, \ldots, N$ is contaminated Gaussian i.i.d. circular complex-valued noise. The goal is to estimate θ_k for $k = 1, 2, \ldots, K$. By separating the real from the imaginary parts, (6) can be reformulated in the form of (4). Using an M-estimator will provide robustness against possible vertical outliers. In [38], the authors introduce a semiparametric approach, where they estimate the distribution of V_n to robustly estimate $\theta_k, k = 1, \ldots, K$. They show the advantage of the semiparametric estimator over the M-estimator. In the absence of knowledge of ω_k, $k = 1, \ldots, K$, one can use the robust estimation for frequency presented in the section "Robust Spectrum Estimation." Further references such as [40] studied the simultaneous estimation of the amplitudes and frequencies based on multiple sensors. The latter estimation falls onto robust nonlinear regression.

As soon as the data contains leverage points, M-estimators become nonrobust with a BP equal to zero. This becomes obvious when observing (5), since the predictors have unbounded influence and leverage points are given full weight. In the case of "good" leverage points, which follow the linear pattern of the majority of the data, this does not necessarily cause a bias in the estimate. "Bad" leverage points, on the other hand, can cause significant damage and will affect M-estimators as can be seen in Figure 3.

An intuitively appealing solution to this problem are generalized M (GM)-estimators, which do not only downweight observations with large residuals, but also the ones with a high leverage. GM-estimators have been used to solve various engineering problems, e.g., the defect detection in hardwood [34] and the linearized state estimation in power systems [33]. The limitation of the GM-estimator is that its BP cannot be 0.5, except for the single regression ($p = 1$). There exist several estimators which tradeoff BP and efficiency, including the S-estimator [42], the least trimmed squares (LTS) and the least median of squares (LMedS) estimators [22]. The LTS and the LMedS have been applied to solving engineering problems such as geolocation of UE for wireless networks [30], robust state estimation of electric power systems [33], and to estimate the optical flow in applications such as video coding or robot navigation [32]. The S-estimate is obtained by the parameter vector that minimizes a robust scale of the residuals. The idea of the LTS- and LMedS-estimators is to perform regression based on a subset of cases where the least-squares fit possesses the smallest sum of squared residuals.

The MM-estimator is becoming increasingly popular with the ever-increasing computational power. It has recently been applied, e.g., to static positioning in ultrasonic systems [31] and to the analysis of oddball event-related potentials in simultaneous electroencephalography (EEG) and functional MRI (fMRI) [35]. The MM-estimate is computed as follows:

Algorithm: MM-estimator

Step 1) Compute an initial consistent high BP estimate $\hat{\boldsymbol{\theta}}_o$ which does not have to have high efficiency (e.g., S- or LTS-estimate).

Step 2) Compute the high BP M-scale of the residuals of Step 1. (e.g., the bi-square M-estimate of scale).

Step 3) Compute an M-estimate of regression, using an iterative procedure starting at $\hat{\boldsymbol{\theta}}_o$.

When the estimators are chosen correctly, see, e.g., [19] for details, the MM-estimate inherits the high BP and other favorable properties, such as regression, scale, and affine equivariance of the initial $\hat{\boldsymbol{\theta}}_o$ and simultaneously high efficiency of the M-estimator in Step 3.

A second approach that also achieves high efficiency and a high BP simultaneously is τ-estimation. The τ-estimate of regression is the parameter vector $\boldsymbol{\theta}$ that minimizes a robust and efficient estimate of the residual scale called the τ-scale [43]. For a comparison of the above-described estimators in a practical setting, see "Practical Evaluation: Localization of a Mobile User Equipment."

ROBUST ESTIMATORS FOR MULTICHANNEL DATA

ROBUST METHODS FOR MULTICHANNEL DATA

Many key signal processing applications use multiple sensors to acquire multichannel data from measurement systems such as

PRACTICAL EVALUATION: LOCALIZATION OF A MOBILE USER EQUIPMENT

We give a practical example for which robust regression techniques are powerful [7], [30]. The localization of a mobile UE, i.e., an object or human being associated with a wireless transmitter device, using different base stations. This is an important task in many civilian and military applications, such as emergency services, yellow page services, and intelligent transport systems, among others [7]. There are different parameters, which can be used for localization; see, e.g., [44] for a comparison.

We give an example of localization based on time of arrival (TOA) measurements, where the nonlinear measurement equation at each base station is

$$y_n = h(\boldsymbol{\theta}) + v_n, \quad n = 1, \ldots, N.$$

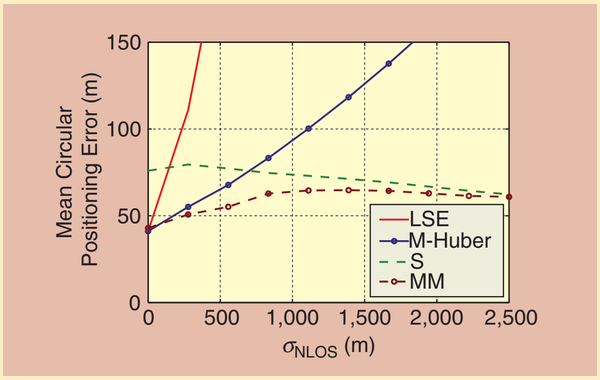
Here, $h = \sqrt{(x - x_{BS,m})^2 + (y - y_{BS,m})^2}$, denotes the distance from the UE to the mth base station, $\boldsymbol{\theta} = (x, y)^\top$ is the position of the UE, and $m = 1, \ldots, M$.

Figure S1 gives the results of a simulation example, where the measurement equation is linearized and the NLOS effects are modeled as i.i.d. random variables with probability density $f_V(v) = (1-\varepsilon) f_G(v; 0, \sigma^2_{\mathrm{LOS}}) + \varepsilon h_{\mathrm{NLOS}}(v)$. The mean circular positioning error (MCPE) using 10,000 Monte Carlo runs is given by

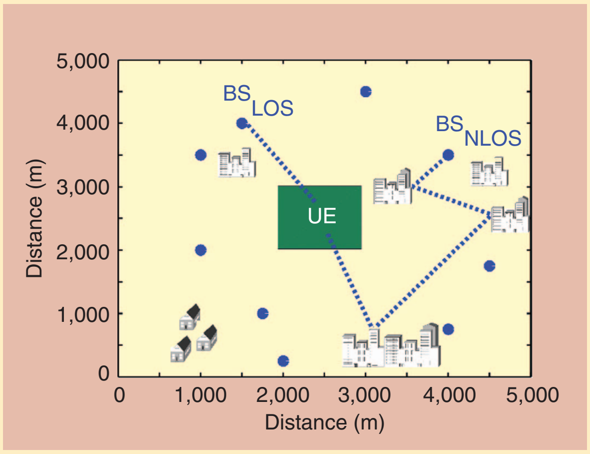
$$\mathrm{MCPE} = \frac{1}{10^4} \sum_{i=1}^{10^4} \sqrt{(\hat{x}^{(i)} - x^{(i)})^2 + (\hat{y}^{(i)} - y^{(i)})^2}.$$

and is displayed with respect to (w.r.t.) the scale of $h_{\mathrm{NLOS}}(v)$. In this example, $M = 10$ base stations are distributed as displayed in Figure S2 as blue circles around the UE, which is positioned with equal probability the green area. Here, $\sigma_{\mathrm{LOS}} = 150$ m, $\varepsilon = 0.4$, $h_{\mathrm{NLOS}}(v)$ is the exponential density and $N = 5$ measurements are assumed to be available at each base station. Four estimators with computational complexity increasing with descending order are displayed. The LSE, an M-estimator with Huber's score function tuned for 95% efficiency w.r.t. the Gaussian model, and the S- and MM-estimators, with parameters as recommended in [45].

For the Gaussian case, which corresponds to $\sigma_{\mathrm{NLOS}} = 0$ m, the LSE and the efficient robust estimators show best performance, while the S-estimator is outperformed. For $\sigma_{\mathrm{NLOS}} \neq 0$, the LSE quickly loses its optimality and is increasingly outdistanced. The M-estimator with Huber's (monotone) score function, while outperforming the LSE for all NLOS-cases, becomes inferior to the S-estimator for $\sigma_{\mathrm{NLOS}} > 750$ m because it only limits the effect of large outliers, instead of giving them zero-weight, as would be the case for a redescending ψ-function. The MM-estimator outperforms its competitors and renders stable performance over all σ_{NLOS}. This is because it is simultaneously efficient and highly robust as discussed previously.



[FIGS1] The localization of a mobile UE in a simulation for which 40% of the TOA measurements are biased due to NLOS propagation. NLOS effects are modeled by an exponential distribution and the mean positioning error of regression estimators is displayed w.r.t. the scale of the exponential distribution.



[FIGS2] In an urban scenario, localization of a mobile UE has to deal with NLOS propagation, which causes severe degradation of nonrobust position estimates.

multiple-input, multiple-output (MIMO) communication systems as well as phased array and MIMO radar systems. Also, many biomedical measurement systems, such as in magnetoencephalography (MEG) and EEG, surveillance and environmental monitoring produce multichannel data. The measurements are vector valued and often realizations of random signals and noise are described by a multivariate probability distribution function. The most commonly used multivariate statistical model is the multivariate Gaussian distribution, which is fully described by first- and second-order statistics, i.e., the mean vector and the covariance matrix. The Gaussian assumption is often justified by the central limit theorem and also convenient due to mathematical tractability [1].

THE MM-ESTIMATOR IS BECOMING INCREASINGLY POPULAR WITH THE EVER-INCREASING COMPUTATIONAL POWER.

One may attempt to derive robust signal processing techniques for multichannel data by simply applying robust methods developed for scalar signals to each signal component independently. This approach is unsuitable as it may lead to unexpected results. For example, in the case of location estimation, the resulting estimate is not necessarily within the convex hull of the data. Moreover, none of the data vector components may be outlying alone but the multivariate observations they form may be far away from the majority of the data. Such outliers would not be detected if robust signal processing is performed on individual single components, which also ignores the underlying correlation among the vector components. The vector components in multichannel data may be correlated and have different variances, which makes detecting outliers more difficult. In such cases estimators that are affine equivariant are highly desirable. By denoting an estimator by a functional $\mathbf{T}(\mathbf{z})$, affine equivariance means in multivariate location estimation that

$$\mathbf{T}(\boldsymbol{B}\mathbf{z}+\mathbf{b}) = \boldsymbol{B}\mathbf{T}(\mathbf{z})+\mathbf{b}$$

and in scatter (covariance) estimation that

$$\mathbf{T}(\boldsymbol{B}\mathbf{z}+\mathbf{b}) = \mathbf{B}\mathbf{T}(\mathbf{z})\boldsymbol{B}^{\mathrm{H}},$$

where $\boldsymbol{B}$ is a $K \times K$ full rank transformation matrix, $\mathbf{b}$ is a $K \times 1$ vector, and $^{\mathrm{H}}$ denotes Hermitian operation. Affine equivariance is desirable, since it means that the estimates are independent of the underlying measurement scale and coordinate system that may vary in practical applications, for example, in color images; see "Practical Evaluation: Impulsive Noise Attenuation in Color Images." Multichannel signal processing systems often require the estimation of location and scatter parameters, such as the mean and the covariance matrix, respectively. Consider the complex and vector valued random variate $\mathbf{Z}$ with mean $\boldsymbol{\mu} = \mathsf{E}[\mathbf{Z}]$ and scatter (covariance) matrix $\boldsymbol{\Sigma} = \mathsf{E}[(\mathbf{Z}-\boldsymbol{\mu})(\mathbf{Z}-\boldsymbol{\mu})^{\mathrm{H}}]$. We use the term *scatter matrix* for $\boldsymbol{\Sigma}$ to include cases where second order moments are not defined, such as for the Cauchy distribution. The scatter matrix is a scaled version of the covariance matrix, if it exists. Due to lack of space, we only treat the case where the complex-valued random vectors are circular, such that the complementary covariance matrix (also pseudo-covariance matrix or relation matrix) [46], $\mathsf{E}[(\mathbf{Z}-\boldsymbol{\mu})(\mathbf{Z}-\boldsymbol{\mu})^{\top}]$, is zero. Robustness in the case of noncircularity is addressed in [47].

