Maximum-likelihood Detection of Impulsive Noise Support for Channel Parameter Estimation

Xavier Mestre, Miquel Payaró
Centre Tecnològic de Telecomunicacions de Catalunya (CTTC/CERCA)
Castelldefels (Barcelona), Spain
xavier.mestre@cttc.cat, miquel.payaro@cttc.es

Deep Shrestha
Signal Theory and Communications Department
Universitat Politècnica de Catalunya (UPC)
Barcelona, Spain
deep.shrestha@upc.edu

***Abstract*—In this paper, we consider the problem of estimating channel parameters in the presence of impulsive noise (IN). To that end, two novel maximum-likelihood based IN support detection techniques are proposed for the cases where the IN is modeled to be a deterministic quantity or a random one. For the deterministic case, an exact closed-form expression for the distribution of the joint likelihood statistic is provided whereas, in the random case, an exact expression of its *asymptotic* distribution is derived. In both cases, the computed distribution of the likelihood statistic enables the joint estimation of the channel parameters and the detection of the IN support with guarantees on the false alarm probability for the samples that are estimated to be in the IN support set. The goodness of the proposed expressions is validated via numerical simulations.**

***Index Terms*—Impulsive noise, support detection, maximum-likelihood estimation and detection.**

I. Introduction

In the time domain, impulsive noise (IN) is characterized by its random occurrence and by having a very brief duration and, potentially, high power [1]. IN is present in many systems (for example audio appliances or power line communications) and it can significantly hinder their performance, unless its effects are mitigated via, e.g., signal processing techniques. Focusing on discrete-time (sampled) systems, one of the main problems when designing digital signal processing schemes to overcome the effects of IN is to accurately identify which signal samples are obliterated by it.

One of the most straightforward approaches to identify IN-corrupted samples is to define a threshold and assume that all signal samples whose power is above the threshold are contaminated by IN. The case where the threshold is fixed is studied in [2]–[4] and more refined schemes that adapt their reference threshold value according to the change in the IN power are proposed in [5]–[7]. More recently, IN support estimation algorithms that exploit the time domain sparsity characteristic of IN based on compressed sensing (CS) and sparse Bayesian learning methods are proposed in [8]–[10]. Based on this idea, a further enhancement in IN support estimation is proposed in [11], where CS-based support estimation is used first as a coarse estimation, which is later refined in a second round by exploiting a priori information on the IN samples distribution. To enhance the precision in IN support detection that can be achieved by CS-based schemes, a basis pursuit (BP) algorithm is proposed in [12], [13]. Some other algorithms, which require a priori information on the sparsity level of the signal and are based on subspace pursuit and compressive sampling matching pursuit (CoSaMP), are proposed in [14], [15]. In situations where it is not possible to have this information before hand (blind signal recovery), a greedy algorithm called sparsity adaptive matching pursuit (SAMP) has been proposed in [16]. The performance of SAMP is further enhanced by the priori-aided SAMP (PA-SAMP) algorithm proposed in [17]. Moreover, the recent work in [18] proposes an improved version of CS-based schemes and greedy algorithms based on OMP and CoSaMP for accurate IN support detection.

In this paper, we capitalize on the work presented in [19] where, under the assumption of known IN support, two techniques for channel parameter estimation (impulse response, background noise power and IN power) were provided for the cases where the IN is modeled as a deterministic or a random quantity. We enhance the work in [19] by deriving the distribution of the two maximum-likelihood (ML) functions, ζ_{DML} for the deterministic case and ζ_{RML} for the random one. These two functions will then be used (i) to detect the IN support with guarantees on the false alarm probability for the samples in the IN support set and (ii) to estimate the channel parameters, which depend on the detected IN support.

The remainder of the paper is organized as follows. In section II, we present the system model under study. In section III, we derive the ML strategy for IN support detection under the deterministic setting. In section IV, we provide an asymptotic ML detection approach for the IN support under the random setting. In section V, we present the numerical validation of the derived results and conclude the paper.

II. System model

Let us consider N uses of a general discrete-time linear vector channel, whose output in the time domain, $\mathbf{y} \in \mathbb{C}^N$, is represented by

$$\mathbf{y} = \sqrt{N}\boldsymbol{\Theta}\mathbf{h} + \mathbf{i} + \mathbf{w}. \tag{1}$$

The work of X. Mestre is supported by Generalitat de Catalunya under grant 2017 SGR 1479 and by the Spanish Government under grant RTI2018-099722-B-I00. The work of M. Payaró is supported by Generalitat de Catalunya under grant 2019 SGR 891.

