Robust Estimation of Noise Standard Deviation in Presence of Signals with Unknown Distributions and Occurrences

Dominique Pastor, Francois-Xavier Socheleau

Abstract

In many applications, $d$-dimensional observations result from the random presence or absence of random signals in independent and additive white Gaussian noise. An estimate of the noise standard deviation can then be very useful to detect or to estimate these signals, especially when standard likelihood theory cannot be applied because of too little prior knowledge about the signal probability distributions. The present paper introduces a new scale estimator of the noise standard deviation when the noisy signals have unknown probability distributions and unknown probabilities of presence less than or equal to one half. The latter assumption can be regarded as to a weak assumption of sparsity. Applied to the detection of non-cooperative radio-communications, this new estimator outperforms the standard MAD and its alternatives as well as the trimmed and winsorized robust scale estimators. The MATLAB code corresponding to the proposed estimator is available at http://perso.telecom-bretagne.eu/pastor.

Index Terms

Robust statistics, sparsity, MAD estimation, trimmed estimation, winsorized estimation, Communication Electronic Support, cognitive radio.
I. INTRODUCTION

In statistical signal processing applications, observations are often $d$-dimensional random vectors assumed to result from the random presence of signals in independent and additive white Gaussian noise (AWGN). As usual, noise is supposed to be centred. In many practical cases, this simple but reasonable model raises two simultaneous difficulties for detecting or estimating the noisy signals. On the one hand, very little is generally known about the signals or about most of their describing parameters so that the probability distributions of the signals are partially or definitely unknown. On the other hand, the noise standard deviation is often unknown and must be estimated so as to process the observations.

Such situations where the distributions of the signals to detect are not fully known and the noise standard deviation must be estimated are frequently encountered in sonar and radar processing because the echoes, received by a sonar or a radar system, result from a convolution between a possibly known transmitted pulse and an often unknown environment. A second typical example, discussed in the experimental part of this paper, is that of spectrum sensing performed in a context of communication electronic support (CES) [1] or cognitive radio [2]. In such applications, the detection of some signals of interest must be performed on the basis of an observation, which is usually a noisy mixture of signals with unknown distributions and occurrences.

Since very little is generally known about the signals, it is often relevant to estimate the noise standard deviation via robust scale estimators. Such estimators are said to be robust in the sense that they are not excessively affected by small departures from the assumed model. In our case, the model is the Gaussian distribution of the noise and the signals are considered as outliers with respect to this model. More specifically, in our specific signal processing context, we can paraphrase [3] and say that 1) “the target distribution [...] is taken to be some unspecified normal distribution” — in fact, with null mean and unknown variance — for the main core of the data, 2) for outliers, “[...] no alternative distribution for these [...] observations is specified” and 3) outliers are signals with unknown distributions and occurrences in noise.

There exists three main families of robust scale estimators: M, L and R [4]. They correspond, respectively, to maximum likelihood type estimates, linear combinations of order statistics and estimates derived from rank tests. For the applications previously listed, L-estimators are usually preferred since they are usually explicit, easy to implement and not computationally costly. There is a wide range of L-estimators of scale. Some are outlined in Appendix A of this paper and more can be
found in [5] or [6]. Among them, the median absolute deviation about the median (MAD) estimator and the trimmed estimator (T-estimator) are very popular. The MAD popularity comes from its high asymptotic breakdown point (50%) and its influence function bounded by the sharpest possible bound among all scale estimators [7]. It was also defined in [4] as “the single most useful ancillary estimate of scale” so that it is often used as an initial value for more efficient robust estimators such as M-estimators. The T-estimator is obtained by excluding some of the extreme values and is very accurate when the outliers are very large and when their number in the observation is known a priori. Despite the good characteristics (high breakdown point, good efficiency etc.) of most of the conventional L-estimators of scale, their performance degrades significantly when the proportion of outliers increases. As an alternative to these estimators, this paper addresses the problem of estimating the noise standard deviation in signal processing applications where the number or the amplitudes of the outliers are too large for these scale estimators to perform well.

In this respect, we hereafter introduce a new type of L-estimator of scale. This estimator is called the $d$-dimensional amplitude trimmed estimator (DATE) because, in contrast to the conventional T-estimator that requires fixing the amount of one-dimensional outliers to remove, it trims the amplitudes or norms of the $d$-dimensional observations by assuming that the signal norms are above some known lower-bound and that the signal probabilities of occurrence are less than one half. These assumptions take into account that, in our signal processing context, a physical model for outliers is that of signals in noise. Crucially, these assumptions bound our lack of prior knowledge about the signals and mean that the signals and noise can be somewhat “separated”. However, these assumptions are also constraints, indeed convenient for statistical signal processing applications, but stronger than those required by standard robust estimators. It is worth emphasizing that such hypotheses are particularly suitable in case of observations returned by a sparse transform, that is, a transform representing a signal by coefficients that are mostly small except a few ones that have large amplitudes.

More specifically, the DATE relies on a limit theorem established below. This theoretical result concerns observations resulting from signals with unknown probability distributions and unknown probabilities of presence less than or equal to one half in independent AWGN. It is similar to [8, Theorem 1] and involves the same type of hypotheses as those mentioned above and aims at bounding our lack of prior knowledge about the amplitudes and occurrences of the signals. The

$^1$as suggested by an anonymous reviewer
computational cost of the DATE is low. It requires prior knowledge of the so-called minimum signal to noise ratio. For signal processing applications, the DATE can be regarded as an alternative to the MAD estimator and other standard robust estimators that do not require prior knowledge on the signals and their distributions. This is the reason why we test the DATE in a spectrum sensing context, which is a rather natural application of such an estimator.

The next section is dedicated to establishing Theorem 1, which guides the design of the DATE. This estimator is presented in Section III, after exhibiting a simple cost derived from Theorem 1 and establishing a convenient lemma for the minimization of this cost. The application to spectrum sensing for cognitive radio and CES is then addressed in Section IV. Issues and prospects opened by the DATE are discussed in the concluding part of the paper, namely, Section V.

II. THEORETICAL BACKGROUND

Signal processing applications often involve sequences of $d$-dimensional real random observations such that each observation is either the sum of some random signal and independent noise or noise alone. In most cases, noise is reasonably assumed to be “white and Gaussian” in that it is Gaussian distributed with null mean and covariance matrix proportional to the $d \times d$ identity matrix $I_d$. Summarizing, we say that each observation results from some signal randomly present or absent in independent AWGN. For reasons described in the introduction, the problem is the estimation of the noise standard deviation in the general case where the probability distributions of the signals are unknown. In this respect, Theorem 1 established below states that, when the observations are independent and the probabilities of presence of the signals are upper-bounded by some value in $[0,1)$, the noise standard deviation is the only positive real number satisfying a specific convergence criterion when the number of observations and the so-called minimum signal to noise ratio tend to infinity. This convergence involves neither the probability distributions nor the probabilities of presence of the signals. This result partly follows from and is very similar to [8, Theorem1]. However, to derive the DATE in Section III, the convergence involved in Theorem 1 is more convenient than that of [8, Theorem1]. Theorem 1 is stated in Section II-B, on the basis of the material presented in the next section and used throughout.

A. Preliminary material

Every random vector and every random variable encountered hereafter is assumed to be real-valued and defined on the same probability space $(\Omega, \mathcal{F}, P)$ for every $\omega \in \Omega$. As usual, if a property $\mathcal{P}$ holds true almost surely, we write $\mathcal{P}$ (a-s). Every random vector considered below is $d$-dimensional.
The set of all \(d\)-dimensional real random vectors, that is, the set of all measurable maps of \(\Omega\) into \(\mathbb{R}^d\), is denoted by \(\mathcal{M}(\Omega,\mathbb{R}^d)\) and the sequences of \(d\)-dimensional random vectors defined on \(\Omega\) is denoted by \(\mathcal{M}(\Omega,\mathbb{R}^d)^N\). In what follows, \(\| \cdot \|\) is the standard Euclidean norm in \(\mathbb{R}^d\). For every given random vector \(Y : \Omega \to \mathbb{R}^d\) and any \(\tau \in \mathbb{R}\), the notation \(I(\|Y\| \leq \tau)\) stands for the indicator function of the event \(\{\|Y\| \leq \tau\}\). If \(Y\) is any \(d\)-dimensional random vector (resp. any random variable), the probability that \(Y\) belongs to some Borel set \(A\) of \(\mathbb{R}^d\) (resp. \(\mathbb{R}\)) is denoted by \(P[Y \in A]\).