Given observations $\mathbf{z}_n$, $n = 1, \ldots, N$, the sample mean and covariance matrix are obtained as

$$\hat{\boldsymbol{\mu}} = \frac{1}{N}\sum_{n=1}^{N}\mathbf{z}_n, \quad \text{and} \quad \hat{\boldsymbol{\Sigma}} = \frac{1}{N}\sum_{n=1}^{N}(\mathbf{z}_n - \hat{\boldsymbol{\mu}})(\mathbf{z}_n - \hat{\boldsymbol{\mu}})^{\mathrm{H}},$$

respectively. The sample mean and the sample covariance matrix perform optimally if the underlying signal is degraded by additive white Gaussian noise. However, their performance is poor if the noise is non-Gaussian. Thus, there is a need for robust techniques as defined in the "Introduction" section.

ROBUST MULTIVARIATE LOCATION AND SCATTER (COVARIANCE) MATRIX ESTIMATION

Robust estimation of the multivariate location parameter is needed in many multichannel filtering problems such as noise attenuation and outlier (impulsive noise) removal in color images, multichannel biomedical measurements, and multimodal imaging as well as sensor array signal processing. Robust estimation of scatter (covariance) matrices is a key to numerous signal processing tasks of optimal multichannel estimation and filtering, such as direction finding, interference cancellation, spatial multiplexing, and signal separation. We report on robust covariance matrix estimation [48] in view of solely the non-Gaussian, heavy-tailed nature of noise and omit robustness in situations of signal model inadequacy. We consider both qualitative and quantitative robustness using the concepts of the IF and the BP.

Multivariate location estimation has been thoroughly investigated in the statistics research community; see [22] and [49] and references therein. All of the estimators that are presented in the following possess the desirable affine equivariance property.

THE MINIMUM VOLUME ELLIPSOID ESTIMATOR AND THE MINIMUM COVARIANCE DETERMINANT ESTIMATOR

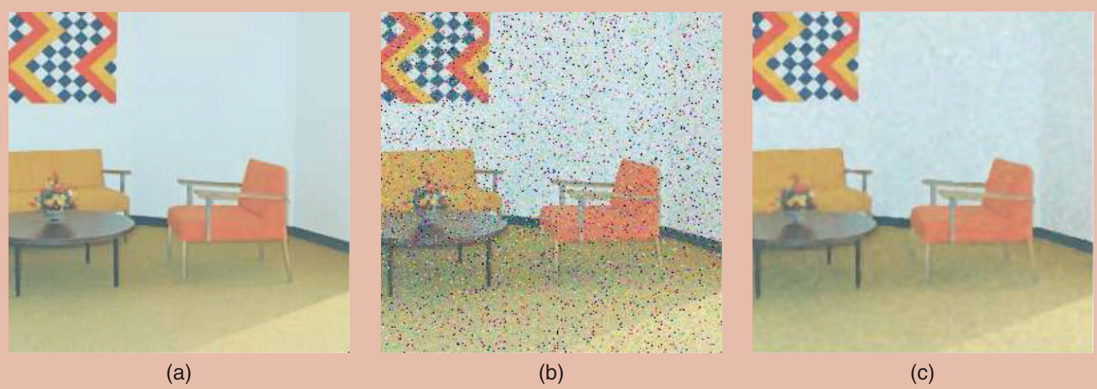
The minimum volume ellipsoid (MVE) estimate of location is defined as the center of the minimum volume ellipsoid covering at least h out of N data points. The minimum covariance determinant (MCD) estimate of location is found by computing the arithmetic mean for h out of N points for which the determinant of the sample covariance matrix is minimal. This means intuitively that the MCD is computed from the "closest" h samples. Given h samples that are not outlying, estimation is then based on the clean data points. The corresponding covariance matrix estimates are the sample covariance matrix of these points. Both the MVE and the MCD estimator require that the data does not lie in a lower than K-dimensional subspace, i.e., the data is in a general position. Both of these estimates employ a robust Mahalanobis distance, which is obtained by replacing the sample mean $\hat{\boldsymbol{\mu}}$ and

PRACTICAL EVALUATION: IMPULSIVE NOISE ATTENUATION IN COLOR IMAGES

Multivariate location estimation has been studied in image and biomedical signal processing; see [17] and [54] for reviews as well as detailed descriptions of a variety of techniques. For example, multivariate extensions of the median filter, such as the vector median [54], and order statistics filters, such as modified trimmed mean filters [55], have been developed.

Let us consider a simple color image processing problem where a color image is contaminated with impulsive noise that needs to be attenuated while preserving the edges in the image. It is well known that color image components are correlated and a robust image processing algorithm should exploit that correlation. The component variances may also be unequal and there are many different coordinate systems that are used for representing color images. It is therefore desirable that the algorithm works independent of these differences, i.e., possesses the equivariance property. The filtering is performed by computing a robust estimate of multivariate location in a small neighborhood (for example 5×5) of pixels and substituting the center pixel value with the estimate. Multivariate generalizations of the median filter, e.g., coordinate-wise or spatial (ℓ_1) median estimators and trimmed mean estimators have been used for this purpose. However, they do not posess the equivariance property.

In Figure S3, an MCD [49] estimate of location is computed using a fast iterative algorithm [56]. The output is obtained by finding h out of N observations that yield a smallest determinant for a standard covariance matrix and then computing the sample mean of those observations. The iterative algorithm finds the new subset of h samples by choosing the observations with the h smallest robust Mahalanobis distances from the current estimate. Their sample mean is the new location estimate and their sample covariance matrix is the scatter estimate. This is repeated until convergence takes place.



[FIGS3] A multichannel color image corrupted by impulsive noise. The noise is attenuated while preserving edges in the image by using a robust filter that computes the MCD location estimates iteratively in a small neighborhood of pixels. (a) Original image, (b) image contaminated by impulsive noise (outliers), and (c) output of the robust filter.

the sample covariance $\hat{\boldsymbol{\Sigma}}$ by their robust counterparts. MCD estimators are more efficient than MVE estimators, hence they are often preferred. A fast iterative algorithm for computing MCD estimates was proposed in [49].

THE S-, τ-AND MM-ESTIMATOR OF LOCATION AND SCATTER

The S-estimator of location and scatter is found by minimizing the determinant of an M-estimate of the covariance matrix. Analogously to the section "Linear Regression Models," where we aimed at minimizing a robust function of the residuals, the intuition behind the multivariate S-estimator is to minimize a robust measure of the distance, i.e., the determinant of an M-estimator of the covariance matrix. The BP of S-estimators is independent of the dimensionality of the data, unlike M-estimators for which the BP decreases. Other estimators, such as the MM- and τ-estimators, (see the section "Linear Regression Models"), have been derived for multivariate location and scatter estimation. A fast algorithm for the MM-estimator is given in [50]. In the sequel, two robust estimators of the scatter matrix are presented in more detail and evaluated using practical examples.

THE M-ESTIMATOR OF COVARIANCE

The M-estimator of covariance based on a sample $\mathbf{z}_1, \ldots, \mathbf{z}_N$ in $\mathbb{C}^K$ is the positive definite Hermitian (PDH) $K \times K$ matrix $\hat{\boldsymbol{\Sigma}}$, which solves

$$\hat{\boldsymbol{\Sigma}} = \frac{1}{N}\sum_{n=1}^{N} w(\mathbf{z}_n^{\mathrm{H}} \hat{\boldsymbol{\Sigma}}^{-1} \mathbf{z}_n)\, \mathbf{z}_n \mathbf{z}_n^{\mathrm{H}},$$

where $w : [0, \infty) \to \mathbb{R}$ is a weighting function. A robust weighting function $w(\cdot)$ is descending to zero, i.e., a highly deviating observation $\mathbf{z}_n$ with large $\|\hat{\boldsymbol{\Sigma}}^{-1/2} \mathbf{z}_n\|^2 = \mathbf{z}_n^{\mathrm{H}} \hat{\boldsymbol{\Sigma}}^{-1} \mathbf{z}_n$ is given less weight.

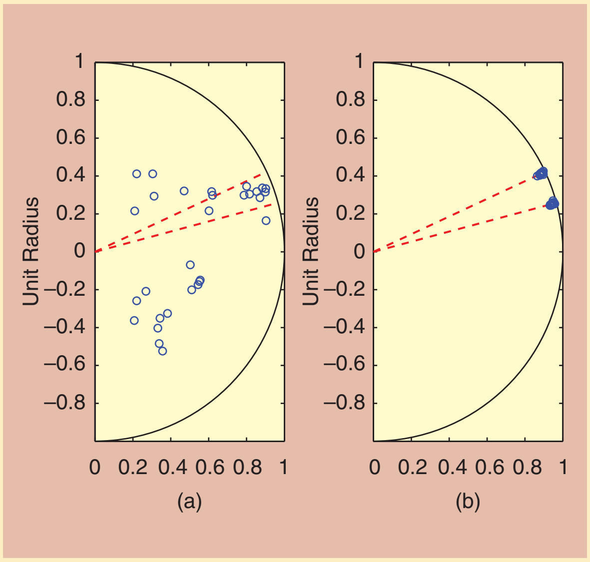
PRACTICAL EVALUATION: DIRECTION OF ARRIVAL ESTIMATION

Define the signal model for an array output of K sensors by

$$\mathbf{Z}(n) = \mathbf{AS}(n) + \mathbf{W}(n), \quad n = 1, \ldots, N, \tag{S1}$$

where for $n = 1, \ldots, N$, $\mathbf{Z}(n) = \mathbf{X}(n) + j\mathbf{Y}(n)$ is the complex K vector-valued observation model, $\mathbf{A}$ is the $K \times P$ system matrix, $\mathbf{S}(n)$ is the P vector-valued signal model and $\mathbf{W}(n)$ is K vector-valued white noise. Depending on the problem at hand, $\mathbf{A}$ may define the matrix of sensor array steering vectors, a MIMO channel in multiantenna communications, or the mixing system in signal separation. Commonly we assume that $P \le K$. Without loss of generality, we assume that $\mathrm{E}[\mathbf{Z}(n)] = 0$. The covariance matrix (if it exists) is then given by $\boldsymbol{\Sigma} = \mathrm{E}[\mathbf{Z}(n)\mathbf{Z}(n)^{\mathrm{H}}] = \mathbf{A}\boldsymbol{\Sigma}_s\mathbf{A}^{\mathrm{H}} + \sigma_w^2\mathbf{I}$, where $\boldsymbol{\Sigma}_s = \mathrm{E}[\mathbf{S}(n)\mathbf{S}(n)^{\mathrm{H}}]$ is a positive definite $P \times P$ signal covariance matrix. The eigenvalue decomposition of a PDH matrix $\boldsymbol{\Sigma}$ is given by $\boldsymbol{\Sigma} = \boldsymbol{\Gamma}\boldsymbol{\Lambda}\boldsymbol{\Gamma}^{\mathrm{H}}$.

An example of robust subspace estimation of direction of arrival in a heavy-tailed noise situation is provided in Figure S4. We used a four-element uniform linear array (ULA) with two Gaussian signals in Cauchy noise at 20 dB signal-to-noise ratio (SNR) impinging from 15° and 25° on the array. With $N = 15$ Root-MUSIC (where MUSIC stands for Multiple Signal Classification) was applied. This exemplifies that the robust method works even for small sample support, which occurs frequenty in practical situations.