978-9-0827-9703-9/19/$31.00 ©2019 IEEE

In (1), the unknown channel impulse reponse (CIR) is denoted by $\mathbf{h} \in \mathbb{C}^L$, where L represents the length of the CIR; $\boldsymbol{\Theta} \in \mathbb{C}^{N \times L}$ denotes a known transformation matrix containing the pilots transmitted to estimate the channel (see further [19], [20]); and $\mathbf{i} \in \mathbb{C}^N$ and $\mathbf{w} \in \mathbb{C}^N$ are the vectors containing the time domain samples of IN and background noise, respectively. The samples of the background noise are assumed to be i.i.d. additive white Gaussian noise random variables, $[\mathbf{w}]_j \sim \mathcal{CN}(0, \sigma_w^2)$, where we used $[\mathbf{w}]_j$ to denote the j^{th} sample of vector $\mathbf{w}$. The IN vector $\mathbf{i}$ is a sparse vector, having only N_{imp} non-zero entries and where N_{imp} is unknown. Furthermore, the indexes within the set $\{1, \ldots, N\}$ that are contaminated by IN are denoted by $\mathcal{A} = \left\{n_1, \ldots, n_{N_{\text{imp}}}\right\}$ such that $N_{\text{imp}} = |\mathcal{A}|$. The signal to noise ratio (SNR) of the system is defined as the ratio σ_s^2/σ_w^2, where σ_s^2 is the power of the transmitted signal and the IN to background noise power ratio (INR) is defined as σ_i^2/σ_w^2, where σ_i^2 is the IN power.

In the following two sections, we consider two different models for the IN, deterministic and random. In the deterministic case, the IN vector $\mathbf{i}$ is treated as an unknown, but constant quantity, that can be seen as the mean of the overall noise term $\mathbf{i} + \mathbf{w} \sim \mathcal{CN}(\mathbf{i}, \sigma_w^2 \mathbf{I}_N)$. In the random case, the samples of $\mathbf{i}$ are assumed to be zero mean i.i.d. Gaussian random variables such that $[\mathbf{i}]_j \sim \mathcal{CN}(0, \sigma_i^2)$ if the j^{th} entry of $\mathbf{i}$ contains an IN sample (i.e., $j \in \mathcal{A}$) and $[\mathbf{i}]_j = 0$ if the entry is IN-free (i.e., $j \notin \mathcal{A}$). In the random case, it will prove useful to define the overall Gaussian noise term $\mathbf{n} = \mathbf{i} + \mathbf{w} \sim \mathcal{CN}(\mathbf{0}, \mathbf{C}_n)$.

III. Deterministic model

Let us assume that $N - |\mathcal{A}| - L > 0$. Then, the negative logarithm of the deterministic maximum likelihood (DML) function is

$$\zeta_{DML}\left(\mathbf{h}, \mathbf{i}_{\mathcal{A}}, \sigma_w^2, \mathcal{A}\right) = \frac{1}{N\sigma_w^2} \sum_{n \in \mathcal{A}} \left| [\mathbf{y}]_n - \sqrt{N} \boldsymbol{\theta}_n^H \mathbf{h} - [\mathbf{i}]_n \right|^2$$

$$+ \frac{1}{N\sigma_w^2} \left(\mathbf{y} - \sqrt{N}\boldsymbol{\Theta}\mathbf{h}\right)^H \mathbf{P}_{\mathcal{A}}^{\perp} \left(\mathbf{y} - \sqrt{N}\boldsymbol{\Theta}\mathbf{h}\right) + \log\left(\sigma_w^2 \pi\right).$$

where $\mathbf{i}_{\mathcal{A}} = [\mathbf{i}]_{n \in \mathcal{A}}$, $\boldsymbol{\theta}_n^H$ is the n-th row in matrix $\boldsymbol{\Theta}$, $\mathbf{P}_{\mathcal{A}} = \sum_{n \in \mathcal{A}} \mathbf{u}_n \mathbf{u}_n^H$, with $[\mathbf{u}_n]_j = \delta_{n-j}$ (being δ_n the Kronecker delta), and $\mathbf{P}_{\mathcal{A}}^{\perp} = \mathbf{I}_N - \mathbf{P}_{\mathcal{A}}$, with $\mathbf{I}_N$ being the identity matrix of dimension $N \times N$. We also define the $N \times |\mathcal{A}|$ selection matrix $\mathbf{U}_{\mathcal{A}}$ as the unique matrix with $\{0, 1\}$ entries that satisfies $\mathbf{P}_{\mathcal{A}} = \mathbf{U}_{\mathcal{A}} \mathbf{U}_{\mathcal{A}}^H$.