Given some positive real number \(\sigma_0\), we say that a sequence \(X = (X_n)_{n \in \mathbb{N}}\) of \(d\)-dimensional real random vectors is a \textit{\(d\)-dimensional white Gaussian noise} (WGN) with standard deviation \(\sigma_0\) if the random vectors \(X_n, \ n \in \mathbb{N}\), are independent and identically Gaussian distributed with null mean vector and covariance matrix \(\sigma_0^2 I_d\). The \textit{minimum signal to noise ratio} (SNR) with respect to noise \(X\) of an element \(\Lambda = (\Lambda_n)_{n \in \mathbb{N}}\) of \(\mathcal{M}(\Omega,\mathbb{R}^d)^N\) is defined as the supremum \(\varrho(\Lambda)\) of the set of those \(\rho \in [0,\infty]\) such that, for every natural number \(n\), \(\|\Lambda_n\|\) is larger than or equal to \(\rho \sigma_0\) (a.s):

\[
\varrho(\Lambda) = \sup\{\rho \in [0,\infty]: \forall n \in \mathbb{N}, \|\Lambda_n\| \geq \rho \sigma_0\ \text{a.s}\}\. (1)
\]

The minimum SNR has some properties that are easy to verify. First, for every given \(\Lambda = (\Lambda_n)_{n \in \mathbb{N}},\ |\Lambda_n| \geq \varrho(\Lambda)\sigma_0\) (a-s) for every \(n \in \mathbb{N}\); second, given \(\rho \in [0,\infty], \varrho(\Lambda) \geq \rho\) if and only if, for every \(n \in \mathbb{N},\ |\Lambda_n| \geq \rho \sigma_0\) (a-s). If \(f\) is some map of \(\mathcal{M}(\Omega,\mathbb{R}^d)^N\) into \(\mathbb{R}\), we will then say that the limit of \(f\) is \(\ell \in \mathbb{R}\) when \(\varrho(\Lambda)\) tends to \(\infty\) and write that \(\lim_{\varrho(\Lambda) \to \infty} f(\Lambda) = \ell\) if

\[
\lim_{\rho \to \infty} \sup\{\|f(\Lambda) - \ell\|: \Lambda \in \mathcal{M}(\Omega,\mathbb{R}^d)^N, \varrho(\Lambda) \geq \rho\} = 0,
\]

that is, if, for any positive real value \(\eta\), there exists some \(\rho_0 \in (0,\infty)\) such that, for every \(\rho \geq \rho_0\) and every \(\Lambda \in \mathcal{M}(\Omega,\mathbb{R}^d)^N\) such that \(\varrho(\Lambda) \geq \rho\), we have \(\|f(\Lambda) - \ell\| \leq \eta\).

Given some non-negative real number \(a\), \(L^a(\Omega,\mathbb{R}^d)\) stands for the set of those \(d\)-dimensional real random vectors \(Y : \Omega \to \mathbb{R}^d\) for which \(\mathbb{E}[\|Y\|^a] < \infty\). We hereafter deal with the set of those elements \(\Lambda = (\Lambda_n)_{n \in \mathbb{N}}\) of \(\mathcal{M}(\Omega,\mathbb{R}^d)^N\) such that \(\Lambda_n \in L^a(\Omega,\mathbb{R}^d)\) for every \(n \in \mathbb{N}\) and \(\sup_{n \in \mathbb{N}} \mathbb{E}[\|\Lambda_n\|^a]\) is finite. This set is hereafter denoted \(\ell^\infty(\mathbb{N},L^a(\Omega,\mathbb{R}^d))\).

A \textit{thresholding function} is any non-decreasing continuous and positive real function \(\theta : [0,\infty) \to (0,\infty)\) such that \(\theta(\rho) = C \rho + \gamma(\rho)\), where \(0 < C < 1\), \(\gamma(\rho)\) is positive for sufficiently large values of \(\rho\) and \(\lim_{\rho \to \infty} \gamma(\rho) = 0\).

\[B. \ A \ limit \ theorem\]

On the basis of the material proposed above, it is easy to model a sequence of observations where random signals are either present or absent in independent AWGN with standard deviation
σ_0. It suffices to consider a sequence \( Y = (Y_n)_{n \in \mathbb{N}} \) such that, for every \( n \in \mathbb{N} \), \( Y_n = \varepsilon_n \Lambda_n + X_n \) where \( \varepsilon = (\varepsilon_n)_{n \in \mathbb{N}} \) is a sequence of random variables valued in \([0, 1]\), \( \Lambda = (\Lambda_n)_{n \in \mathbb{N}} \) stands for some sequence of \( d \)-dimensional real random vectors and \( X = (X_n)_{n \in \mathbb{N}} \) is some \( d \)-dimensional WGN with standard deviation \( \sigma_0 \). We write \( Y = \varepsilon \Lambda + X \) and, in this model, \( Y \) is the sequence of observations, \( \Lambda \) is the sequence of signals that can be observed in independent AWGN represented by \( X \) and \( \varepsilon_n \) models the possible occurrence of \( \Lambda_n \). To state the following result, from which our estimator derives, we must introduce the following notation: given a sequence \((x_n)_{n \in \mathbb{N}}\) of non-negative real values, two non negative real values \( r \) and \( s \) such that \( r > s \geq 0 \), some non-negative real value \( \tau \) and some sample size \( N \in \mathbb{N} \), we define the ratio between the \( r \)th-order and the \( s \)th-order sample moments of the data \( x_1, x_2, \ldots, x_N \) censored with threshold height \( \tau \) by

\[
M_{[x_1, x_2, \ldots, x_N]}(\tau) = \frac{\sum_{n=1}^{N} x_n^r I(x_n \leq \tau)}{\sum_{n=1}^{N} x_n^s I(x_n \leq \tau)}. \tag{3}
\]

This ratio can always be defined in \([0, \infty[\) because it can be extended by continuity by setting \( \sum_{n=1}^{N} x_n^0/\sum_{n=1}^{N} x_n^s = 0 \) for \((x_1, \ldots, x_N) = (0, \ldots, 0)\). We can now state the following result, which partially derives from [8, Theorem 1].

**Theorem 1:** Let \( Y = (Y_n)_{n \in \mathbb{N}} \in \mathcal{M}(\Omega, \mathbb{R}^d)^\mathbb{N} \) be a sequence of independent random vectors such that \( Y = \varepsilon \Lambda + X \) where \( \Lambda = (\Lambda_n)_{n \in \mathbb{N}} \in \mathcal{M}(\Omega, \mathbb{R}^d)^\mathbb{N} \) is a sequence of \( d \)-dimensional random signals, \( X = (X_n)_{n \in \mathbb{N}} \) is some \( d \)-dimensional WGN with standard deviation \( \sigma_0 \) and \( \varepsilon = (\varepsilon_n)_{n \in \mathbb{N}} \) is a sequence of random variables valued in \([0, 1]\) respectively.

Assume that

(A1) for every \( n \in \mathbb{N} \), \( \Lambda_n \), \( X_n \) and \( \varepsilon_n \) are independent;

(A2) the set of priors \( \{P[\varepsilon_n = 1]: n \in \mathbb{N}\} \) has an an upper bound \( p \in [0, 1) \) and the random variables \( \varepsilon_n \), \( n \in \mathbb{N} \), are independent;

(A3) there exists some \( v \in (0, \infty) \) such that \( \Lambda \in \ell^\infty([0, 1], \mathcal{L}^v(\Omega, \mathbb{R}^d)) \).