[FIGS4] Direction of arrival estimation using Root-MUSIC for two sources impinging on a four-element uniform linear array. The subspaces are computed based on the sample (a) covariance matrix and the (b) M-estimator of the covariance matrix, which is the MLE for a t_1 distribution. The red dashed lines depict the true directions of arrival, while blue circles are the estimates.

Using the complex t-distribution as a heavy-tailed model, we obtain t_ν M-estimators as an example of M-estimators, with the weight function

$$w(s) = w_\nu(s) = \frac{2K + \nu}{\nu + 2s},$$

where $\nu > 0$. This M-estimator is also the MLE if the data is complex K vector-valued t_ν-distributed. The estimates may conveniently be computed with the following iterative formula:

$$\hat{\boldsymbol{\Sigma}}_{k+1} = \frac{1}{N}\sum_{n=1}^{N} w_\nu(\mathbf{z}_n^{\mathrm{H}}\,\hat{\boldsymbol{\Sigma}}_k^{-1}\mathbf{z}_n)\,\mathbf{z}_n\mathbf{z}_n^{\mathrm{H}},$$

which converges to the unique solution $\hat{\boldsymbol{\Sigma}}$ under mild regularity conditions on the data (see [51]). For an application of this estimator to direction of arrival estimation using Root-MUSIC; see "Practical Evaluation: Direction of Arrival Estimation." An interesting special case is obtained with $\nu = 0$ and a weight function $w(s) = K/s$, i.e.,

$$\hat{\boldsymbol{\Sigma}} = \frac{K}{N}\sum_{n=1}^{N}\frac{\mathbf{z}_n\mathbf{z}_n^{\mathrm{H}}}{\mathbf{z}_n^{\mathrm{H}}\,\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{z}_n},$$

excluding all $\mathbf{z}_n = \mathbf{0}$.

Multivariate generalizations of sign and ranks from nonparametric statistics can also be used to derive robust estimators for the covariance matrix [52]. We define the spatial sign function $\mathcal{S}$ for a complex K vector-valued $\mathbf{z}$ as follows:

$$\mathcal{S}(\mathbf{z}) = \begin{cases} \dfrac{\mathbf{z}}{\|\mathbf{z}\|}, & \mathbf{z} \ne \mathbf{0} \\ \mathbf{0}, & \mathbf{z} = \mathbf{0}, \end{cases}$$

where $\|\mathbf{z}\| = (\mathbf{z}^{\mathrm{H}}\mathbf{z})^{1/2}$. The spatial sign of $\mathbf{z}$ is a unit length vector to the direction of $\mathbf{z}$ and hence a natural generalization of the univariate sign function. For a complex K vector-valued set of data $\mathbf{z}_1, \ldots, \mathbf{z}_N$, the sample spatial sign covariance matrix (SCM) is

$$\hat{\boldsymbol{\Sigma}}_1 = \frac{1}{N}\sum_{n=1}^{N}\mathcal{S}(\mathbf{z}_n)\mathcal{S}(\mathbf{z}_n)^{\mathrm{H}}.$$

This method is particularly useful in subspace estimation since it estimates the eigenvectors in a convergent manner; see "Practical Evaluation: Robust High-Resolution Frequency Estimation" for an application of the SCM to the estimation of line spectra.

THEORETICAL EVALUATION

We are interested in studying the qualitative and quantitative robustness of scatter matrix estimators. The concept of the BP is typically used to establish quantitative robustness; see the section "Basic Concepts of Robustness." In the context of covariance estimation, one definition could be the smallest number k out of N observations that could make either the largest eigenvalue over all bounds or the smallest eigenvalue arbitrarily close to zero, and consequently the matrix ill conditioned.

PRACTICAL EVALUATION: ROBUST HIGH-RESOLUTION FREQUENCY ESTIMATION

Parametric methods such as MUSIC [78] or ESPRIT for the estimation of line spectra can be robustified by employing robust eigendecomposition of the covariance matrix. Robust and consistent estimates of the covariance matrix and associated subspaces may be obtained as described in the section "Robust Estimators for Multichannel Data."

In the following, we give an example where $x_0, \ldots, x_{N-1}$ are observations from a random signal, X_n, $n \in \mathbb{Z}$, modeled by p complex-valued sinusoids buried in complex circular white noise,

$$X_n = \sum_{i=1}^{p} A_i e^{j(\omega_i n + \Phi_i)} + W_n, \quad n = 0, \ldots, N-1,$$

where A_i and ω_i, $i = 1, \ldots, p$, are real-valued parameters to be estimated and Φ_i are independent random variates for $i = 1, \ldots, p$, uniformly distributed on $[\pi, \pi)$. The above model can be written in matrix form as follows:

$$\mathbf{X} = \mathbf{AS} + \mathbf{W},$$

where $\mathbf{X} = (X_0, \ldots, X_{N-1})^\top$, $\mathbf{A}$ is an $N \times p$ complex-valued matrix, with elements $(\mathbf{A})_{n,i} = e^{j\omega_i n}$ for $n = 0, \ldots, N-1$ and $i = 1, \ldots, p$. $\mathbf{S} = (S_1, \ldots, S_p)^\top$ is a complex-valued $p \times 1$ vector with elements $S_i = A_i e^{j\Phi_i}$ and $\mathbf{W} = (W_0, \ldots, W_{N-1})^\top$. Assuming that X_n and W_n are uncorrelated for all $n \in \mathbb{Z}$, the signal model above leads to the covariance model

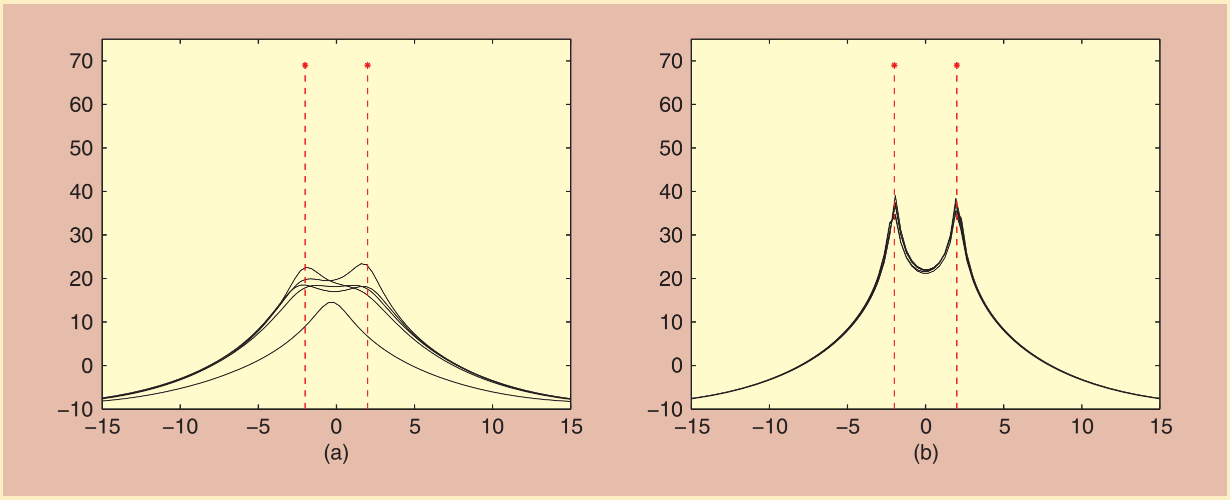
$$\boldsymbol{\Sigma}_x = \mathbf{A}\boldsymbol{\Sigma}_s\mathbf{A}^{\mathrm{H}} + \sigma_w^2\mathbf{I} = \boldsymbol{\Gamma}\boldsymbol{\Lambda}\boldsymbol{\Gamma}^{\mathrm{H}} = \boldsymbol{\Gamma}_s\boldsymbol{\Lambda}_s\boldsymbol{\Gamma}_s^{\mathrm{H}} + \boldsymbol{\Gamma}_w\boldsymbol{\Lambda}_w\boldsymbol{\Gamma}_w^{\mathrm{H}},$$

where $\boldsymbol{\Sigma}_x$ is $N \times N$ matrix-valued, $\boldsymbol{\Sigma}_s = E[\mathbf{SS}^{\mathrm{H}}]$ is $p \times p$ matrix-valued and the r.h.s. of the second equality stems from eigendecomposition of $\boldsymbol{\Sigma}_x$. Evaluation of the so-called MUSIC pseudo-spectrum

$$\hat{P}_{\mathrm{MUSIC}} = \frac{1}{\mathbf{q}^{\mathrm{H}}\mathbf{V}_w\mathbf{V}_w^{\mathrm{H}}\mathbf{q}}$$

leads to sharp peaks at $\mathbf{q} = \mathbf{a}_i$, where $\mathbf{a}_i$ is the ith column of $\mathbf{A}$ for $i = 1, \ldots, p$. A standard estimator for the covariance matrix is the sample covariance matrix. This estimator is very sensitive even to small departures from the Gaussian noise model.

Robust estimators of the covariance matrix may be formed to obtain a reliable performance in the non-Gaussian noise case. Such estimators have been discussed in the section "Robust Estimators for Multichannel Data." One may use the complex-valued spatial sign function $\mathcal{S}(\mathbf{x})$ to form a robust estimate of the covariance matrix. It was proven that the spatial SCM gives convergent estimates of noise and signal subspace basis vectors [79] if the noise is Gaussian. An example of frequency estimation using the spatial SCM and the MUSIC estimator in the case where the noise is Cauchy distributed is shown in Figure S5.



[FIGS5] MUSIC frequency estimates of two narrowband signals buried in Cauchy noise. The MUSIC pseudospectrum based on (a) the sample covariance matrix and (b) the spatial SCM is shown. The estimator based on the sample covariance matrix gives clearly biased estimates and fails to resolve the two sources in non-Gaussian noise.

Loosely speaking, qualitative robustness means that small departures from the underlying assumptions should cause only a small change in the performance of the estimator. Qualitative robustness is typically studied using the IF; see the section "Basic Concepts of Robustness." A particularly interesting form of the IF is given in terms of eigenvalue or eigenvector functionals of a scatter matrix. To present these functionals, we first introduce some additional distribution models and notation. We start by defining complex multivariate distributions where contours of equal probability are ellipses. A random vector $\mathbf{Z} = \mathbf{X} + j\mathbf{Y} \in \mathbb{C}^K$ is said to have a complex elliptically symmetric (CES) distribution with parameters

$$\boldsymbol{\mu} = \boldsymbol{\mu}_1 + j\boldsymbol{\mu}_2 \in \mathbb{C}^K \quad \text{and} \quad \boldsymbol{\Sigma} = \boldsymbol{\Sigma}_1 + j\boldsymbol{\Sigma}_2 \in \mathrm{PDH}(K)$$

if $\mathbf{Z}_R = (\mathbf{X}^\top, \mathbf{Y}^\top)^\top$ has a real elliptically symmetric distribution with parameters $\boldsymbol{\mu}_R = (\boldsymbol{\mu}_1^\top, \boldsymbol{\mu}_2^\top)^\top \in \mathbb{R}^{2K}$ and

$$\boldsymbol{\Sigma}_R = \frac{1}{2}\begin{pmatrix} \boldsymbol{\Sigma}_1 & -\boldsymbol{\Sigma}_2 \\ \boldsymbol{\Sigma}_2 & \boldsymbol{\Sigma}_1 \end{pmatrix} \in \mathrm{PDS}(2K),$$

where PDS stands for positive definite symmetric. If the probability density function of $\mathbf{Z}_R$ is given by

$$f_Z^R(\mathbf{z}_R;\, \boldsymbol{\mu}_R, \boldsymbol{\Sigma}_R) = |\boldsymbol{\Sigma}_R|^{-1/2} g\big((\mathbf{z}_R - \boldsymbol{\mu}_R)^\top \boldsymbol{\Sigma}_R^{-1}(\mathbf{z}_R - \boldsymbol{\mu}_R)\big),$$

for some nonnegative function $g(\cdot)$, then the pdf of $\mathbf{Z}$ is

$$f_Z^C(\mathbf{z};\, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = 2^K |\boldsymbol{\Sigma}|^{-1} g\big(2(\mathbf{z}-\boldsymbol{\mu})^{\mathrm{H}} \boldsymbol{\Sigma}^{-1}(\mathbf{z}-\boldsymbol{\mu})\big).$$

Sometimes $g(\cdot)$ is called the density generator function since it uniquely distinguishes a distribution from another [53]. Some well-known distributions can be presented using this model, e.g., the complex K-variate normal, with

$$g(t) = (2\pi)^{-K} \exp(-t/2)$$

and the complex K-variate t-distribution with ν degrees of freedom, with

$$g(t) \propto c(1 + t/\nu)^{-(2K+\nu)/2},$$

which is the complex multivariate Cauchy distribution for $\nu = 1$.