Let us denote by $\hat{\mathbf{h}}_{\mathcal{A}}$, $\hat{\boldsymbol{\imath}}_{\mathcal{A}}$, and $\hat{\sigma}_{w,\mathcal{A}}^2$ the ML estimators of the CIR, the IN amplitude and the noise power, respectively, where we made explicit the dependence of these estimators on the assumed impulse support, $\mathcal{A}$. These estimators can be obtained by differentiating the cost function $\zeta_{DML}\left(\mathbf{h}, \mathbf{i}_{\mathcal{A}}, \sigma_w^2, \mathcal{A}\right)$ with respect to the corresponding variable and equating it to zero, which yields

$$\hat{\mathbf{h}}_{\mathcal{A}} = \frac{1}{\sqrt{N}} \left(\boldsymbol{\Theta}^H \mathbf{P}_{\mathcal{A}}^{\perp} \boldsymbol{\Theta}\right)^{-1} \boldsymbol{\Theta}^H \mathbf{P}_{\mathcal{A}}^{\perp} \mathbf{y},$$

$$\hat{\boldsymbol{\imath}}_{\mathcal{A}} = \mathbf{U}_{\mathcal{A}}^H \left(\mathbf{y} - \sqrt{N}\boldsymbol{\Theta}^H \hat{\mathbf{h}}_{\mathcal{A}}\right), \text{ and}$$

$$\hat{\sigma}_{w,\mathcal{A}}^2 = \frac{1}{N} \mathbf{y}^H \left(\mathbf{P}_{\mathcal{A}}^{\perp} - \mathbf{P}_{\mathcal{A}}^{\perp} \boldsymbol{\Theta} \left(\boldsymbol{\Theta}^H \mathbf{P}_{\mathcal{A}}^{\perp} \boldsymbol{\Theta}\right)^{-1} \boldsymbol{\Theta}^H \mathbf{P}_{\mathcal{A}}^{\perp}\right) \mathbf{y}.$$

We define by $\zeta_{DML}^{(\mathcal{A})}$ the statistic that is obtained by replacing all these estimators back into the DML cost function, that is

$$\zeta_{DML}^{(\mathcal{A})} = \zeta_{DML}\left(\hat{\mathbf{h}}_{\mathcal{A}}, \hat{\imath}_{\mathcal{A}}, \hat{\sigma}_{w,\mathcal{A}}^2, \mathcal{A}\right) = \log\left(\mathbf{y}^H \mathcal{R}_{\mathcal{A}}^{\perp} \mathbf{y}\right) + C,$$

where C denotes a constant independent of the support $\mathcal{A}$ and where we have defined $\mathcal{R}_{\mathcal{A}}^{\perp} = \mathbf{P}_{\mathcal{A}}^{\perp} - \mathbf{P}_{\mathcal{A}}^{\perp} \boldsymbol{\Theta} \left(\boldsymbol{\Theta}^H \mathbf{P}_{\mathcal{A}}^{\perp} \boldsymbol{\Theta}\right)^{-1} \boldsymbol{\Theta}^H \mathbf{P}_{\mathcal{A}}^{\perp}$.

Our approach for detecting the support of the IN will follow a greedy procedure that will compare the statistic $\zeta_{DML}^{(\mathcal{A})}$ with the statistic $\zeta_{DML}^{(\mathcal{A} \cup \{j\})}$ for every $j \notin \mathcal{A}$. Recalling that these values are negative log-likelihoods, a value of $\zeta_{DML}^{(\mathcal{A} \cup \{j\})}$ that is much lower than $\zeta_{DML}^{(\mathcal{A})}$ will indicate that the support $\mathcal{A} \cup \{j\}$ is much more probable than the support $\mathcal{A}$. Then, our proposed algorithm incorporates the j^{th} sample into the support $\mathcal{A}$ when the difference between negative log-likelihoods $\zeta_{DML}^{(\mathcal{A})} - \zeta_{DML}^{(\mathcal{A} \cup \{j\})}$ is sufficiently high. Noting that the function $-\log(1-x)$ is monotonically increasing for $x \in (0, 1)$, this is equivalent to a sufficiently large value of the statistic [21]:

$$\mathcal{T}_{\mathcal{A}}(j) = \frac{\mathbf{y}^H \mathcal{R}_{\mathcal{A}}^{\perp} \mathbf{u}_j \mathbf{u}_j^H \mathcal{R}_{\mathcal{A}}^{\perp} \mathbf{y}}{\mathbf{y}^H \mathcal{R}_{\mathcal{A}}^{\perp} \mathbf{y} \mathbf{u}_j^H \mathcal{R}_{\mathcal{A}}^{\perp} \mathbf{u}_j}. \tag{2}$$