Then, for any two given non-negative real numbers \( r \) and \( s \) such that \( v/2 \geq r > s \geq 0 \) and any thresholding function \( \theta \), \( \sigma_0 \) is the unique positive real number \( \sigma \) such that, for every \( \beta_0 \in (0, 1] \),

\[
\lim_{\beta(\Lambda) \to \infty} \limsup_{N \to \infty} \left| \frac{\|M_{[\|Y_1\|,\|Y_2\|,\ldots,\|Y_N\|]}(\beta \sigma \theta(\rho(\Lambda))) - \lambda \sigma^{r-s}\|}{\lambda} \right| = 0 \tag{4}
\]

uniformly in \( \beta \in [\beta_0, 1] \), where \( \lambda = 2^{\frac{r-s}{2}} \Gamma\left(\frac{d+2r}{2}\right)/\Gamma\left(\frac{d+s}{2}\right) \) and \( \Gamma \) is the standard Gamma function.

**Proof:** See appendix B.
At this stage, some comments are required. To begin with, we recall that assumption (A1) models observations due to the random presence or absence of signals in independent AWGN: for any given \( k \in \mathbb{N} \), the independence between the signal \( \Lambda_k \) and noise \( X_k \) is usual and, on the other hand, the independence between \( \varepsilon_k \) and \( \Lambda_k \) models that the state of nature about the presence or absence of any signal is chosen independently of the possible signals. Assumption (A2) can be regarded as a weak assumption of sparsity because it imposes that the signals are not always present, without imposing necessary small probabilities of occurrence for the signals. In particular, in the sequel, the case \( p = 1/2 \) will be considered. It corresponds to the case of signals less often present than absent, which is a reasonable assumption in practice. Assumption (A3) means that the \( \nu \)th powers of the signal norms remain bounded by some value. The specific case \( \nu = 2 \) is of most interest in practice since it corresponds to random signals with finite energy upper-bounded by some finite value.

The result of Theorem 1 can be seen as a method of moments, as illustrated by the following heuristic approach. Let us consider the sample moments

\[
\mu_{N,\gamma}(\tau) = \frac{1}{N} \sum_{n=1}^{N} \| Y_n \|^\gamma \mathbb{I}(\| Y_n \| \leq \tau), \quad \gamma \in [0, \infty)
\]

for any \( \tau \in [0, \infty) \) and any \( N \in \mathbb{N} \). If the random vectors \( \Lambda_n \) are independent and identically distributed (iid), the strong law of large numbers makes it possible to justify the approximation

\[
M\{\| Y_1 \|, \| Y_2 \|, \ldots, \| Y_N \|\}(\tau) = \mu_{N,r}(\tau) / \mu_{N,s}(\tau) \approx \mathbb{E}\left[ \| Y_n \|^\gamma \mathbb{I}(\| Y_n \| \leq \tau) \right] / \mathbb{E}\left[ \| Y_n \|^s \mathbb{I}(\| Y_n \| \leq \tau) \right].
\]

On the other hand, when each \( \| \Lambda_n \| / \sigma_0 \) is large enough, we can expect the existence of a suitable threshold height \( \tau \) capable of distinguishing observations where the signals are present from observations where noise only is present, so that we can approximatively write, thanks to this threshold,

\[
\mathbb{E}\left[ \| Y_n \|^\gamma \mathbb{I}(\| Y_n \| \leq \tau) \right] \approx \mathbb{E}\left[ \| X_n \|^\gamma \mathbb{I}(\| X_n \| \leq \tau) \right] \mathbb{P}[\varepsilon_n = 0]
\]

for any \( \gamma \in [0, \infty) \). By combining this approximation with (5) and taking into account that \( \| X_k \|_2^2 \) has the centered chi-2 distribution with \( d \) degrees of freedom, we derive that, in a certain sense to specify:

\[
M\{\| Y_1 \|, \| Y_2 \|, \ldots, \| Y_N \|\}(\tau) \approx \sigma_0^{r-s} Y_r(\tau/\sigma_0) / Y_s(\tau/\sigma_0),
\]

when \( N \) and the norms of the signals are large enough. In the equation above, \( Y_q \) is defined for any \( q \in [0, \infty) \) as the map of \([0, \infty)\) into \([0, \infty)\) that assigns to every \( x \in [0, \infty) \) the value

\[
Y_q(x) = \int_0^x t^{q+d-1} e^{-t^2/2} dt.
\]

When the signal norms are large, the threshold \( \tau \) must be large as well. It then turns out that the ratio in the right hand side of (7) tends to \( \lambda \) when \( \tau \) tends to \( \infty \). We then conclude that
M_{\|Y_1\|,\|Y_2\|,...,\|Y_N\|}(\tau) must be close to \lambda \sigma_0^{r-s}. The suitable convergence criterion, according to which we can really approximate M_{\|Y_1\|,\|Y_2\|,...,\|Y_N\|}(\tau) by \lambda \sigma_0^{r-s} when the signal norms and the number N of observations are large enough, cannot be trivial because (5) and (7) relate to different types of convergence. It is a surprising and unforeseen fact that the convergence of M_{\|Y_1\|,\|Y_2\|,...,\|Y_N\|}(\tau) to \lambda \sigma_0^{r-s} is satisfied even when the signals \Lambda_n are not iid, whatever their probability distributions, if we assume (A1), (A2) and (A3) and resort to the notions of minimum amplitude and thresholding function, which make it possible to choose a suitable threshold height \tau for censoring the data. To conclude this heuristic presentation, we can say that although the existence of some convergence could be somewhat guessed, the specific form of the convergence criterion (4) was not intuitive.

III. THE \textit{d}-DIMENSIONAL AMPLITUDE TRIMMED ESTIMATOR (DATE)

The theoretical result given above states that M_{\|Y_1\|,\|Y_2\|,...,\|Y_N\|}(\tau) tends to \lambda \sigma_0^{r-s}, according to a specific convergence, when the sample size N and the minimum SNR \varphi(\Lambda) tend to \infty. The question addressed in this section is then the design, guided by this theoretical result, of an estimator of the noise standard deviation \sigma_0. In fact, the convergence involved in (4) cannot be implemented as such and must be adapted in practice with some acceptable simplification. We thus proceed in three steps. First, Section III-A exposes some preliminary results to transform the theoretical convergence of Theorem 1 to a discrete cost suitable for practical situations. A general form for this discrete cost can readily be given at the beginning of Section III-A. However, this cost involves a minimization routine, which can be computationally heavy. Therefore, instead of studying this discrete cost in its general form, we proceed further by restricting our attention to a particular case where the minimization of this cost reduces to solving an equation, for which we can easily derive the existence and the value of a solution via lemma 1 below. The DATE derives from this lemma and is presented in Section III-B. The relevance of the DATE will be illustrated in Section IV.

A. Preliminary results for robust noise standard deviation estimation derived from Theorem 1

On the basis of Theorem 1, the same type of reasoning as in [8, Sec. 3.1] leads to estimate the noise standard deviation \sigma_0 by seeking a possibly local minimum of

$$
\arg\min_{\sigma \in [\sigma_{\min}, \sigma_{\max}]} \sup_{\ell \in \{1,\ldots,L\}} \left| M_{\|Y_1\|,\|Y_2\|,...,\|Y_N\|}(\beta_{\ell} \sigma \varphi(\Lambda)) - \lambda \sigma^{r-s} \right|, \tag{9}
$$

where \beta_{\ell} = \ell / L with \ell = 1, 2, \ldots, L and [\sigma_{\min}, \sigma_{\max}] is a suitable search interval. Such a discrete cost requires a minimization routine, which can be computationally heavy. In the present paper, we follow
a different path by considering the discrete cost \( (9) \) when \( L = 1 \) and deriving a specific solution that involves no minimization routine in this particular case.

When \( L = 1 \), if there exists a solution in \( \sigma \in [\sigma_{\min}, \sigma_{\max}] \) to the equation

\[
M_{\{Y_1, Y_2, \ldots, Y_N\}}(\sigma \theta(\rho(\Lambda))) = \lambda \sigma^{r-s}
\]

(10)

this solution basically minimizes the discrete cost \( (9) \). The existence and value of solutions to equations such as (10) are ruled by the following lemma 1. The construction of the DATE then derives from the result established by lemma 1.