For any affine equivariant scatter matrix functional $\mathbf{C}(F) \in \mathrm{PDH}(K)$, there exist functions $\alpha, \beta: \mathbb{R}^+ \to \mathbb{R}$ such that the IF of $\mathbf{C}(F)$ at a CES distribution F is

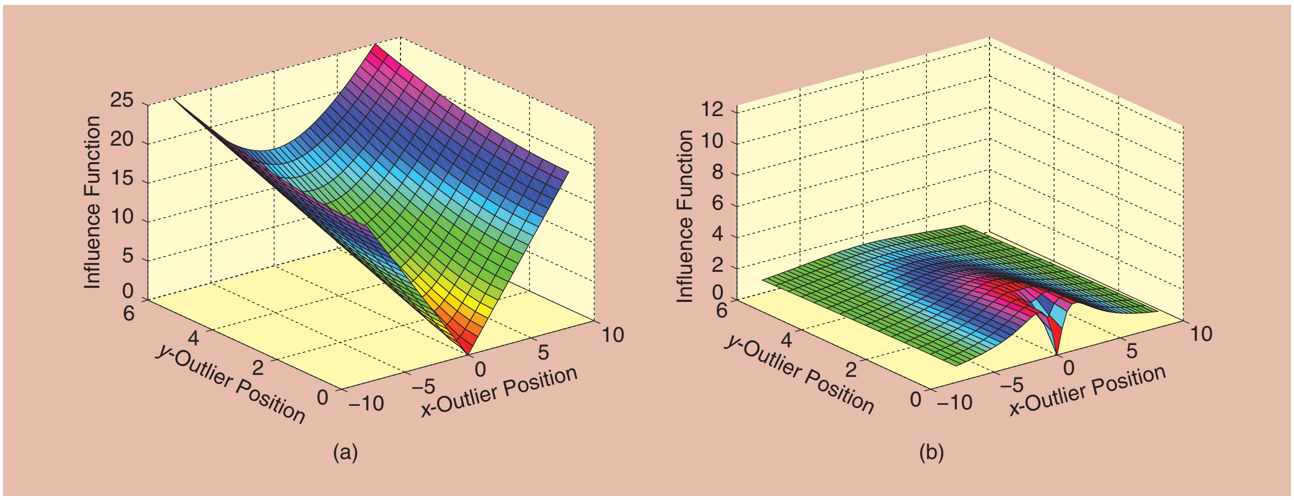
$$\mathrm{IF}(\mathbf{z}; \mathbf{C}, F) = \alpha(r)\boldsymbol{\Sigma}^{1/2}(\mathbf{w}\mathbf{w}^{\mathrm{H}} - (1/K)\mathbf{I})\boldsymbol{\Sigma}^{1/2} + \beta(r)\boldsymbol{\Sigma},$$

where $r^2 = \mathbf{z}^{\mathrm{H}}\boldsymbol{\Sigma}^{-1}\mathbf{z}$ and $\mathbf{w} = \boldsymbol{\Sigma}^{-1/2}\mathbf{z}/r$. The IF of the scatter functional is bounded iff the corresponding "weight functions" α and β are bounded. For the sample covariance matrix, $\alpha(r)$ and $\beta(r)$ are quadratic in r and consequently the IF is unbounded and the estimator is not robust.

The eigenvalue decomposition of the covariance or scatter matrix is extensively employed in sensor array signal processing applications. It allows for identifying signal and noise subspaces, high-resolution parameter estimation, source signal enumeration, and decorrelating signals, for example. Eigenvalues and eigenvectors may also be used in studying qualitative robustness. By expressing IFs in terms of eigenvalue or eigenvector functionals, one sees how sensitive the basis vectors of the subspaces are to outliers, and how the eigenvalue spectrum behaves when the underlying assumption on the noise model does not hold. The functional representations of the sample eigenvalues $\hat{\lambda}_i$ and sample eigenvectors $\hat{\boldsymbol{\gamma}}_i$, $i = 1, \ldots, K$ of $\hat{\boldsymbol{\Sigma}}$ are obtained from the eigenvalue decomposition of the scatter functional $\mathbf{C}(F) = \mathbf{G}(F)\mathbf{L}(F)\,\mathbf{G}(F)^{\mathrm{H}}$, where $\mathbf{G}(F) = (\mathbf{g}_i(F) \cdots \mathbf{g}_K(F))$ and $\mathbf{L}(F) = \mathrm{diag}\{l_1(F), \ldots, l_K(F)\}$, with $\mathbf{g}_i(F) \in \mathbb{C}^K$ and $l_i(F) \in \mathbb{R}$ being the eigenvector and eigenvalue functionals of $\mathbf{C}(F)$, respectively. Since, for a constant c, $\mathbf{C}(F) = c\boldsymbol{\Sigma}$, it immediately follows that $\mathbf{g}_i(F) = \boldsymbol{\gamma}_i$ and $l_i(F) = c\lambda_i$, for $i = 1, \ldots, K$. As an example, consider how sensitive eigenvalues of scatter matrix estimates are to outliers. Let the IF for the eigenvector functional $\mathbf{g}_i(F)$ of $\mathbf{C}(F)$ corresponding to a simple eigenvalue λ_i be given by

$$\mathrm{IF}(\mathbf{z}; \mathbf{g}_i, F) = \frac{\alpha(r)}{c} \sum_{j=1, j\neq i}^{K} \frac{\sqrt{\lambda_j \lambda_i}}{\lambda_i - \lambda_j} \boldsymbol{w}_j \boldsymbol{w}_i^* \boldsymbol{\gamma}_j,$$

where $\boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_K$ denote the eigenvectors of the scatter matrix $\boldsymbol{\Sigma}$ and w^* is the complex conjugate of w. Figure 4 depicts $\|\mathrm{IF}(\mathbf{z}; \mathbf{g}_1, F)\|$ for the sample covariance estimator [Figure 4(a)] and the t_1 M-estimator [Figure 4(b)] at the bivariate complex normal distribution ($\lambda_1 = 1$, $\lambda_2 = 0.6$). It can be seen that the IF is smooth and bounded for the t_1 M-estimator, hence the method is qualitatively robust. The IF for the sample covariance matrix is not bounded (quadratic in the extent of outlyingness) which means that it is very sensitive to even



[FIG4] The IFs for the largest eigenvalue of two covariance estimators at the bivariate complex normal distribution. The IF of the (a) sample covariance estimator increases quadratically, while that of the (b) t_1 M-estimator is smooth and bounded, which shows its qualitative robustness.

small fraction of outliers. Similarly, the IF can be given in terms of eigenvectors; see [51].

ROBUST ESTIMATORS FOR DEPENDENT DATA

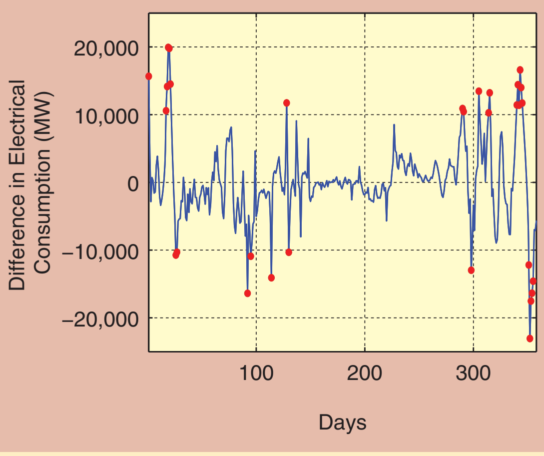
There is a large body of literature on robust estimation methods for engineering practice under the assumption of i.i.d. observations. However, methods for dependent data are limited. Although the first contributions were made in the mid-1970s and 1980s [13], [57] there has not been much progress on methods for dependent data for a long time. This is mainly due to the fact that existing robust estimators and concepts for i.i.d. data are not easily extendable to dependent data, and thus new robust approaches are sought for. Recently, research increased significantly in this area and some novel estimators have been proposed. In this section, we describe the most interesting methods and concepts that are useful to engineering problems. To illustrate the applicability of these methods, an example of electricity consumption forecasting is treated [8], [58]; see "Practical Evaluation: Electricity Consumption Forecasting." Forecasting the electric load is an important task carried out by electric companies that produce electricity or manage and operate electrical power transmission systems. Daily short-term load forecasts allow these companies to ensure the balance between supply and demand, which at the

PRACTICAL EVALUATION: ELECTRICITY CONSUMPTION FORECASTING

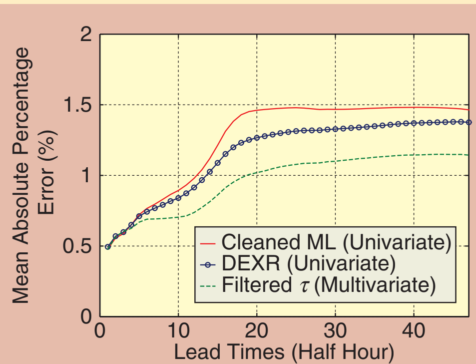
In the following, we treat a practical problem of electricity consumption forecasting. Figure S6 shows the one-week difference in consumption in France at 7:00 a.m. in 2005.

To forecast the electricity consumption, the load time series is first corrected from the weather influence by using a regression model where the explanatory variables are the temperature and the nebulosity. The nebulosity is a measure of the cloud cover in real time. In this step, a robust regression estimation should be considered. The obtained residual series encompasses some seasonalities (daily, weekly, yearly) and is generally modeled by a seasonal AR integrated moving average (SARIMA) model. Double differencing leads to a stationary process by removing the daily and weekly seasonalities. The presence of outliers, which are marked in red, is clear from Figure S6. The goal is to improve the short-term load forecasting quality in the presence of outliers due to special days, such as nonworking days and holidays, soccer's World Cup, and storms by choosing the best possible robust estimator. In [8], different robust estimators were compared and their practical effectiveness for short-term load forecasting was studied. For example, an interesting criterion for comparing the quality of different estimation methods is the error between the observation at day n and its forecast from day $n-1$. Let us now briefly illustrate some of the practical results obtained in [8]. The robust mean squared error or the mean squared error evaluated over normal days of the 200 out-of-sample, one-day-ahead forecast error for different robust methods for the French electrical consumption at 8:00 a.m. is computed for the RME, the filtered τ-estimator (F-τ), the GM-estimator, and the cleaned ML; the obtained results are 0.696, 0.708, 1.109, and 1.632 $(\times 10^6\{\mathrm{MW}^2\})$, respectively. It can be seen that some of the sophisticated robust estimators discussed above outperform the cleaned ML, which is used by the majority of electricity companies. Both the RME and the filtered-τ clearly behave better than the GM estimator.

The evaluation of the forecasting quality is generally assessed by plotting a robust version of the mean absolute percentage error (MAPE) [8] with respect to the forecasting horizon. Figure S7 [8] illustrates the robust MAPE of the univariate cleaned maximum likelihood, a univariate robust version of the double exponential smoothing (DEXR) and the multivariate filtered τ-estimator. More details about this application can be found in [8].