Now, the first question that we need to answer is, how do we choose a threshold value α_j so that we decide that the j^{th} sample is contaminated by IN if $\mathcal{T}_{\mathcal{A}}(j) > \alpha_j$? Clearly, in order to best fix this threshold, we should take into account all the statistics $\{\mathcal{T}_{\mathcal{A}}(j), j \notin \mathcal{A}\}$ and not only the j^{th} statistic alone. In other words, we should consider the problem as a multiple hypothesis test.

A. Multiple hypothesis test

Let us consider the multiple hypothesis test composed of $N - |\mathcal{A}|$ binary decisions corresponding to the samples not included in $\mathcal{A}$. The j^{th} binary test is defined for every $j \notin \mathcal{A}$ as

$$H_0(j) : [\mathbf{y} - \sqrt{N}\boldsymbol{\Theta}\mathbf{h}]_j \sim \mathcal{CN}\left(0, \sigma_w^2\right)$$
$$H_1(j) : [\mathbf{y} - \sqrt{N}\boldsymbol{\Theta}\mathbf{h}]_j \sim \mathcal{CN}\left([\mathbf{i}]_j, \sigma_w^2\right).$$

The null hypothesis $H_0(j)$ assumes that the j^{th} sample is free from IN, whereas the alternative one $H_1(j)$ considers IN contamination with complex amplitude $[\mathbf{i}]_j$. For each binary test ($H_0(j)$ versus $H_1(j)$) we may construct the generalized likelihood ratio statistic in (2) and we decide that the alternative is correct if $\mathcal{T}_{\mathcal{A}}(j) > \alpha_j$ for a given threshold α_j. We will now sketch how the threshold values α_j for each $j \notin \mathcal{A}$ can be selected. For details, see [21].

Conventional binary hypothesis tests select the threshold value α_j in order to have a certain probability of false alarm, defined as the probability of incorrectly deciding for $H_1(j)$ when the correct hypothesis is $H_0(j)$, namely $\text{FA}_j = \mathbb{P}_{H_0(j)}(\mathcal{T}_{\mathcal{A}}(j) > \alpha_j)$, where $\mathbb{P}_{H_0(j)}(\cdot)$ is the probability of a certain event under the hypothesis $H_0(j)$. To fix this probability, we must investigate the statistical behavior of the statistic $\mathcal{T}_{\mathcal{A}}(j)$ under $H_0(j)$, meaning that $\mathbf{y}_p - \sqrt{N}\mathbf{\Theta}\mathbf{h} \sim \mathcal{CN}\left(0, \sigma_w^2 \mathbf{I}_N\right)$. In [21], we show that

$$\frac{\mathcal{T}_{\mathcal{A}}(j)}{1-\mathcal{T}_{\mathcal{A}}(j)} \sim \frac{\chi_2^2}{\chi^2_{2(N-|\mathcal{A}|-L-1)}} = \frac{1}{N-|\mathcal{A}|-L-1} F\left(2, 2\left(N-|\mathcal{A}|-L-1\right)\right), \tag{3}$$

where $F(d_1, d_2)$ is the Snedecor F-distribution. We can now compute the false alarm probability for the j^{th} binary test, $\mathrm{FA}_j = \mathbb{P}_{H_0(j)}\left(\mathcal{T}_{\mathcal{A}}(j) \geq \alpha_j\right)$, in closed form as:

$$\mathrm{FA}_j = \mathbb{P}_{H_0(j)}\left(\frac{\mathcal{T}_{\mathcal{A}}(j)}{1-\mathcal{T}_{\mathcal{A}}(j)} \geq \frac{\alpha_j}{1-\alpha_j}\right) = \left(1-\alpha_j\right)^{N-|\mathcal{A}|-L-1},$$

where we have used that $\mathbb{P}\left(F\left(2,2d\right) \leq x\right) = 1-(d/(x+d))^d$.