**Lemma 1:** Given any \( N \in \mathbb{N} \) and any \( N \) non-negative real numbers \( x_1, x_2, \ldots, x_N \), let us define

\[
M^*_{\{x_1, x_2, \ldots, x_N\}}(n) = \begin{cases} \sum_{k=1}^{n} x'(k) / \sum_{k=1}^{n} x^s(k) & \text{if } n \neq 0 \\ 0 & \text{if } n = 0, \end{cases}
\]

(11)

for any \( n \in \{1, 2, \ldots, N\} \), where \( x(1) \leq x(2) \leq \ldots \leq x(N) \) are the values \( x_1, x_2, \ldots, x_N \) sorted by ascending order. Set \( x(0) = 0 \) and \( x(N+1) = \infty \).

(i) There exists a positive solution in \( \sigma \) to the equation

\[
M_{\{x_1, x_2, \ldots, x_N\}}(\sigma \tau) = K \sigma^{r-s},
\]

(12)

where \( K \) and \( \tau \) are any two positive real numbers, if and only if there exists some integer \( n \in \{1, 2, \ldots, N\} \) such that

\[
x(n) \leq \left( \frac{M^*_{\{x_1, x_2, \ldots, x_N\}}(n)}{K} \right)^{\frac{1}{r-s}} \tau < x(n+1);
\]

(13)

(ii) If the inequalities in (13) are satisfied by some \( n \in \{1, 2, \ldots, N\} \), then \( \left( M^*_{\{x_1, x_2, \ldots, x_N\}}(n)/K \right)^{\frac{1}{r-s}} \) is a solution \( \sigma \) to (12).

(iii) If \( K \leq \tau^{r-s} \), there exists \( n \in \{1, 2, \ldots, N\} \) such that (13) holds true.

**Proof:**

**Proof of statement (i):** Suppose the existence of \( \sigma \) such that (12) holds true. Let \( n_{\text{max}} \) be the largest integer \( n \in \{0, 1, \ldots, N\} \) such that \( x(n) \leq \sigma \tau < x(n+1) \) We then have \( M_{\{x_1, x_2, \ldots, x_N\}}(\sigma \tau) = M^*_{\{x_1, x_2, \ldots, x_N\}}(n_{\text{max}}) = K \sigma^{r-s} \). It follows that \( n_{\text{max}} \) is necessarily larger than or equal to 1 since \( K \sigma^{r-s} > 0 \) and that \( \sigma = \left( M^*_{\{x_1, x_2, \ldots, x_N\}}(n_{\text{max}})/K \right)^{\frac{1}{r-s}} \). Therefore, there exists some element \( n \in \{1, 2, \ldots, N\} \), actually equal to \( n_{\text{max}} \), such that (13) holds true.
Proof of statement (ii): Conversely, suppose the existence of some integer \( n \in \{1, 2, \ldots, N\} \) such that (13) holds true. Set \( \sigma = \left( M^*_{[x_1, x_2, \ldots, x_N]}(n)/K \right)^{1/r} \). We then have \( x(n) \leq \sigma \tau < x(n+1) \) so that \( M^*_{[x_1, x_2, \ldots, x_N]}(\sigma \tau) = K \sigma^{r-s} \).

Proof of statement (iii): A routine exercise shows that the sequence \( M^*_{[x_1, x_2, \ldots, x_N]}(n) \) is non decreasing with \( n \) because \( r > s \). For notational sake, set \( \mu_n = \left( M^*_{[x_1, x_2, \ldots, x_N]}(k)/K \right)^{1/r} \) for any \( n \in \{1, 2, \ldots, N\} \).

We argue by contradiction, assuming that \( K \leq \tau^{r-s} \) and no integer in \( \{1, 2, \ldots, N\} \) satisfies (13). Let \( n \) be some element of \( \{0, 2, \ldots, N-1\} \) such that \( x(n+1) \leq \mu_n \). We then have \( x(n+1) \leq \mu_{n+1} \) because of the non-decreasingness of \( \mu_n \) with \( n \). Since no integer in \( \{1, 2, \ldots, N\} \) is assumed to verify (13), it follows that \( x(n+2) \leq \mu_{n+1} \). According to a straightforward recurrence, we obtain that \( x(n+1) \leq \mu_N \), a contradiction. Now, let \( n \) be some element of \( \{0, 2, \ldots, N-1\} \) such that \( x(n) > \mu_n \). We thus have \( x(n) > \mu_{n-1} \) because \( \mu_n \) is non-decreasing with \( n \). Therefore, \( x(n-1) > \mu_{n-1} \), since it is assumed that no integer in \( \{1, 2, \ldots, N\} \) satisfies (13). A straightforward recurrence leads to \( x(1) > \mu_1 \), a contradiction with \( K \leq \tau^{r-s} \).

B. The DATE

Suppose we are given a sequence \( Y = (Y_k)_{k \in \mathbb{N}} \in \mathcal{M}(\Omega, \mathbb{R}^d)^{\mathbb{N}} \) that satisfies the assumptions of Theorem 1 and a non-negative real value \( \rho \) such that \( \rho(\Lambda) \geq \rho \). To estimate the noise standard deviation \( \sigma_0 \) by a solution in \( \sigma \) to (10), lemma 1 applies to determine whether such a solution exists and to calculate the value of this solution if it actually exists. Specifically, denoting by \( Y_1, Y_2, \ldots, Y_N \) the sequence of observations \( Y_1, Y_2, \ldots, Y_N \) sorted by increasing norm, it follows from lemma 1 that:

[Existence] if \( \lambda \leq \theta(\rho)^{r-s} \), there exists some integer \( n \in \{1, 2, \ldots, N\} \) such that
\[
\|Y(n)\| \leq \left( M^*_{[\|Y_1\|, \|Y_2\|, \ldots, \|Y_N\|]}(n)/\lambda \right)^{1/r} \theta(\rho) < \|Y(n+1)\|,
\]
which implies that a solution \( \sigma \) to (10) exists;

[Value] if (14) is satisfied by some \( n \in \{1, 2, \ldots, N\} \), a solution \( \sigma \) to (10) is then
\[
\sigma^*_{[Y_1, Y_2, \ldots, Y_N]} = \left( M^*_{[\|Y_1\|, \|Y_2\|, \ldots, \|Y_N\|]}(n)/\lambda \right)^{1/r}.
\]

[Existence] and [Value] are the cornerstones of the estimator proposed and studied below. To proceed further for practical usage, we must now choose a suitable thresholding function \( \theta \) and appropriate values for \( r \) and \( s \).

In many signal processing applications, signals of interest are less often present than absent, so that we can reasonably restrict our attention to the case where (A2) is satisfied with \( p = 1/2 \). It is worth
noticing that, when the amplitudes of the signals are large enough, we can say that the sequence of observations \((Y_n)_{n \in \mathbb{N}}\) is weakly sparse in the sense that most of the random vectors \(Y_k\) are due to noise alone and at most half of them contain signals. The choice for the thresholding function then derives from the assumption that \(p = 1/2\). Specifically, for every given \(\rho \in [0, \infty)\), let \(\xi(\rho)\) be the unique positive solution for \(x\) in the equation \(\phi \Gamma \left( \frac{d+1}{2} \right) / \Gamma \left( \frac{d}{2} \right) \leq \xi(\rho)\), where \(\phi \Gamma\) is the generalized hypergeometric function [9, p. 275]. This map \(\xi\) is a thresholding function with \(C = 1/2\) ([8], [10]). Under assumption (A2) with \(p = 1/2\), that is, when signals are less often present than absent, \(\xi\) is then particularly relevant because it follows from [10, Theorem VII-1] that the thresholding test with threshold height \(\sigma_0 \xi(\rho)\) is capable of making a decision on the value of any \(\varepsilon_k\) with a probability of error that decreases rapidly with \(\rho\). By thresholding test with threshold height \(\sigma_0 \xi(\rho)\), we mean the test that returns 1 when the norm of any given observation exceeds \(\sigma_0 \xi(\rho)\) and 0 otherwise. The use of the thresholding function \(\xi\) thus optimizes the approximation of (6) and, therefore, favors the convergence in (4).