[FIGS6] An illustration of the occurrence of outliers (marked in red) in the one-week difference in electrical consumption in France at 7:00 a.m. in 2005.



[FIGS7] A plot of the MAPE versus the 48 half hours of a day (00:00, . . . , 23:30) for France (200 days post-sample period). The univariate cleaned maximum likelihood is the method used by the majority of electricity companies. The univariate robust DEXR gives some improvement but is outperformed by the multivariate filtered τ-estimator, which exploits the correlation structure of the multichannel data.

same time minimizes the costs and maximizes the security and reliability of an electrical power transmission system [59]. Before we describe this problem and propose solutions, we first illustrate the main concepts and methods for robust estimation in the dependent data case.

OUTLIERS IN CORRELATED SIGNALS

Different structural types of outliers for correlated signals have been defined in the literature [60]. They are: 1) additive outliers, which are external errors that affect only the current observation; 2) innovation outliers, which are internal errors in the noise sequence (Z_n) that drives, for example, the filter of an autoregressive moving average (ARMA) model; and 3) replacement outliers, which are outliers drawn from a second replacement process which can be independent from the nominal process. The additive or replacement outliers are more frequent in practice and much more difficult to deal with in the estimation than innovative outliers; see the section "Robust Estimation for ARMA Models" and Figure 5. In terms of their appearance time configuration, outliers can also be classified into 1) isolated independent outliers, which appear separately, scattered and isolated and 2) patchy outliers, which appear grouped in chains and constitute several outlying blocks. Some other types of outlier models exist in the literature, such as level shift, variance change, and seasonal outliers [61], to mention a few.

One of the most popular approaches the signal processing practitioner uses to model dependent data is the ARMA model [62]. This is not much different in robust statistics.

ROBUST ESTIMATION FOR ARMA MODELS

An ARMA(p, q) model is defined as

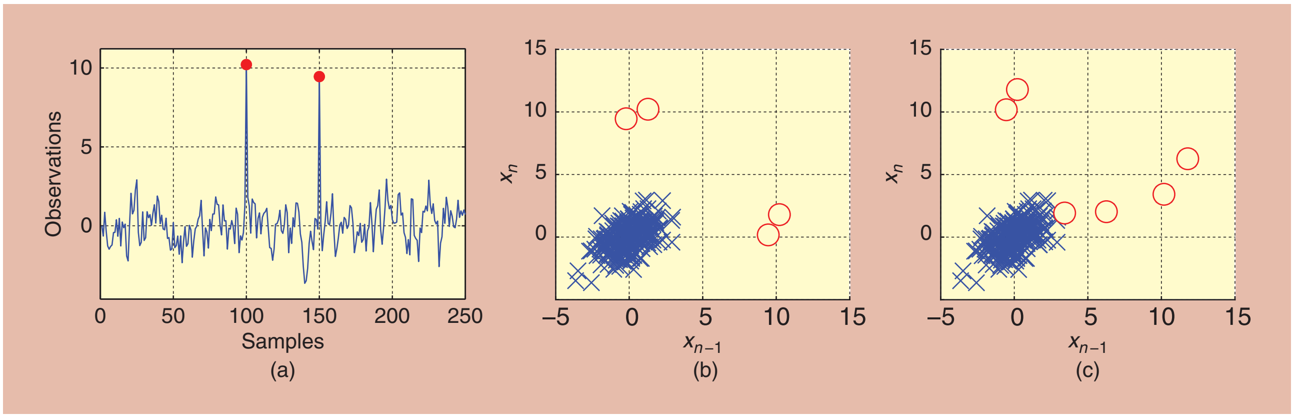
$$X_n + \sum_{k=1}^{p} a_k X_{n-k} = Z_n + \sum_{k=1}^{q} b_k Z_{n-k}, \quad n \in \mathbb{Z}, \tag{7}$$

where Z_n is a sequence of zero-mean i.i.d. random variables with finite variance σ_Z^2 (i.e., white noise). An autoregressive (AR) model AR(p) obeys (7) if $q = 1$ and $b_1 = 0$. A convenient representation of (7) is $A(z)X_n = B(z)Z_n$. Here $A(z) = 1 + a_1 z^{-1} + \cdots + a_p z^{-p}$, $B(z) = 1 + b_1 z^{-1} + \cdots + b_q z^{-q}$ where z^{-1} is the lag operator defined by $z^{-l}X_n = X_{n-l}, l \in \mathbb{Z}$. The polynomial $A(z)$ has no zeros z_k with $|z_k| > 1$, which ensures stationarity of the process X_n. Invertibility of the system is obtained when the same condition is satisfied for $B(z)$. In the following, we focus on the estimation in the presence of outliers, which consists in estimating the parameter vector $\mathbf{c} = (a_1, \ldots, a_p, b_1, \ldots, b_q)^\top$ and σ_Z based on the contaminated observations $y_1, \ldots, y_N$.

First, let us illustrate the presence of additive and innovation outliers in a simple case of an AR(1) model. This will give some insight into the impact of both types of outliers on the classical estimation. Figure 5(a) depicts an example for an AR(1) process and its two-dimensional regression representation in the case of additive [Figure 5(b)] and innovation outliers [Figure 5(c)]. We remark that two additive outliers in the AR(1) process generate four outliers in the regression representation which are marked as red circles in the plot. Two of them are y-axis outliers and two bad leverage points. Leverage points, as explained in the section "Linear Regression Models," are highly influential observations on the classical estimation and hence may cause a large bias to the estimator $\hat{a}$ which is the tangent of the regression line. On the other hand, two innovative outliers in the AR(1) process generate two y-axis outliers but several good leverage points in the regression representation. Since the good leverage points obey the model, they will generally compensate the bad effect of the y-axis outliers and even further improve classical estimation. In the sequel, we use the term "outlier" for both additive or replacement outliers. Let us now revisit some of the interesting robust estimators for ARMA models.

THE CLEANED MLE

This is the most popular method among practicing engineers. The method uses maximum likelihood estimation after 3-σ rejection. The 3-σ rejection is a simple rule of thumb that rejects observations beyond three times the known or estimated standard deviation. Its justification is explained by the fact that the fraction of observations beyond this limit is very small for a Gaussian distribution. Indeed, we have



[FIG5] (a) AR(1) process with two additive outliers marked red. (b) Representation of an AR(1) with two additive outliers as a regression. (c) Representation of an AR(1) with two innovation outliers as a regression. Leverage points in the regression that are generated by the outliers are marked as red circles.

$P(|X_n - \mu| > 3\sigma) = 0.003$ when $X_n \sim \mathcal{N}(\mu, \sigma^2)$. Robust estimates of the mean μ and the standard deviation σ such as the sample median and the normalized median absolute deviation (3) should be used. Using classical estimators will lead to what is known as the masking effect, which means that some outliers will be masked or will not be detected because of other outliers that inflate the estimate of σ and cause a bias to $\hat{\mu}$. After rejecting outliers, an MLE that handles missing observations is used [63]. Several engineering practitioners believe that repeating the 3-σ with classical estimators several times can improve detection of outliers. However, this is not true; in fact, this approach can even reject good observations in the subsequent steps. Furthermore, even the 3-σ with robust estimators does not provide the best performance since it neglects the time correlation and hence lacks robustness.

[ONE OF THE MOST POPULAR APPROACHES THE SIGNAL PROCESSING PRACTITIONER USES TO MODEL DEPENDENT DATA IS THE ARMA MODEL.]

THE FILTERED GM AND THE FILTERED M-ESTIMATOR

The GM-estimator was introduced to robustify the M-estimator against leverage points. It is popular in power systems for state estimation [33]. However, its major inconvenience is the lack of robustness when the order of the autoregression is high. In fact, for an AR(p), its BP decreases with an increasing p [19]. Furthermore, it is not robust for ARMA models. When combined with the robust filter cleaner, its performance improves, but it is still not satisfactory. The filtered M-estimator is a special case of the filtered GM-estimator when the weighting function to downweight leverage points is the identity weight, which means that they suffer from the same drawbacks.

THE RESIDUAL AUTOCOVARIANCE AND THE TRUNCATED RA-ESTIMATOR

Both the residual autocovariance (RA)- and truncated-RA estimators were introduced in [64]. The idea behind the RA is to robustify M-estimators using robust estimation of the autocovariance of the residuals. The RA is not robust for ARMA models since one outlier has an impact on a very large number of residuals. To solve this problem, Bustos and Yohai [64] proposed to use the truncated residuals in the estimation procedure, to obtain the so-called truncated RA estimator. The truncated RA-estimator is robust but not efficient for a Gaussian ARMA process.

THE FILTERED S-ESTIMATOR

As in the case of LSE, which can be seen as the estimate minimizing the standard deviation of the residuals, the intuitive idea of the filtered S-estimator consists in minimizing a robust scale of the robust prediction residuals. The filtered S-estimator is robust for ARMA models. However, it is not efficient when highly robust; on the other hand, it lacks robustness when highly efficient.

THE FILTERED τ-ESTIMATOR

The filtered τ-estimator proposed by Yohai and Zamar [43] is obtained by minimizing the scale τ-estimate of the robust prediction residuals. It ensures a high BP of 0.5 and a controllable normal efficiency (0.95) [19]. The filtered τ-estimator has a good performance, but it suffers from a rather difficult implementation.

THE RATIO OF MEDIANS ESTIMATOR AND THE MEDIAN OF RATIOS ESTIMATOR

The ratio of medians estimator (RME) and median of ratios estimator (MRE) methods proposed in [8] and [65] use robust autocorrelation estimates based on sample medians coupled with a robust filter cleaner, which rejects outlying observations. This improves the efficiency and prevents outlier propagation. To estimate a contaminated Gaussian stationary ARMA(p, q) , we can use the following algorithm:

Algorithm: RME and MRE

Step 1) Fit a high order AR(p^*) using the RME or MRE, where the order $p^* > p$ and is obtained by a robust order selection criterion. The autocorrelations are linked to the ratio of medians of the processes $Y_n Y_{n-k}$, Y_n^2 by a relation based on the Bessel-K function [8]. On the other hand, the MRE uses the median of Y_n / Y_{n-k} to estimate the correlation.

Step 2) Discard the outliers by filtering the signal using a robust filter cleaner [74] with the estimated parameters of the high order AR(p^*) and apply a classical ML-based estimation method of ARMA models that handles missing data [63].

These methods offer good performance in practice and are easy to implement. However, their BP is limited to 0.25.

THE FILTERED HELLINGER-BASED ESTIMATOR

The filtered Hellinger-based estimator [8] is computed in two steps such as in the RME and MRE procedures. The estimation of the high-order AR in Step 1 above is carried out by a Levinson-Durbin algorithm in which the partial autocorrelation is estimated via minimization of a robust Hellinger scale [8]. The estimator offers a good performance, but it is complicated to implement and requires significant computing power.

THE BOUNDED INNOVATION PROPAGATION-ARMA ESTIMATOR

The bounded innovation propagation-ARMA (BIP-ARMA) estimator was recently proposed in [66] and uses the BIP-ARMA model

$$Y_n = Z_n + \sum_{i=1}^{\infty} \xi_i \sigma \eta\left(\frac{Z_{n-i}}{\sigma}\right).$$

Herein, η is an even and bounded function, ξ_i's are the coefficients of the polynomial $A^{-1}(z^{-1})B(z^{-1})$, and σ is the residual scale. In [66], it is shown that this model is equivalent to using a robust filter cleaner. This means that the effect of an additive outlier is limited to the region where it occurs. Combining this model with an M- or MM-estimator gives a robust estimator for the parameters of an ARMA(p, q) model, which ensures a high BP (0.5) and a controllable normal efficiency under Gaussianity (0.95) [66]. The estimator has the advantage of being consistent and having a tractable asymptotic expression and confidence intervals, which is not the case with the filter cleaner.