Therefore, we can fix the threshold α_j that guarantees a certain false alarm probability FA_j as $\alpha_j = 1 - \mathrm{FA}_j^{1/(N-|\mathcal{A}|-L-1)}$.

IV. Random model

A similar greedy algorithm for the detection of IN support as presented above can also be formulated under a random Gaussian model for the IN. In this case, the system model (1) becomes $\mathbf{y} = \sqrt{N}\mathbf{\Theta}\mathbf{h} + \mathbf{n}$ with $\mathbf{n} = \mathbf{i} + \mathbf{w} \sim \mathcal{CN}(\mathbf{0}, \mathbf{C}_n)$ and the negative logarithm of the random maximum likelihood (RML) function is

$$\zeta_{RML}\left(\mathbf{h}, \sigma^2, \sigma_w^2, \mathcal{A}\right) = \frac{1}{N}\log\det \mathbf{C}_n + \frac{1}{N}\left(\mathbf{y} - \sqrt{N}\mathbf{\Theta}\mathbf{h}\right)^H \mathbf{C}_n^{-1}\left(\mathbf{y} - \sqrt{N}\mathbf{\Theta}\mathbf{h}\right),$$

where $\mathbf{C}_n = \sigma^2 \mathbf{P}_{\mathcal{A}} + \sigma_w^2 \mathbf{P}_{\mathcal{A}}^{\perp}$ and $\sigma^2 = \sigma_i^2 + \sigma_w^2$. By taking the RML channel estimate (denoted here also $\hat{\mathbf{h}}_{\mathcal{A}}$, with a slight abuse of notation) and inserting it back into the cost function we obtain [21]

$$\zeta_{RML}\left(\hat{\mathbf{h}}_{\mathcal{A}}, \sigma^2, \sigma_w^2, \mathcal{A}\right) = \frac{|\mathcal{A}|}{N}\log\sigma^2 + \left(1 - \frac{|\mathcal{A}|}{N}\right)\log\sigma_w^2 + \frac{1}{N\sigma_w^2}\mathbf{y}^H \mathbf{P}_{\mathbf{\Theta}}^{\perp}\mathbf{y} - \left(\frac{\sigma^2}{\sigma_w^2} - 1\right)\frac{1}{N}\sum_{i=1}^{|\mathcal{A}|}\frac{\mathbf{y}^H \mathbf{P}_{\mathbf{\Theta}}^{\perp}\mathbf{U}_{\mathcal{A}}\mathbf{u}_{\mathcal{A}}(i)\,\mathbf{u}_{\mathcal{A}}^H(i)\,\mathbf{U}_{\mathcal{A}}^H \mathbf{P}_{\mathbf{\Theta}}^{\perp}\mathbf{y}}{\sigma_w^2 + (\sigma^2 - \sigma_w^2)\,\lambda_{\mathcal{A}}(i)}, \tag{4}$$

where $\mathbf{U}_{\mathcal{A}}$ is the same as defined in section III and we have used the eigendecomposition $\mathbf{U}_{\mathcal{A}}^H \mathbf{P}_{\mathbf{\Theta}}^{\perp}\mathbf{U}_{\mathcal{A}} = \sum_{i=1}^{|\mathcal{A}|}\lambda_{\mathcal{A}}(i)\,\mathbf{u}_{\mathcal{A}}(i)\,\mathbf{u}_{\mathcal{A}}^H(i)$.

The proposed estimator for the IN support will follow the same greedy approach as the one for the deterministic case, based on the difference $\zeta_{RML}^{(\mathcal{A})} - \zeta_{RML}^{(\mathcal{A}\cup\{j\})}$. The main idea is to compute $\zeta_{RML}^{(\mathcal{A}\cup\{j\})}$ for all $j \notin \mathcal{A}$. Assuming that the value of the negative log-likelihood $\zeta_{RML}^{(\mathcal{A}\cup\{j\})}$ is sufficiently lower than the value of $\zeta_{RML}^{(\mathcal{A})}$, we will have evidence that the model that incorporates the j^{th} sample into the IN support is more likely. This will lead us to the conclusion that the j^{th} sample should be incorporated into the model. Following the dertivations in [21], $\zeta_{RML}^{(\mathcal{A}\cup\{j\})}$ can be expressed as