We fix \(s = 0\) and \(r = 1\) and, thus, \(v \geq 2\) in (A3), because this choice guarantees the existence of some \(n \in \{1, 2, \ldots, N\}\) satisfying (14) for any \(\rho \in [0, \infty)\). This can be seen as follows. When \(s = 0\) and \(r = 1\), the required condition in [Existence] is that \(\sqrt{2\Gamma \left( \frac{d+1}{2} \right) / \Gamma \left( \frac{d}{2} \right)} \leq \xi(\rho)\). This inequality holds true for any \(d\) and any \(\rho\) because \(\sqrt{2\Gamma \left( \frac{d+1}{2} \right) / \Gamma \left( \frac{d}{2} \right)} \leq \sqrt{d} [11, p. 403]\) and \(\sqrt{d} \leq \xi(\rho) [10, p. 235]\). Our choice for \(s\) and \(r\) is compatible with the very important practical case of signals with finite energy, for which \(v = 2\) in (A3). Fixing \(s = 0\) and \(r = 1\) is also computationally beneficial since the power \(1/(r - s)\) is not used any longer in (10) and lemma 1.

Although there exists \(n \in \{1, 2, \ldots, n_{\min}\}\) such that (14) holds true, the fact that the signal probabilities of presence are less than or equal to one half \((p = 1/2)\) can be taken into account to avoid choosing a too small \(n\) leading to an overestimate of the noise standard deviation via [Value]. Specifically, let us consider \(N\) observations \(Y_1, Y_2, \ldots, Y_N\) with \(p = 1/2\) and some positive real value \(Q\) less than or equal \(1 - \frac{N}{4N/2 - 1}\). Set

\[ h = 1/\sqrt{4N(1-Q)}. \]  

According to Bienaymé-Chebyshev’s inequality, and since the probabilities of presence of the signals are assumed to be less than or equal to one half, the probability that the number of observations due to noise alone be above

\[ n_{\min} = N/2 - hN, \]  

is larger than or equal to \(Q\). The experimental results presented in the sequel were obtained by
choosing \( Q = 0.95 \). The DATE is then specified as follows, on the basis of the foregoing.

**The DATE:**

**Input:**
- A finite subsequence sequence \( \{Y_1, Y_2, \ldots, Y_N\} \) of a sequence \( Y = (Y_k)_{n \in \mathbb{N}} \) of independent \( d \)-dimensional real random vectors satisfying assumptions (A1-A3) of Theorem 1 with \( \nu \geq 2 \) and \( p = 0.5 \).
- A lower bound \( \rho \) for the minimum SNR \( \rho(\Lambda) \)
- The probability value \( Q \leq 1 - \frac{N}{4(N/2-1)^2} \)

**Output:** The estimate \( \sigma^*_{\{Y_1, Y_2, \ldots, Y_N\}} \) of the noise standard deviation.

**Computation of \( \sigma^*_{\{Y_1, Y_2, \ldots, Y_N\}} \):**

1) **[Search interval]:** Compute \( n_{\text{min}} \) according to (17) and (16)

2) **[Existence]:**

   IF there exists a smallest integer \( n \) in \( \{n_{\text{min}}, \ldots, N\} \) such that
   \[
   \|Y(n)\| \leq \left(M^*_{\{\|Y_1\|, \|Y_2\|, \ldots, \|Y_N\|\}}(n)/\lambda\right)\xi(\rho) < \|Y(n+1)\|
   \]
   with
   \[
   M^*_{\{\|Y_1\|, \|Y_2\|, \ldots, \|Y_N\|\}}(n) = \begin{cases} \frac{1}{n} \sum_{k=1}^{n} \|Y(k)\| & \text{if } n \neq 0 \\ 0 & \text{if } n = 0 \end{cases}
   \]
   set \( n^* = n \).

   ELSE, set \( n^* = n_{\text{min}} \).

3) **[Value]:**
   \[
   \sigma^*_{\{Y_1, Y_2, \ldots, Y_N\}} = M^*_{\{\|Y_1\|, \|Y_2\|, \ldots, \|Y_N\|\}}(n^*)/\lambda
   \]

**Remark 1:** The DATE belongs to the family of trimmed estimators. However, it is important to emphasize that there are two fundamental differences between the DATE and the conventional trimmed estimator T depicted in Appendix A-C. The first one is that the DATE applies to \( d \)-dimensional vectors, whereas the T-estimator is limited to \( d = 1 \). The second difference is that the conventional T-estimator requires fixing the amount of outliers to remove whereas the DATE trims the amplitudes or norms of the observations by assuming that the signal norms are above some known lower-bound and that the signal probabilities of occurrence are less than one half.
Remark 2: In case of multiple solutions of (18), the DATE picks the smallest one. This choice obeys the following rationale. As enhanced in Section III-C, the DATE can be regarded as an outlier detector. Because of the theoretical results on which the DATE relies, the observations \( Y_{(m)} \) with \( m \geq n \) where \( n \) is a solution to (18) are considered as pertaining to some signal. It follows that a too large value for \( n \) might induce the detection of too few outliers. Seeing that our purpose is to design an estimator capable of performing better than the standard robust scale estimators that may fail because of signal probabilities of presence less than but close to \( 1/2 \), the detection performed by the DATE must not be too severe. We thus choose the smallest \( n^* \) satisfying (18) to avoid a detrimental overestimation of the noise standard deviation. Note that the number of solutions to (18) decreases by increasing the value of \( n_{\min} \) and, thus, diminishing the probability value \( Q \).

Remark 3: The one and the two dimensional cases. These cases are of most importance in many signal processing applications. For instance, one-dimensional observations can be time samples or coefficients returned by some transform such as the wavelet transform or the Discrete Fourier Transform. Two-dimensional random vectors, or, equivalently, complex random variables can be the complex values provided by the standard \( I \) and \( Q \) decomposition of standard receivers in radar, sonar and telecommunication systems. They can also result, as in Section IV dedicated to spectrum sensing, from the Discrete Fourier Transform (DFT) of the received signal. It is then worth recalling that, for one- and two-dimensional observations, the expression of \( \xi \) simplifies. In fact, for any \( \rho \in [0, \infty) \), it follows from [10, Remark V.3 and Section 2] that \( \xi(\rho) = \cosh^{-1}(\rho^{2/2}) = \frac{1}{2}\rho + \frac{1}{\rho}\log(1 + \sqrt{1 - e^{-\rho^2}}) \) if \( d = 1 \) and that \( \xi(\rho) = I_0^{-1}(\rho^{2/2})/\rho \) if \( d = 2 \), where \( I_0 \) is the zeroth order modified Bessel function of the first kind. Note that \( \lambda = 0.7979 \) for \( d = 1 \) and \( \lambda = 1.2533 \) for \( d = 2 \).

Remark 4: On the choice for the SNR lower bound \( \rho \) in applications. If the signals have a minimum signal to noise ratio \( \rho(\Lambda) \) above or equal to \( \rho \), probabilities of presence less than or equal to \( 1/2 \) and \( \rho \) and \( N \) are both large enough, the DATE should return an accurate enough estimate \( \sigma_{(Y_1,Y_2,...,Y_N)}^* \) of the noise standard deviation \( \sigma_0 \). However, in practice, how can we choose \( \rho \), insofar as signals may have amplitudes that are not bounded away from zero? The experimental results presented in [12] for synthetic signals as well as Fourier and wavelet transforms of real speech signals and images pinpointed the following. Under the experimental conditions of [12], the empirical bias and normalized mean square error (NMSE) of the DATE when \( \rho \) varies was minimized for \( \rho \) around 4, even in case of signals with signal to noise ratio below this value. In other words, the performance of the DATE adjusted with \( \rho = 4 \) and, thus, for signals with a minimum SNR above 4, was not significantly altered by the presence of signals with smaller SNR. This is rather natural
because small signal to noise ratios can be expected not to really affect the estimation. This is partly justified by [8, Theorem 1], where statement (ii) addresses the case of small signal amplitudes. With regard to the foregoing, we thus recommend using a value around $\rho = 4$ for practical problems in the one- and two-dimensional cases. In our application to spectrum sensing addressed in Section IV, we proceeded exactly this way by fixing $\rho = 4$, once for all, without trying to optimize it in this specific context. For many applications, this value can certainly be adjusted on the basis of a representative and sufficiently large database of synthetic or, even better, real observations.