THEORETICAL EVALUATION

We can compare the different methods for dependent data using some theoretical measures. These are the IF, MBC, and BP. The asymptotic efficiency under the nominal model is also of importance and should be considered, too.

- *The IF:* The definition of the IF for i.i.d. data has been extended to the correlated case by Martin and Yohai in [67] and Künsch in [68]. The two definitions are different, but they are mathematically related [19]. The IFs of some basic robust estimators such as the GM, RA in the case of an AR(1) have been evaluated theoretically in [67]. The IF for the dependent data case is very limited in the statistical literature. Furthermore, there is no practical evaluation for an AR process of order superior than $p = 1$ or for an ARMA(p, q) process with $p \neq 0$ and $q \neq 0$. This is explained by the complexity introduced by the correlation where for an AR(p), for example, one should consider the joint distribution of $(Y_n, Y_{n-1}, \ldots, Y_{n-p})$. The IF curves also change depending on the contamination model, i.e., whether the outliers are isolated, patchy, of replacement, or additive. It is also important to remark that the definition is more general than in the i.i.d. case, since a contamination process does not have to be represented by a Dirac distribution. It can be, for example, a Gaussian process with a different variance and correlation [65], [67]. Furthermore, the validity of the relation of the asymptotic variance or efficiency under the nominal model to the IF is not obvious with the IF given by Martin and Yohai.
- *The MBC:* The same definition as in the i.i.d. case holds for the dependent data case, however, it is more complex and harder to compute, e.g., for an AR(1) process, the joint distribution of (Y_n, Y_{n-1}) should be considered and depends on the type of outliers. The MBC is generally obtained using Monte-Carlo simulations.
- *The BP:* The definition of the BP in the correlated case has not been established in the literature yet [69]. However, the intuitive idea that the BP is a certain percentage of outliers, beyond which the estimator gets stuck at some value (fixed maximum bias) even when adding more outliers, can be used in many practical situations to estimate the BP.

In Figure 6, we depict the IF of the robust MRE and the classical LSE [65] for an AR(1) process. The contamination is a Gaussian replacement process with variance κ^2. The ratio of the standard deviations of the contaminated process over the clean process is denoted as η (i.e., $\eta = \kappa/\sigma$). Figure 6 confirms the robustness and the nonrobustness of the MRE and LSE, respectively [65].

ROBUST FILTERS

Robust filtering constitutes an important field of research with a long and rich history, reaching back even before the mathematical formalization of robust statistics by Huber in 1964. Considering the large amount material on robust filtering, this topic would warrant a tutorial of its own, thus we can only give a brief overview on recent developments without any claim of completeness. Classical and frequently used robust filters are the robust Wiener and Kalman filters, see [13] for a survey up to 1985 and [70], [71] and references therein for some recent advances on robust Kalman filtering. Examples of robust extended Kalman filtering applied to geolocation tracking in wireless networks can be found in [7] and [30]. Apart from these classical filters, there exist further robust filtering frameworks, which have been recently applied to signal processing problems. We can only touch upon some of the principles [72], [73]. In Figure 7, we sketch the principle of robust Kalman filtering for $\mathbf{V}_n$ and $\mathbf{W}_n$ independently distributed and $\mathbf{V}_n$ containing outliers.

The robust estimation of $\hat{\mathbf{X}}_{n|n}$ is obtained for a scalar Y_n, for example, via

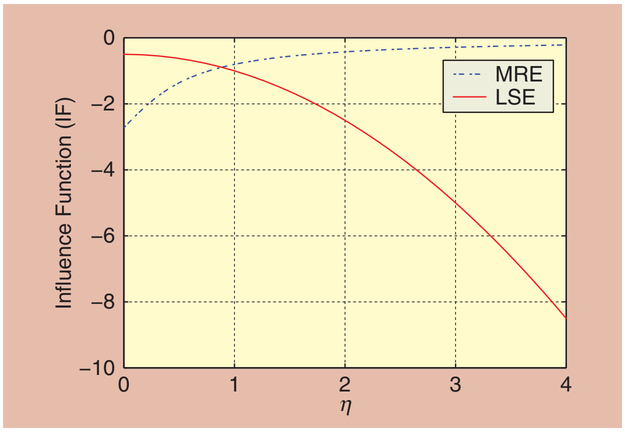
$$\hat{\mathbf{X}}_{n|n} = \hat{\mathbf{X}}_{n|n-1} + \frac{1}{\hat{\sigma}_n}\hat{\boldsymbol{\Sigma}}_{n|n-1}\mathbf{C}^{\mathrm{T}}\psi\left(\frac{\hat{V}_n}{\hat{\sigma}_n}\right),$$

where $\hat{\sigma}_n$ is a robust scale estimate of $\hat{V}_n$ and $\psi(\cdot)$ bounds the influence of $\hat{V}_n$ (see the section "Robust Estimators for Single-Channel Data").

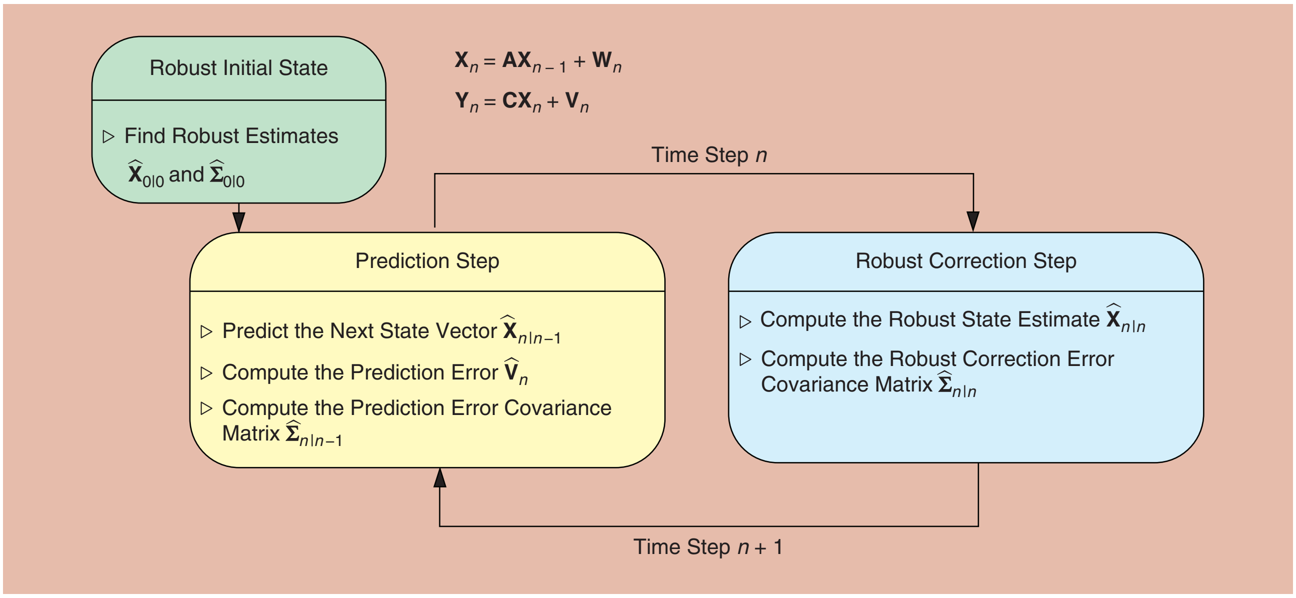
ROBUST SPECTRUM ESTIMATION

A frequently used nonparametric method to estimate the power spectral density (PSD) of a discrete-time signal $X(n)$ that is observed for $n = 0, \ldots N-1$, is based on the periodogram

$$I_{XX}(e^{j\omega}) = \frac{1}{N}\left|\sum_{n=0}^{N-1} X(n)e^{-j\omega n}\right|^2.$$



[FIG6] An example of the IF versus $\eta > 0$, where η is the ratio of the standard deviation of the clean and the contaminated processes for an AR(1) at $\alpha = 0.5$ for a Gaussian replacement outlier process. With growing η, the influence of an outlier on the LSE, which is the MLE for the Gaussian model, grows in an unbounded fashion. For the MRE, which uses robust autocorrelation estimates coupled with a robust filter cleaner, on the other hand, the IF remains bounded.



[FIG7] Principle of the recursive robust Kalman filter. Robust multivariate scatter estimates are used in the initialization of $\hat{\boldsymbol{\Sigma}}_{0|0}$. The effects of $\hat{V}_n$ are bounded in the correction step.

The periodogram is not consistent; its variance can be reduced by averaging periodograms of adjacent blocks or smoothing neighboring periodogram ordinates. In the case of averaged periodograms (Bartlett's estimator), the estimator is computed for M periodograms from equal-length contiguous segments of the data

$$\hat{C}_{XX}(e^{j\omega}) = \frac{1}{M}\sum_{m=1}^{M} I_{XX}(e^{j\omega}; m). \quad (8)$$

A parametric method of PSD estimation, which uses an AR approximation is given by

$$\hat{C}_{XX}^{\mathrm{AR}}(e^{j\omega}) = \frac{\hat{\sigma}^2}{|1-\sum_{k=1}^{\hat{p}} \hat{a}_k e^{-j\omega k}|^2}. \quad (9)$$

Here, $\hat{p}$ is the estimated order of the AR(p) process, $\hat{\mathbf{c}} = (\hat{a}_1, \ldots, \hat{a}_{\hat{p}})^{\top}$ is the coefficient estimate, and $\hat{\sigma}^2$ residual variance estimate. It is well known that these classic estimates of spectra lack robustness against heavy-tailed noise. Both bias and variance of the above estimators are inflated by a considerable amount [74]; see "Practical Evaluation: Simple Robust Nonparametric Spectrum Estimation Example."

One way to introduce robustness into spectral estimation is to use robust filter cleaners [13], [74], [75] that "clean" the data, based on a robust prediction in the state-transition representation as in the section "Robust Filters." Based on the cleaned data, a robust PSD estimate is then obtained using (8) or (9).

A second approach, which has been discussed recently, is to interpret the periodogram as the LSE solution of a (nonlinear) harmonic regression. For details, see [76] and references therein. When the harmonic regression is solved using M-estimation, a so-called M-periodogram is obtained. A special case, for which the asymptotic distribution has been calculated [76], is the L_p-norm regression with $p \in (1, 2)$, where $p = 1$ results in the Laplace periodgram, which coincides with the min-max optimal robust M-periodogram. It has been recently applied to the robust discovery of periodically expressed genes [77]. A tradeoff between robustness and efficiency can be obtained by an appropriate selection of p.