$$\zeta_{RML}\left(\hat{\mathbf{h}}_{\mathcal{A}\cup\{j\}}, \sigma^2, \sigma_w^2, \mathcal{A}\cup\{j\}\right) = \frac{|\mathcal{A}|+1}{N}\log\sigma^2 + \left(1 - \frac{|\mathcal{A}|+1}{N}\right)\log\sigma_w^2 + \frac{1}{N\sigma_w^2}\mathbf{y}^H \mathcal{R}_{\mathcal{A}}(\alpha)\,\mathbf{y} - \frac{1}{N\sigma_w^2}\frac{\mathbf{y}^H \mathcal{R}_{\mathcal{A}}(\alpha)\,\mathbf{u}_j \mathbf{u}_j^H \mathcal{R}_{\mathcal{A}}(\alpha)\,\mathbf{y}}{\alpha + \mathbf{u}_j^H \mathcal{R}_{\mathcal{A}}(\alpha)\,\mathbf{u}_j},$$

where $\mathcal{R}_{\mathcal{A}}(x) = \mathbf{P}_{\Theta}^{\perp} - \mathbf{P}_{\Theta}^{\perp}\mathbf{U}_{\mathcal{A}}(x\mathbf{I}_{|\mathcal{A}|} + \mathbf{U}_{\mathcal{A}}^H \mathbf{P}_{\Theta}^{\perp}\mathbf{U}_{\mathcal{A}})^{-1}\mathbf{U}_{\mathcal{A}}^H \mathbf{P}_{\Theta}^{\perp}$ and $\alpha = \sigma_w^2/(\sigma^2 - \sigma_w^2)$. Observe that, according to the definition of $\sigma^2 = \sigma_i^2 + \sigma_w^2$, one always has $\sigma^2 \geq \sigma_w^2$ so that $\alpha \in (0, \infty)$. Also, observe that we have $\mathcal{R}_{\mathcal{A}}^{\perp} = \mathcal{R}_{\mathcal{A}}(0)$.

For the optimization of the cost function ζ_{RML} with respect to the pair $\left(\sigma^2, \sigma_w^2\right)$ we need to take the derivatives and equate them to zero. Following the derivations in [21] and defining the quantity

$$\mathcal{R}_{\mathcal{A}\cup\{j\}}(x) = \mathcal{R}_{\mathcal{A}}(x) - \frac{\mathcal{R}_{\mathcal{A}}(x)\,\mathbf{u}_j \mathbf{u}_j^H \mathcal{R}_{\mathcal{A}}(x)}{x + \mathbf{u}_j^H \mathcal{R}_{\mathcal{A}}(x)\,\mathbf{u}_j}, \tag{5}$$

the minimum will be attained at the pair of points:

$$\hat{\sigma}^2_{\mathcal{A}\cup\{j\}} = \frac{\beta^2_{\mathcal{A}\cup\{j\}}}{\left(\beta_{\mathcal{A}\cup\{j\}} - 1\right)\left(|\mathcal{A}|+1\right)}\mathbf{y}^H \mathcal{R}_{\mathcal{A}\cup\{j\}}\left(\alpha_{\mathcal{A}\cup\{j\}}\right) \times \left(\mathbf{I}_N - \mathcal{R}_{\mathcal{A}\cup\{j\}}\left(\alpha_{\mathcal{A}\cup\{j\}}\right)\right)\mathbf{y}$$

$$\hat{\sigma}^2_{w,\mathcal{A}\cup\{j\}} = \frac{1}{\left(\beta_{\mathcal{A}\cup\{j\}} - 1\right)\left(N - |\mathcal{A}| - 1\right)}\mathbf{y}^H \mathcal{R}_{\mathcal{A}\cup\{j\}} \times \left(\alpha_{\mathcal{A}\cup\{j\}}\right)\left(\mathbf{I}_N - \beta_{\mathcal{A}\cup\{j\}}\mathcal{R}_{\mathcal{A}\cup\{j\}}\left(\alpha_{\mathcal{A}\cup\{j\}}\right)\right)\mathbf{y},$$

where $\alpha_{\mathcal{A}} = (\beta_{\mathcal{A}} - 1)^{-1}$ and $\beta_{\mathcal{A}}$ is a solution to the equation in β:

$$\frac{1}{N-|\mathcal{A}|-1}\mathbf{y}^H \mathcal{R}_{\mathcal{A}\cup\{j\}}(\alpha)\left(\mathbf{I}_N - \beta\mathcal{R}_{\mathcal{A}\cup\{j\}}(\alpha)\right)\mathbf{y} = \frac{\beta}{|\mathcal{A}|+1}\mathbf{y}^H \mathcal{R}_{\mathcal{A}\cup\{j\}}(\alpha)\left(\mathbf{I}_N - \mathcal{R}_{\mathcal{A}\cup\{j\}}(\alpha)\right)\mathbf{y},$$