C. Discussion on the robustness of the DATE

The DATE is more specific and constrained than standard robust scale estimators since it is designed with respect to the noise centred Gaussian distribution and for “outliers” that are, in fact, noisy signals. However, from the signal processing point of view, it is robust in the sense that, in contrast to standard robust estimators such as the MAD or the T-estimators, it can cope with signals that have probabilities of occurrence close to 1/2 from below. It is then natural to get some insight into the robustness of the DATE in the “neighborhood” of our predefined white and Gaussian model. In this respect, among the several criteria proposed by the theory of robust estimation to measure the robustness of an estimator around its target distribution, the influence function (IF) and the breakdown point are very popular. In this section, we then discuss to what extent these criteria contribute to better understand the behavior of the DATE.

The IF reflects the bias of any estimator $\hat{\Theta}$ at the underlying distribution $F$ caused by the addition of a few outliers at point $\kappa$, standardized by the amount $\eta$ of contamination, i.e.

$$\text{IF}_{\hat{\Theta}}(\kappa, F) = \lim_{\eta \to 0} \frac{\hat{\Theta}((1 - \eta)F + \eta \delta_{\kappa}) - \hat{\Theta}(F)}{\eta}$$

(21)

where $\delta_{\kappa}$ is the point-mass at $\kappa$. From the IF, several robustness measures can be defined such as

- the gross-error sensitivity: $\sup_{\kappa} \{||\text{IF}_{\hat{\Theta}}(\kappa, F)||\}$,
- the rejection point: $\inf_{\zeta > 0} \{\text{IF}_{\hat{\Theta}}(\kappa, F) = 0, |\kappa| > \zeta\}$.

To get quantitative results on these measures, the closed-form expression of the IF is required. However, the DATE belonging to the family of L-estimators, its IF depends on some quantile [4, p. 56] that is here indirectly determined by variable $n^*$ calculated by solving (18). Since $n^*$ depends on the observation vector, it is therefore a random variable whose distribution has yet to be determined. Consequently, the IF expression cannot be derived. Although a quantitative analysis cannot be provided, qualitative conclusions can be drawn from the so-called sensitivity curve. The sensitivity
curve can be seen as a finite sample version of the IF [14, p. 43]. It measures the effect of different locations of an outlier $\kappa$ on any estimate $\hat{\Theta}$ based on the sample $Y = Y_1, \cdots, Y_N$ and is defined as $SC_m(\kappa) = (m+1)(\hat{\Theta}(Y_1, \cdots, Y_N, \kappa) - \hat{\Theta}(Y_1, \cdots, Y_N))$. Figure 1 shows the sensitivity curve of the DATE as well as the IF of the MAD and the trimmed estimator, when the model distribution is Gaussian with $d = 1$. From this figure, it can first be conjectured that the gross-error sensitivity of the DATE is finite and relatively small, which is expected from a robust estimator. The gross-error sensitivity of the MAD is equal to 1.167, which is the smallest value that can be obtained for any scale estimator in the case of the normal distribution. Moreover, the rejection point of the DATE appears to be finite ($\approx 2$), in contrast to the MAD whose rejection point is infinite. Although the rejection point helps to get some insight into the behavior of an estimator, a finite rejection point is not required for robust estimators. Estimators with a finite rejection point are said to be redescending and are well protected against very large outliers. However, a finite rejection point can affect the efficiency of the estimator since samples near the tail of a distribution may be ignored. The T-estimator also has a finite rejection point that corresponds to the $(1 - \mu)$-quantile of the normal distribution where $\mu$ is the amount of trimmed data (see Appendix A).

This finite rejection point also indicates that the DATE can be used to detect outliers as defined in our context, that is, signals in AWGN. We can give some further details on this point. If there is an integer $n^*$ such that (18) holds true, the estimate of the noise standard deviation is given by (20) and it follows that, $\|Y_\ell\| \geq \sigma_{Y_1, Y_2, \cdots, Y_N}^* \xi(\rho), \ell \in \{n^* + 1, \ldots, N\}$. On the one hand, the signals are assumed to have probabilities of presence less than or equal to 1/2 and norms larger than or equal to $\rho$. On the other hand, $\sigma_{Y_1, Y_2, \cdots, Y_N}^* \xi(\rho)$ approximates $\sigma_0 \xi(\rho)$ reasonably well. Therefore, the observations $Y_\ell$ with $\ell \in \{n^* + 1, \ldots, N\}$ are those whose norms exceed a rather good estimate of $\sigma_0 \xi(\rho)$. The threshold height $\sigma_0 \xi(\rho)$ makes it possible [10, Theorem VII-1] to detect our signals with a probability of error that has a sharp upper bound, which is also a sharp upper bound for the probability of error of the minimum probability of error (MPE) test (see [15, Sec. II. B], among others, for the definition of the MPE test). This upper bound is sharp because it is attained for any with probability of presence equal to 1/2 and uniform distribution on $\rho S^{d-1}$. Therefore, the DATE can be regarded, and possibly used, as a detector of the signals that are either present or absent in noise. This detector thus decides that $\epsilon_\ell$ equals 1 — and, thus, that $Y_\ell$ pertains to some signal — for every $\ell \geq n^* + 1$ and decides that $\epsilon_\ell$ equals 0 — and, thus, that $Y_\ell$ is noise alone — for every $\ell \leq n^*$.

Another measure of robustness is the breakdown point of an estimator. Several definitions for this measure are available in the literature [16]. In contrast to most scale estimators such as the MAD, the
breakdown point of the DATE depends on the actual realization of the observation vector $Y$ so that it cannot easily be derived in the general case. This is once again due to the random nature of $n^*$ determined by (18). Further theoretical studies have yet to be performed to address the derivation of the breakdown point.

IV. APPLICATION TO SPECTRUM SENSING

In a cognitive radio context or for CES applications, spectrum sensing aims at monitoring the radio-frequency bands of interest to detect either communication systems or spectrum holes (i.e. idle frequency channels). In both cases, most of the detection algorithms rely on prior knowledge of the noise standard deviation. In these applications, detection usually results from the processing of a complex-valued observation represented in the time-frequency domain (i.e. at the output of a short-time Fourier transform), that is $Y_{k,n} = \varepsilon_{k,n}\Lambda_{k,n} + X_{k,n}$, where $k$ is the time frame index and $n$ the DFT bin number, $X_{k,n}$ is complex white Gaussian noise, $X_{k,n} \sim \mathcal{CN}(0,2\sigma_0^2)$ and $\Lambda_{k,n}$ is the received signal. This signal is here modeled for the numerical simulations by $\Lambda_{k,n} = \sqrt{E_s}a_{k,n}H_{k,n}$, where $E_s$ is the signal power, the $a_{k,n}$ represents the transmitted data symbol and $H_{k,n}$ is the propagation channel. For the sake of generality, the $a_{k,n}$’s are assumed to be iid, zero-mean and uniformly distributed with $\mathbb{E}[|a_{k,n}|^2] = 1$; $H_{k,n}$ is an iid Rayleigh fading channel in the frequency domain and a Gauss-Markov process in the time domain with $\mathbb{E}[H_{k,n}H_{k-1,n}^*] = 0.9$. Note that because of channel memory in the time domain, the iid assumption on the $Y_{k,n}$ required by the theory which the DATE is based on, is not fulfilled here. Despite this slight deviation from the ideal observation model, it is shown hereafter that the DATE performs well. The DATE performance is measured as a function of the probability $P[\varepsilon_{k,n} = 1]$ as well as a function of the actual average signal-to-noise ratio (ASNR) defined as ASNR $= 10 \log_{10}(E_s/(2\sigma_0^2))$. The ASNR represents the simulated average power ratio between $\Lambda_{k,n}$ and $X_{k,n}$, which has to be differentiated from the minimum SNR $\rho$, the input parameter of the DATE. In the numerical analyses, $\rho$ has been set to 4 in accordance to Remark 4.