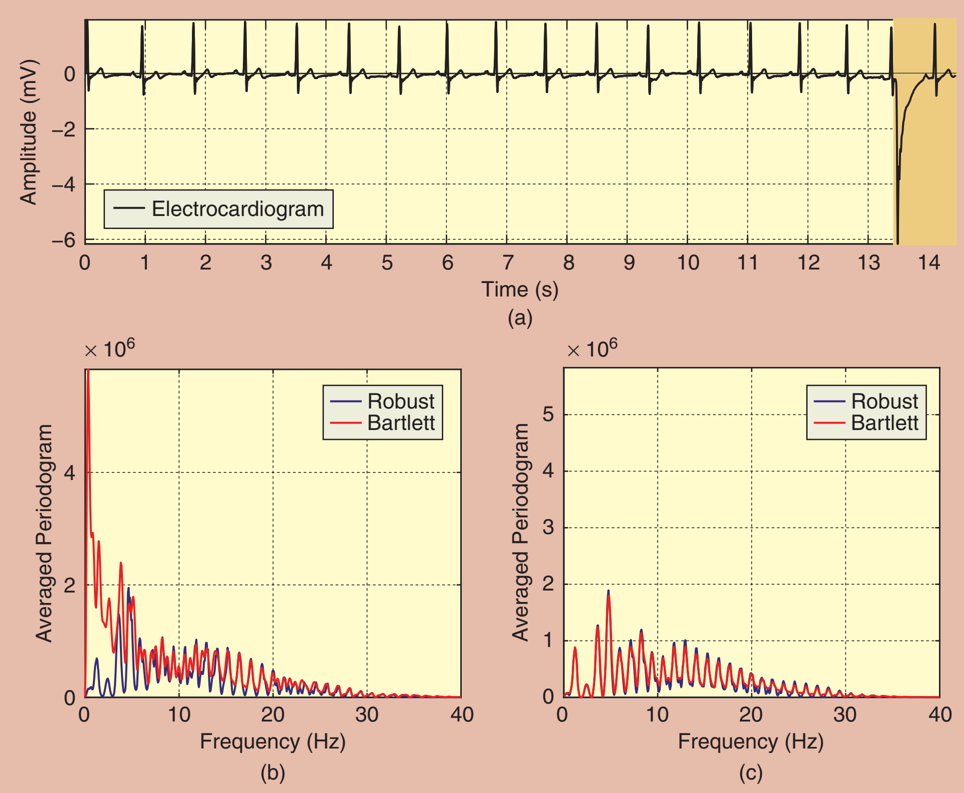
CONCLUSIONS AND PERSPECTIVES

Our objective has been to give a tutorial-style treatment of fundamental concepts of robust statistics for signal processing and to introduce in an accessible manner the lesser-known aspects and most recent trends of robust statistics to the signal processing practitioner. These include robust methods for multichannel complex-valued data and also methods for the challenging case of dependent data. Great attempts have been made to support the rather difficult concepts with examples and real-life applications. Robustness has become an important area of engineering practice, and more emphasis has been given to the design and development of robust systems in recent years. Optimal systems are desirable, however, optimality is harder to achieve with the type of signal disturbance, such as impulsive interference, encountered in today's applications. We have defined measures of robustness, such as the BP, the IF, and the MBC. Using these concepts, we have encouraged the use of robust estimators, such as the M-estimator, for location parameters instead of the conventionally used intuitive sample mean. Parameter estimation in linear regression has been considered because of its importance of modeling many practical problems as linear regression in impulsive interference, such as the given example of geolocation of UE in NLOS observations in wireless communications. We have treated the

PRACTICAL EVALUATION: SIMPLE ROBUST NONPARAMETRIC SPECTRUM ESTIMATION EXAMPLE

In the following, we describe a further nonparametric robust approach [19], and give an illustrative example of an electrocardiogram (ECG) signal. We mention that this simple and computationally cheap robust method can only be expected to work well in situations where less than half of the time segments contain outliers. When using (8), even a small fraction of outliers in one of the segments can dominate and thus spoil the estimate. Hence, Thomson et al. [74] suggested to replace the sample average in (8) by a robust location estimate, e.g., the median, which is approximately min-max bias optimal for this case [19].

Figure S8(a), depicts a locally stationary ECG recording of about 14 s, which is sampled with $f_s = 256$ Hz, detrended and lowpass-filtered with a cutoff frequency of 40 Hz, containing a measurement artifact due to movement of the subject for $t \approx 13.5-14$ s denoted by yellow in the plot. Figure S8(b) shows the averaged periodograms for the whole sample, while (c) depicts the estimates without the artifact. It can be noticed that both estimators give nearly identical results for the clean measurements [Figure S8(c)], while they differ, when computed for the complete ECG recording. The robust estimator provides results similar to clean data case, while Bartlett's estimator changes drastically.



[FIGS8] An illustrative example, where robust nonparametric PSD estimation is applied to (a) an ECG recording containing measurement artifacts due to movement of the subject for $t \approx 13.5-14$ s marked yellow in the plot. Part (b) depicts the PSD estimates for the complete data, while (c) shows estimates without the part containing outliers.

multichannel data case that takes an important place and is becoming increasingly important in signal processing practice, and included the complex-valued case, as encountered, for example, in sensor array processing. The basic concepts of robustness introduced in the independent data case cannot be straightforwardly extended to the dependent data case, which forms the most practical case for signal processing practitioners. We have given definitions of robustness in this case and introduced numerous robust estimators. Extensive applications have been discussed. These include electricity consumption forecast, which is an obvious candidate for robust estimation, and robust spectrum estimation. An area that has not been studied intensively in the literature is robust model selection. This is a topic of great potential for future research, in particular when robust estimation in model-based signal processing, e.g., for ARMA, is used. We left this part out due to space limitations. Also, for the same reasons, we did not touch upon robust methods for machine learning, which we believe is becoming an increasingly important area in machine learning.

AUTHORS

Abdelhak M. Zoubir (zoubir@spg.tu-darmstadt.de) received his Dr.-Ing. degree from Ruhr-Universität Bochum, Germany. He was with Queensland University of Technology, Australia, from 1992 to 1998. He then joined Curtin University of Technology, Australia, as a professor of telecommunications and was interim head of the School of Electrical and Computer Engineering from 2001 to 2003. Since 2003, he has been a professor of signal processing at Technische Universität Darmstadt, Germany. He is an IEEE Distinguished Lecturer (Class of 2010–2011), past chair of the Signal Processing Theory and Methods Technical Committee of the IEEE Signal Processing Society, and he is the editor-in-chief of *IEEE Signal Processing Magazine*. His research interest lies in statistical methods for signal processing applied to telecommunications, radar, sonar, car engine monitoring, and biomedicine. He has published over 300 journal and conference papers in these areas. He is a Fellow of the IEEE.

Visa Koivunen (visa.koivunen@aalto.fi) received the D.Sc degree from the University of Oulu, Finland. He was a visiting researcher at the University of Pennsylvania from 1992 to 1995. Since 1999, he has been a professor at Aalto University (Helsinki University of Technology, Finland), where he is currently an academy professor. He is vice chair and one of the principal investigators in the Smart Radios and Wireless Systems Centre of Excellence in Research nominated by the Academy of Finland. He has been an adjunct faculty member at the University of Pennsylvania and a visiting fellow at Nokia Research Center and Princeton University. His research interests include statistical, communications, and array signal processing. He received the 2007 IEEE Signal Processing Society Best Paper Award. He is an Editorial Board member of *IEEE Transactions on Signal Processing* and a Fellow of the IEEE.

Yacine Chakhchoukh (ychakh@spg.tu-darmstadt.de) received the engineering degree (with honors) from École Polytechnique d'Alger, in 2004, the M.S. degree (with honors) in control and signal processing, and the Ph.D. degree (with honors) from the University of Paris-Sud XI, Paris, in 2005 and 2009, respectively. From June 2006 to September 2009, he was with the French transmission system operator Gestionnaire du Réseau de Transport d'Électricité, working on statistical robust signal processing for load forecasting. He was a postdoctoral research fellow with the Signal Processing Group at the Institute of Telecommunications, Technische Universität Darmstadt, Germany, from 2010 to 2011, where he worked on statistical robust estimation applied to practical problems such as communications.

Michael Muma (muma@spg.tu-darmstadt.de) received the Dipl.-Ing. degree in electrical engineering and information technology from the Darmstadt University of Technology. He completed his diploma thesis with the Contact Lens and Visual Optics Laboratory, School of Optometry, Brisbane, Australia, in the role of cardiopulmonary signals in the dynamics of the eye's wavefront aberrations. He is currently working toward his Ph.D. degree in the Signal Processing Group, Institute of Telecommunications, Darmstadt University of Technology. His research is on robust statistics for signal processing with an emphasis on correlated data and model selection. He has been chair of the IEEE Signal Processing Society Signal Processing Theory and Methods Student Subcommittee since 2011.

REFERENCES

[1] K. Kim and G. Shevlyakov, "Why Gaussianity?" *IEEE Signal Processing Mag.*, vol. 25, no. 2, pp. 102–113, 2008.

[2] T. K. Blankenship, D. M. Kriztman, and T. S. Rappaport, "Measurements and simulation of radio frequency impulsive noise in hospitals and clinics," in *Proc. IEEE 47th Vehicular Technology Conf.*, 1997, vol. 3, pp. 1942–1946.

[3] Y. I. Abramovich and P. Turcaj, "Impulsive noise mitigation in spatial and temporal domains for surface-wave over-the horizon radar," DTIC Document, Cooperative Research Centre for Sensor Signal and Information Processing, Mawson Lakes, Australia 1999.

[4] D. Middleton, "Non-Gaussian noise models in signal processing for telecommunications: New methods and results for class A and class B noise models," *IEEE Trans. Inform. Theory*, vol. 45, no. 4, p. 1129, 1999, pp. 1129–1149.

[5] P. C. Etter, *Underwater Acoustic Modeling and Simulation*. New York: Taylor & Francis, 2003.

[6] D. C. Alexander, G. J. Barker, and S. R. Arridge, "Detection and modeling of non-Gaussian apparent diffusion coefficient profiles in human brain data," *Magn. Reson. Med.*, vol. 48, no. 2, pp. 331–340, 2002.

[7] U. Hammes, E. Wolsztynski, and A. M. Zoubir, "Robust tracking and geolocation for wireless networks in NLOS environments," *IEEE J. Select. Topics Signal Processing*, vol. 3, no. 5, pp. 889–901, Oct. 2009.

[8] Y. Chakhchoukh, P. Panciatici, and L. Mili, "Electric load forecasting based on statistical robust methods," *IEEE Trans. Power Syst.*, vol. 26, no. 3, pp. 982–991, 2010.

[9] M. Nassar, K. Gulati, A. K. Sujeeth, N. Aghasadeghi, B. L. Evans, and K. R. Tinsley, "Mitigating near-field interference in laptop embedded wireless transceivers," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, 2008, pp. 1405–1408.

[10] P. J. Huber and E. M. Ronchetti, *Robust Statistics*. Hoboken, NJ: Wiley, 2009.

[11] S. M. Stigler, "Simon Newcomb, Percy Daniell and the history of robust estimation 1885–1920," *J. Amer. Stat. Assoc.*, vol. 68, no. 344, pp. 872–879, 1973.

[12] F. R. Hampel, E. M. Ronchetti, P. J. Rousseeuw, and W. A. Stahel, *Robust Statistics: The Approach Based on Influence Functions*. New York: Wiley, 1986.

[13] S. A. Kassam and H. V. Poor, "Robust techniques for signal processing: A survey," *Proc. IEEE,* vol. 73, no. 3, pp. 433–481, 1985.

[14] S. A. Kassam, *Signal Detection in Non-Gaussian Noise* (Springer Texts in Electrical Engineering). New York: Springer-Verlag, 1988.

[15] H. V. Poor, *An Introduction to Signal Detection and Estimation*. Berlin: Springer-Verlag, 1994.

[16] C. V. Stewart, "Robust parameter estimation in computer vision," *SIAM Rev.*, vol. 41, no. 3, pp. 513–537, 1999.

[17] G. Arce, *Nonlinear Signal Processing: A Statistical Approach*. Hoboken, NJ: Wiley, 2004.

[18] P. C. Chandrasekharan, *Robust Control of Linear Dynamical Systems*. New York: Academic, 1996.

[19] R. A. Maronna, R. D. Martin, and V. J. Yohai, *Robust Statistics: Theory and Methods* (Wiley Series in Probability and Statistics). Hoboken, NJ: Wiley, 2006.

[20] E. Eskin, "Anomaly detection over noisy data using learned probability distributions," in *Proc. Int. Conf. Machine Learning*, 2000, pp. 255–262.

[21] S. Chawla, D. Hand, and V. Dhar, "Outlier detection special issue," *Data Min. Knowl. Discov.*, vol. 20, no. 2, pp. 189–190, 2010.