if there exists one with $\beta_{\mathcal{A}} \geq 1$ or, otherwise, $\hat{\sigma}^2_{\mathcal{A}\cup\{j\}} = \hat{\sigma}^2_{w,\mathcal{A}\cup\{j\}} = \frac{1}{N}\mathbf{y}^H \mathbf{P}_{\mathbf{\Theta}}^{\perp}\mathbf{y}$. Note that all these equations can be formulated in terms of the eigendecomposition of $\mathcal{R}_{\mathcal{A}}(x)$ by using the identity in (5).

A. Multiple hypothesis test

Following the same approach as in the deterministic case, one can now formulate a greedy algorithm for the detection of the IN support. As before, given a support set $\mathcal{A}$, we ask ourselves whether there exist other samples outside $\mathcal{A}$ that are contaminated by IN. To solve this, we consider the multiple hypothesis test composed of $N - |\mathcal{A}|$ binary decisions corresponding to each of the j^{th} samples not included in $\mathcal{A}$. The j^{th} test is defined for every $j \notin \mathcal{A}$ as

$$H_0(j): \left(\mathbf{y} - \sqrt{N}\mathbf{\Theta}\mathbf{h}\right) \sim \mathcal{CN}\left(\mathbf{0}, \sigma^2 \mathbf{P}_{\mathcal{A}} + \sigma_w^2 \mathbf{P}_{\mathcal{A}}^{\perp}\right)$$

$$H_1(j): \left(\mathbf{y} - \sqrt{N}\mathbf{\Theta}\mathbf{h}\right) \sim \mathcal{CN}\left(\mathbf{0}, \sigma^2 \mathbf{P}_{\mathcal{A}\cup\{j\}} + \sigma_w^2 \mathbf{P}_{\mathcal{A}\cup\{j\}}^{\perp}\right).$$

Note that, as in the case in section III, the null hypothesis $H_0(j)$ assumes that the j^{th} sample is free from impulsive noise, whereas the alternative one $H_1(j)$ considers that the j^{th} sample has larger variance σ^2. For each binary test ($H_0(j)$ versus $H_1(j)$) we decide that the alternative is correct if

$$\zeta_{RML}^{(\mathcal{A})} - \zeta_{RML}^{(\mathcal{A}\cup\{j\})} > \alpha_j \tag{6}$$

for a given threshold α_j (related to the probability of false alarm as shown by the end of this section) and where we define $\zeta_{RML}^{(\mathcal{A})}$ as the value of the negative log-likelihood function after replacing all parameters with their RML estimates, namely $\zeta_{RML}^{(\mathcal{A})} = \zeta_{RML}(\hat{\mathbf{h}}_{\mathcal{A}}, \hat{\sigma}_{\mathcal{A}}^2, \hat{\sigma}_{w,\mathcal{A}}^2, \mathcal{A})$.

Therefore, from (6), we need to investigate the statistical behavior of the test $\zeta_{RML}^{(\mathcal{A})} - \zeta_{RML}^{(\mathcal{A}\cup\{j\})}$ under the null hypothesis. Given the complex form of the RML estimator and the fact that it does not accept a closed form expression, it is in general extremely difficult to characterize the statistical law of $\zeta_{RML}^{(\mathcal{A})}$. In order to solve this issue, we will take here an asymptotic approach and analyze the behavior of this statistic when both $N, |\mathcal{A}| \to \infty$ at the same rate. The main statistical assumptions that are made for the asymptotic analysis are:

($\mathbf{As1}$) The noise term is zero mean, circularly symmetric and Gaussian distributed with covariance $\bar{\mathbf{C}}_n$, that is $\mathbf{n} \sim \mathcal{CN}\left(0, \bar{\mathbf{C}}_n\right)$.

($\mathbf{As2}$) Both N and $|\mathcal{A}|$ converge to infinity at the same rate: $0 < \liminf_N |\mathcal{A}|/N \leq \limsup_N |\mathcal{A}|/N < 1$.