Figure 2 compares the performance of the DATE with a set of conventional L-estimators detailed in Appendix A.2 These results are obtained in an “average scenario” where ASNR equals 10 dB and the observation is limited to 128 DFT bins and 16 time frames. Despite the theoretical ground of the DATE that bounds the probability of the signal presence to 1/2, the NMSE of the estimated $\sigma_0$ is

2Note that this estimators are originally designed for real-valued random variables. However, since the real and imaginary parts of $Y_{k,n}$ are independent in our scenario, the noise standard deviation of $X_{k,n}$ is estimated on the observation vector made of the concatenation of the 2 dimensions of the $Y_{k,n}$’s.
here plotted for $0 \leq P[\epsilon_{k,n} = 1] \leq 1$. This is motivated by the application context of spectrum sensing algorithms that can face situations where $P[\epsilon_{k,n} = 1] > 1/2$. As shown in [17], the occupancy rate of the radio-frequency spectrum varies from one band to another. CES systems mainly focus on military bands such as the 30-88 MHz, 225-400 MHz or 960-1240 MHz bands where the occupancy rate can be less than 10% in peace zones to more than 50% on theatres of operation. Cognitive radio systems mainly focus on TV and ISM bands where the occupancy rate is usually high (from 15% to 70% [17]).

Figure 2 clearly shows that the DATE largely outperforms the trimmed and the winsorized estimators, whatever the value of $\mu$ might be. It also outperforms the MAD and its alternatives in the range of probability of presence relevant for spectrum sensing applications. This very good robustness is paid back by a poor efficiency (there is a 17 dB NMSE loss in comparison to $Q_n$ for $P[\epsilon_{k,n} = 1] = 0$).

As discussed in Section III-C, the poor efficiency of the DATE is consistent with the finite rejection point of the sensitivity curve observed in Figure 1. It is also worth noticing the seemingly existence of an optimum for the NMSE of the DATE. This was not expected theoretically and deserves some attention in forthcoming work.

As detection is usually the most critical operation in spectrum sensing, the proposed estimator is indirectly evaluated in Figure 3 through the performance of a classical constant false alarm rate (CFAR) detector. This figure compares the true detection rate ($P_{det}$) for various theoretical false alarm rates ($P_{fa}$), when the noise standard deviation is either perfectly known or when it is estimated using the DATE or the MAD. The DATE is here compared with the MAD since none of these two estimators require prior knowledge of the outlier number, in contrast to trimmed or winsorized estimators. The decision on detection is made by comparing $|Y_{k,n}|^2$ to a positive threshold aimed at guaranteeing a specified false alarm rate. Given that noise is complex-valued and Gaussian, $|Y_{k,n}|^2/\sigma_0^2$ follows a chi-square distribution with 2 degrees of freedom when $\epsilon_{k,n} = 0$. Therefore, when $\sigma_0$ is known, the detector decides that $\epsilon_{k,n} = 1$ if $|Y_{k,n}|^2 > -2\sigma_0^2\ln(P_{fa})$ and that $\epsilon_{k,n} = 0$, otherwise. It is usual to summarize this decision-making on the value of $\epsilon_{k,n}$ by writing $|Y_{k,n}|^2 \geq_{\epsilon_{k,n} = 0}^{\epsilon_{k,n} = 1} -2\sigma_0^2\ln(P_{fa})$.

When $\sigma_0$ is estimated by the DATE or the MAD, we replace $\sigma_0$ by its estimate. Figure 3 confirms the benefit of the DATE for spectrum sensing applications. In the example given, the MAD over-estimates $\sigma_0$ so that the true detection rate is far below the one obtained with a perfect knowledge $\sigma_0$. This has to be compared with the detection rate of the DATE, which is very similar to the ideal one.

3 CES or cognitive radio systems may use detector structures other than CFAR ones. This kind of detector structure is one of the most popular and is therefore used for illustration purposes.
V. CONCLUSION AND PERSPECTIVES

The DATE is a new estimator of the noise standard deviation for signal processing applications where $d$-dimensional observations involving signals with unknown distributions must be processed so as to extract knowledge about these same signals. The theoretical derivation of this estimator is made under weak sparsity hypotheses that bound our lack of prior knowledge about the signal distributions and probabilities of presence. In contrast to most robust scale estimators, the DATE applies whatever the value of $d$ and has been shown to perform very well in spectrum sensing, a rather natural application. In [12], an application to denoising by wavelet shrinkage is also presented. In this application as well, the DATE outperforms the MAD. However, in denoising by wavelet shrinkage, the improvement yielded by the DATE is not so noticeable as in spectrum sensing because wavelet transforms of standard images are rather sparse in that most of the signal wavelet coefficients are small [18], which does not hinder the DATE but is beneficial to the MAD. The MATLAB code for this application is available at http://perso.telecom-bretagne.eu/pastor, where the interested reader will also find an application to speech signals in AWGN. In this case, the short-time Fourier transform is used to obtain a time-frequency representation where the noisy speech signals are less often present than absent. There are however too many large coefficients pertaining to the speech signal for the MAD to provide a good estimate of the noise standard deviation and the DATE significantly outperforms the MAD. Radio-astronomy can also be a field of application of the DATE, since cosmic radio emissions (usually modeled as Gaussian) are perturbed by man-made radio-frequency interferences of unknown distribution [6]. In this case, the DATE could make it possible to estimate the standard deviation of the signals of interest.

The DATE raises several questions and open new prospects in robust statistical signal processing. A theoretical analysis of the bias, MSE, efficiency and consistency of the DATE should be undertaken. In order to get better insight into the behavior of the DATE, it could also be relevant to take into account that the noise standard deviation is also the solution of a limit equation when signals have very small amplitude and priors upper-bounded by some value in $[0, 1)$ (see [8, Theorem 1, statement (ii)]). Such a result could be exploited to justify the good performance of the DATE, even when the signal amplitudes or the signal SNRs are not big. The Gaussian assumption being an approximation, forthcoming work could address the robustness of the DATE to deviation from normality. For signal processing applications, statistical properties of wavelet transforms such as those summarized in [19] could be very helpful in combination with the DATE to deal with non white and/or non Gaussian
noise. Finally, as already enhanced in Section III-C, the DATE can be used to detect signals. However, this capability of the DATE has not yet been exploited, even in Section IV where the detection is performed via a very standard approach. The link between the problem of detecting signals in AWGN and the scale estimation performed by the DATE deserves further attention.

VI. ACKNOWLEDGEMENT

The authors are very grateful to the reviewers and the editor for the several comments that contributed to improve the paper. In particular, the authors are very thankful to the anonymous reviewer, whose in-depth knowledge in robust statistics helped in better introducing and discussing the DATE in comparison with standard robust estimators

APPENDIX A

PRESENTATION OF SOME CONVENTIONAL ROBUST ESTIMATORS

In this appendix, $Y_1, \cdots, Y_N$ are $N$ random variables representing the observations, $b$ is a constant needed to make the estimators Fisher-consistent at the model distribution, $\text{med}_i Y_i$ stands for the median value of the observation and $Y_{(1)} \leq Y_{(2)} \leq \cdots \leq Y_{(N)}$ are the order statistics.

A. MAD

A widely used robust scale estimator is the median absolute deviation about the median (MAD) defined in [13], [7] by $\text{MAD} = b \times \text{med}_i |Y_i - \text{med}_j Y_j|$. To make the MAD consistent at the normal distribution $b \approx 1.4826$.

B. $S_n$ and $Q_n$ estimates

As alternatives to the MAD, Rousseeuw and Croux proposed in [7] two estimators that offer a better efficiency than the MAD, while guaranteeing a 50% breakdown point. These estimators are $S_n = b \times \text{med}_i (\text{med}_j |Y_i - Y_j|)$ with $b \approx 1.1926$ for the normal distribution and $Q_n = b \times \{ |Y_i - Y_j|, i \leq j \}_{(k)}$, where $k \approx \left( \frac{N}{2} \right) / 4$ and $b \approx 2.2219$ to make $Q_n$ consistent at the normal distribution.