[22] P. J. Rousseeuw and A. M. Leroy, *Robust Regression and Outlier Detection*. New York: Wiley, 1987.

[23] V. Hodge and J. Austin, "A survey of outlier detection methodologies," *Artif. Intell. Rev.*, vol. 22, no. 2, pp. 85–126, 2004.

[24] C. L. Nikias and M. Shao, *Signal Processing with Alpha-Stable Distributions and Applications*. New York: Wiley, 1995.

[25] F. Ruggeri, "Bayesian robustness," *Newslett. Eur. Working Group Multiple Criteria Decis. Aiding*, vol. 3, no. 17, pp. 6–10, 2008.

[26] R. Brcich, C. L. Brown, and A. M. Zoubir, "An adaptive robust estimator for scale in contaminated distributions," in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP)*. New York: IEEE, 2004, pp. 1049–1052.

[27] X. Wang and H. V. Poor, "Robust multiuser detection in non-Gaussian channels," *IEEE Trans. Signal Processing,* vol. 47, no. 2, pp. 289–305, 1999.

[28] A. M. Zoubir and R. F. Brcich, "Multiuser detection in heavy tailed noise," *Digital Signal Processing,* vol. 12, no. 2–3, pp. 262–273, 2002.

[29] T.A. Kumar and K. D. Rao, "A new M-estimator based robust multiuser detection in flat-fading non-Gaussian channels," *IEEE Trans. Commun.*, vol. 57, no. 7, pp. 1908–1913, 2009.

[30] I. Guvenc and C. C. Chong, "A survey on TOA based wireless localization and NLOS mitigation techniques," *IEEE Commun. Surveys Tutorials,* vol. 11, no. 3, pp. 107–124, 2009.

[31] J. C. Prieto, C. Croux, and A. R. Jiménez, "RoPEUS: A new robust algorithm for static positioning in ultrasonic systems," *Sensors*, vol. 9, no. 6, pp. 4211–4229, 2009.

[32] M. Ye, R. M. Haralick, and L. G. Shapiro, "Estimating piecewise-smooth optical flow with global matching and graduated optimization," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 25, no. 12, pp. 1625–1630, 2003.

[33] L. Mili, M. G. Cheniae, and P. J. Rousseeuw, "Robust state estimation of electric power systems," *IEEE Trans. Circuits Syst. I*, vol. 41, no. 5, pp. 349–358, 2002.

[34] L. Thomas and L. Mili, "A robust GM-estimator for the automated detection of external defects on barked hardwood logs and stems," *IEEE Trans. Signal Processing*, vol. 55, no. 7, pp. 3568–3576, 2007.

[35] C. G. Bénar, D. Schön, S. Grimault, B. Nazarian, B. Burle, M. Roth, J. M. Badier, P. Marquis, C. Liegeois-Chauvel, and J. L. Anton, "Single-trial analysis of oddball event-related potentials in simultaneous EEG-fMRI," *Hum. Brain Mapp.*, vol. 28, no. 7, pp. 602–613, 2007.

[36] D. S. Pham and A. M. Zoubir, "A sequential algorithm for robust parameter estimation," *IEEE Signal Processing Lett.*, vol. 12, no. 1, pp. 21–24, 2004.

[37] I. Markovsky and S. Van Huffel, "Overview of total least-squares methods," *Signal Process.*, vol. 87, no. 10, pp. 2283–2302, 2007.

[38] U. Hammes, E. Wolsztynski, and A.M. Zoubir, "Transformation-based robust semiparametric estimation," *IEEE Signal Processing Lett.*, vol. 15, no. 1, pp. 845–848, 2008.

[39] D. S. Pham, A. M. Zoubir, R. F. Brcic, and Y. H. Leung, "A nonlinear M-estimation approach to robust asynchronous multiuser detection in non-Gaussian noise," *IEEE Trans. Signal Processing*, vol. 55, no. 5, pp. 1624–1633, 2007.

[40] P. Stoica, H. Li, and J. Li, "Amplitude estimation of sinusoidal signals: Survey, new results, and an application," *IEEE Trans. Signal Processing*, vol. 48, no. 2, pp. 338–352, 2002.

[41] A. F. Siegel, "Robust regression using repeated medians," *Biometrika*, vol. 69, no. 1, pp. 242–244, 1982.

[42] P. J. Rousseeuw and V. J. Yohai, "Robust regression by means of S-estimators," *Robust Nonlinear Time Ser. Anal.*, vol. 26, no. 1, pp. 256–272, 1984.

[43] V. J. Yohai and R. H. Zamar, "High breakdown-point estimates of regression by means of the minimization of an efficient scale," *J. Amer. Stat. Assoc.*, vol. 83, no. 402, pp. 406–413, 1988.

[44] F. Gustafsson and F. Gunnarsson, "Mobile positioning using wireless networks: Possibilities and fundamental limitations based on available wireless network measurements," *IEEE Signal Processing Mag.*, vol. 22, no. 4, pp. 41–53, 2005.

[45] M. Salibian-Barrera and V. J. Yohai, "A fast algorithm for S-regression estimates," *J. Comput. Graph. Stat.*, vol. 15, no. 2, pp. 414–427, 2006.

[46] P. Schreier and L. Scharf, *Statistical Signal Processing of Complex-Valued Data: The Theory of Improper and Noncircular Signals*. Cambridge, U.K.: Cambridge Univ. Press, 2010.

[47] E. Ollila and V. Koivunen, "Complex ICA using generalized uncorrelating transform," *Signal Process.*, vol. 89, no. 4, pp. 365–377, 2009.

[48] R. J. Kozick and B. M. Sadler, "Maximum-likelihood array processing in non-Gaussian noise with Gaussian mixtures," *IEEE Trans. Signal Processing*, vol. 48, no. 12, pp. 3520–3535, 2002.

[49] M. Hubert, P. J. Rousseeuw, and S. Van Aelst, "High-breakdown robust multivariate methods," *Stat. Sci.*, vol. 23, no. 1, pp. 92–119, 2008.

[50] M. Salibian-Barrera, S. Van Aelst, and G. Willems, "Principal components analysis based on multivariate MM estimators with fast and robust bootstrap," *J. Amer. Stat. Assoc.,* vol. 101, no. 475, pp. 1198–1211, 2006.

[51] E. Ollila and V. Koivunen, "Influence functions for array covariance matrix estimators," in *Proc. IEEE Workshop Statistical Signal Processing (SSP)*. New York: IEEE, 2003, pp. 462–465.

[52] S. Visuri, V. Koivunen, and H. Oja, "Rank and sign covariance matrices," *J. Stat. Plann. Inference*, vol. 97, no. 2, pp. 557–575, 2000.

[53] K. T. Fang, S. Kotz, and K. W. Ng, *Symmetric Multivariate and Related Distributions*. London, U.K.: Chapman & Hall, London, 1990.

[54] J. Astola, P. Haavisto, and Y. Neuvo, "Vector median filters," *Proc. IEEE*, vol. 78, no. 4, pp. 678–689, 1990.

[55] V. Koivunen, N. Himayat, and S. A. Kassam, "Nonlinear filtering techniques for multivariate images—Design and robustness characterization," *Signal Process.*, vol. 57, no. 1, pp. 81–91, 1997.

[56] V. Koivunen, "Nonlinear filtering of multivariate images under robust error criterion," *IEEE Trans. Image Processing*, vol. 5, no. 6, pp. 1054–1060, June 1996.

[57] S. Kassam and J. Thomas, "A class of nonparametric detectors for dependent input data," *IEEE Trans. Inform. Theory*, vol. 21, no. 4, pp. 431–437, 1975.

[58] Y. Chakhchoukh, "A new robust estimation method for ARMA models," *IEEE Trans. Signal Processing*, vol. 58, no. 7, pp. 3512–3522, July 2010.

[59] R. Weron, *Modeling and Forecasting Electricity Loads and Prices: A Statistical Approach*. Hoboken, NJ: Wiley 2006.

[60] A. J. Fox, "Outliers in time series," *J. Roy. Stat. Soc. B*, vol. 34, pp. 350–363, 1972.

[61] R. S. Tsay, "Outliers, level shifts, and variance changes in time series," *Int. J. Forecast.*, vol. 7, no. 3, pp. 1–20, 1988.

[62] G. E. P. Box, G. M. Jenkins, and G. C. Reinsel, *Time Series Analysis: Forecasting and Control* (Wiley Series in Probability and Statistics), 4th ed. Hoboken, NJ: Wiley, 2008.

[63] R. H. Jones, "Maximum likelihood fitting of ARMA models to time series with missing observations," *Technometrics*, vol. 22, no. 3, pp. 389–395, 1980.

[64] O. H. Bustos and V. J. Yohai, "Robust estimates for ARMA models," *J. Amer. Stat. Assoc.*, vol. 81, no. 393, pp. 155–168, 1986.

[65] Y. Chakhchoukh, P. Panciatici, and P. Bondon, "Robust estimation of SARIMA models: Application to short-term load forecasting," in *Proc. IEEE Workshop Statistical Signal Processing (SSP)*, Cardiff, U.K., Aug. 2009, pp. 77–80.

[66] N. Muler, D. Pena, and V. J. Yohai, "Robust estimation for ARMA models," *Ann. Stat.*, vol. 37, no. 2, pp. 816–840, 2009.

[67] R. D. Martin and V. J. Yohai, "Influence functionals for time series," *Ann. Stat.*, vol. 14, no. 3, pp. 781–855, 1986.

[68] H. Künsch, "Infinitesimal robustness for autoregressive processes," *Ann. Stat.*, vol. 12, no. 3, pp. 843–863, 1984.

[69] M. G. Genton and A. Lucas, "Comprehensive definitions of breakdown points for independent and dependent observations," *J. Roy. Stat. Soc. B*, vol. 65, no. 1, pp. 81–94, 2003.

[70] Y. Yang, H. He, and G. Xu, "Adaptively robust filtering for kinematic geodetic positioning," *J. Geodesy*, vol. 75, no. 2, pp. 109–116, 2001.

[71] M. A. Gandhi and L. Mili, "Robust Kalman filter based on a generalized maximum-likelihood-type estimator," *IEEE Trans. Signal Processing*, vol. 58, no. 5, pp. 2509–2520, 2010.

[72] T. C. Aysal and K. E. Barner, "Meridian filtering for robust signal processing," *IEEE Trans. Signal Processing*, vol. 55, no. 8, pp. 3949–3962, 2007.

[73] H. Dong, Z. Wang, and H. Gao, "Robust H_∞ filtering for a class of nonlinear networked systems with multiple stochastic communication delays and packet dropouts," *IEEE Trans. Signal Processing*, vol. 58, no. 4, pp. 1957–1966, 2010.

[74] R. D. Martin and D. J. Thomson, "Robust-resistant spectrum estimation," *Proc. IEEE*, vol. 70, no. 9, pp. 1097–1115, 1982.

[75] B. Spangl and R. Dutter, "Estimating spectral density functions robustly," *REVSTAT–Stat. J.*, vol. 5, no. 1, pp. 41–61, 2007.

[76] T. H. Li, "A nonlinear method for robust spectral analysis," *IEEE Trans. Signal Processing*, vol. 58, no. 5, pp. 2466–2474, 2010.

[77] K. Liang, X. Wang and T. H. Li, "Robust discovery of periodically expressed genes using the Laplace periodogram," *BMC Bioinform.*, vol 10, no. 1, pp. 1–15, 2009.

[78] R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. Antennas Propagat.*, vol. 34, no. 3, pp. 276–280, 1986.

[79] S. Visuri, H. Oja, and V. Koivunen, "Nonparametric method for subspace based frequency estimation," in *Proc. 10th European Signal Processing Conf. (EUSIPCO 2000)*, Tampere, Finland, 2000, pp. 1261–1264.

[SP]