C. Trimmed estimate (T-estimate)

Let $\mu$ denote the chosen amount of trimming with $0 \leq \mu \leq 1/2$ and $k = \lfloor \mu N \rfloor$ where $\lfloor x \rfloor$ denotes the largest integer less than or equal to $x \in \mathbb{R}$. The sample trimmed standard deviation is computed by removing the $k$ largest and the $k$ smallest data and using the remaining values [4, Chap. 5]:

$$T = \sqrt{\frac{b}{N-2k} \sum_{j=k}^{N-k} \left( Y_{(j)} - \frac{1}{N-2k} \sum_{i=k}^{N-k} Y_{(i)} \right)^2}.$$
The constant $b$ depends on the model distribution as well as on $\mu$. For a Gaussian model, Table I gives some values of $b$ for different $\mu$. The breakdown point of the $\mu$-trimmed estimator is $\mu$.

**Table I**

<table>
<thead>
<tr>
<th>$\mu$</th>
<th>0.05</th>
<th>0.1</th>
<th>0.2</th>
<th>0.4</th>
</tr>
</thead>
<tbody>
<tr>
<td>$b$</td>
<td>1.6</td>
<td>2.28</td>
<td>4.66</td>
<td>47.14</td>
</tr>
</tbody>
</table>

**D. Winsorized estimate**

For the chosen $0 \leq \mu \leq 1/2$ and $k = \lfloor \mu N \rfloor$, winsorization$^4$ of the sorted data consists in setting

$$w_j = \begin{cases} 
Y_{(k+1)}, & \text{if } Y_{(j)} \leq Y_{(k+1)} \\
Y_{(j)}, & \text{if } Y_{(k+1)} < Y_{(j)} < Y_{(N-k)} \\
Y_{(N-k)}, & \text{otherwise.}
\end{cases}$$

The winsorized standard deviation estimate is then computed as follows:

$$W = \sqrt{\frac{b}{N-1} \sum_{j=1}^{N} \left( w_j - \frac{1}{N} \sum_{i=1}^{N} w_i \right)^2}.$$ 

For a Gaussian model, Table II gives some values of $b$ for different $\mu$.

**Table II**

<table>
<thead>
<tr>
<th>$\mu$</th>
<th>0.05</th>
<th>0.1</th>
<th>0.2</th>
<th>0.4</th>
</tr>
</thead>
<tbody>
<tr>
<td>$b$</td>
<td>1.2</td>
<td>1.47</td>
<td>2.43</td>
<td>17.99</td>
</tr>
</tbody>
</table>

**APPENDIX B**

**PROOF OF THEOREM 1**

It follows from [8, Theorem 1] that $\sigma_0$ is such that, for every $\beta_0 \in (0, 1]$,

$$\lim_{\varphi(\Lambda) \to \infty} \limsup_{N \to \infty} M_{\{||Y_1||, \ldots, ||Y_N||\}}(\beta \sigma_0 \theta(\varphi(\Lambda))) - \sigma_0^{-s} \frac{Y_r(\beta \theta(\varphi(\Lambda)))}{Y_s(\beta \theta(\varphi(\Lambda)))} = 0 \quad (22)$$

$^4$As noted by Dixon [20], this procedure has been called winsorization in honour of Charles P. Winsor.
uniformly in \( \beta \in [\beta_0, 1] \), where \( Y_q \) is defined for any \( q \in [0, \infty) \) by (8). On the other hand, it follows from [21, 3.326 (2), p. 337] that, given any \( q \in [0, \infty) \), \( \lim_{x \to \infty} Y_q(x) = 2^{q+d} \Gamma\left(\frac{q+d}{2}\right) \), so that \( \lim_{\rho(\Lambda) \to \infty} Y_{\rho(\rho(\Lambda))}/Y_{\rho(\rho(\Lambda))} = \lambda \). From this equality and (22), \( \sigma_0 \) satisfies (4).

It remains to prove that \( \sigma_0 \) is actually the unique positive real satisfying (4). This is achieved by mimicking the proof in [8, Appendix A.5] with some slight simplification. Assume the existence of two positive real numbers \( \sigma_1 \geq \sigma_2 > 0 \) that both satisfy (4). Let \( \beta_1 \) and \( \beta_2 \) two elements of (0,1) such that \( \beta_1 \sigma_2 = \beta_2 \sigma_1 \). We have

\[
\lambda |\sigma_1^{r-s} - \sigma_2^{r-s}| \leq \frac{1}{N} \left( \sum_{n=1}^{N} \| Y_n \| Y_{[0,\beta_2\sigma_1\theta(\rho(\Lambda))]}(\| Y_n(\omega) \|) \right) - \lambda \sigma_1^{r-s} + \frac{1}{N} \left( \sum_{n=1}^{N} \| Y_n \| Y_{[0,\beta_1\sigma_2\theta(\rho(\Lambda))]}(\| Y_n(\omega) \|) \right) - \lambda \sigma_2^{r-s}
\]

for any given pair \( (N, \omega) \in \mathbb{N} \times \Omega \). From this inequality, we derive that

\[
\lambda |\sigma_1^{r-s} - \sigma_2^{r-s}| \leq \left\| \limsup_{N \to \infty} \mathbb{M}(Y_1 || Y_2 || \ldots || Y_N) (\sigma_1 \beta_2 \theta(\rho(\Lambda))) - \lambda \sigma_1^{r-s} \right\|_\infty + \left\| \limsup_{N \to \infty} \mathbb{M}(Y_1 || Y_2 || \ldots || Y_N) (\sigma_2 \beta_1 \theta(\rho(\Lambda))) - \lambda \sigma_2^{r-s} \right\|_\infty.
\]

Since \( \sigma_1 \) and \( \sigma_2 \) are both assumed to satisfy (4), then, by choosing \( \rho(\Lambda) \) large enough, it follows from (2) that the right hand side in the inequality above can be rendered arbitrarily small. Therefore, we conclude that \( \sigma_1 = \sigma_2 \).

REFERENCES


**Dominique Pastor** was born in Cahors, France, 1963. He graduated from Telecom Bretagne (Brest, France) in 1986 and from the University of Rennes (France) in 1997 (Ph. D.). From 1987 until 2000, he was with Thales. In particular, between 1990 and 1998, he was with Thales Avionics where his research concerned speech processing for application to speech recognition systems in military fast jet cockpits and, from 1998 to 2000, he was with Thales Nederland where he worked on the detection of radar targets in sea clutter. In September 2000, he joined Altran Technologies Nederland as a senior consultant. Since September 2002, he is with Institut Telecom, where he is currently Professor at Telecom Bretagne. His current research interests concern alternatives to likelihood theory, for statistical processing of random signals with unknown distributions.

January 10, 2012
Francois-Xavier Socheleau (S’08, M’12) received the Eng. degree from the Ecole Supérieure d’Électronique de l’Ouest (ESEO), Angers, France, in 2001 and the Ph.D. degree from Telecom Bretagne, Brest, France, in 2011. From 2001 to 2004, he was a Research Engineer at Thales Communications, France, where he worked on the interception of noncooperative communications. From 2005 to 2007, he was employed at Navman Wireless (New Zealand/U.K.) as an R&D Engineer. Since October 2011, he is an Assistant Professor at ENSTA Bretagne. His general research interests lie in the field of signal processing and wireless communications with an emphasis on blind and robust estimation, channel characterization and underwater acoustics.

Fig. 1. Sensitivity curve of the DATE ($m=10000$, $\rho=4$) and influence function of the MAD and the 10% trimmed estimator with a one-dimensional Gaussian model.
Fig. 2. Normalized mean square error comparison of various robust noise standard deviation estimators with ASNR=10 dB. (a) Comparison with trimmed estimators, (b) comparison with winsorized estimators and (c) comparison with the MAD and its alternatives.
Fig. 3. Comparison of the receiver operating characteristics when the noise variance is perfectly known and when it is estimated via the DATE and the MAD. The priors $P[\epsilon_k n = 1]$ all equal 1/2.