($\mathbf{As3}$) The eigenvalues of the matrix $\mathbf{\Theta}^H\mathbf{\Theta}$ are contained in a compact interval of the positive real axis for all N, that is $0 < \inf_N \mathbf{\Theta}^H\mathbf{\Theta} < \sup_N \mathbf{\Theta}^H\mathbf{\Theta} < \infty$. Furthermore, if $\mathbf{P}_{\mathcal{A}}$ is the $N \times N$ selection matrix corresponding to the set $\mathcal{A} \subset [N]$ we have $\inf_N \mathbf{\Theta}^H\mathbf{P}_{\mathcal{A}}\mathbf{\Theta} > 0$ and $\inf_N \mathbf{\Theta}^H\mathbf{P}_{\mathcal{A}}^{\perp}\mathbf{\Theta} > 0$.

($\mathbf{As4}$) The norm of the rows of $\mathbf{\Theta}$ decays uniformly to zero as $O\left(N^{-1}\right)$ or faster, that is $\sup_N \max_{j=1,\ldots,N} N\left\|\mathbf{u}_j^H\mathbf{\Theta}\right\| < \infty$.

($\mathbf{As5}$) It holds that

$$\limsup_N \left(\frac{1}{N}\mathrm{tr}\left(\bar{\mathbf{C}}_n\right) - \frac{1}{|\mathcal{A}|}\mathrm{tr}\left(\bar{\mathbf{C}}_n\mathbf{P}_{\mathcal{A}}\right)\right) < 0.$$

If $\hat{\mathbf{h}}_{\mathcal{A}}$ denotes the RML channel estimator when the support $\mathcal{A}$ is assumed and denoting as $\bar{\mathbf{h}}$ the true channel impulse response, under ($\mathbf{As1}$) $-$ ($\mathbf{As3}$) we have

$$\hat{\mathbf{h}}_{\mathcal{A}} = \bar{\mathbf{h}} + \frac{1}{\sqrt{N}}\left(\mathbf{\Theta}^H\widetilde{\mathbf{C}}_{n,\mathcal{A}}^{-1}\mathbf{\Theta}\right)^{-1}\mathbf{\Theta}^H\widetilde{\mathbf{C}}_{n,\mathcal{A}}^{-1}\mathbf{n} + o_p\left(\frac{1}{\sqrt{N}}\right), \tag{7}$$

with $\widetilde{\mathbf{C}}_{n,\mathcal{A}} = \widetilde{\sigma}_{\mathcal{A}}^2\mathbf{P}_{\mathcal{A}} + \widetilde{\sigma}_{w,\mathcal{A}}^2\mathbf{P}_{\mathcal{A}}^{\perp}$, $\widetilde{\sigma}_{\mathcal{A}}^2 = \mathrm{tr}\left(\mathbf{P}_{\mathcal{A}}\bar{\mathbf{C}}_n\right)/|\mathcal{A}|$, and $\widetilde{\sigma}_{w,\mathcal{A}}^2 = \mathrm{tr}\left(\mathbf{P}_{\mathcal{A}}^{\perp}\bar{\mathbf{C}}_n\right)/(N - |\mathcal{A}|)$. This result can be used to establish the following proposition, which provides an asymptotic description of the statistic $\zeta_{RML}^{(\mathcal{A})} - \zeta_{RML}^{(\mathcal{A}\cup\{j\})}$.

Proposition 1: Under ($\mathbf{As1}$) $-$ ($\mathbf{As5}$), we have

$$\begin{aligned}\zeta_{RML}^{(\mathcal{A})} - \zeta_{RML}^{(\mathcal{A}\cup\{j\})} &= \frac{1}{N}\log\left(\frac{\widetilde{\sigma}_{w,\mathcal{A}}^2}{\widetilde{\sigma}_{\mathcal{A}}^2}\right) \\ &+ \frac{1}{N}\left[\bar{\mathbf{C}}_n\right]_{jj}\left(\frac{\widetilde{\sigma}_{\mathcal{A}}^2 - \widetilde{\sigma}_{w,\mathcal{A}}^2}{\widetilde{\sigma}_{w,\mathcal{A}}^2\widetilde{\sigma}_{\mathcal{A}}^2}\right) \\ &+ \frac{1}{N}\mathbf{n}^H\mathcal{Q}_{\mathcal{A}}^{\perp}\Psi_{\mathcal{A}}(j)\,\mathcal{Q}_{\mathcal{A}}^{\perp}\mathbf{n} + o_p\left(\frac{1}{N}\right)\end{aligned}$$

where we have defined

$$\mathcal{Q}_{\mathcal{A}}^{\perp} = \widetilde{\mathbf{C}}_{n,\mathcal{A}}^{-1} - \widetilde{\mathbf{C}}_{n,\mathcal{A}}^{-1}\mathbf{\Theta}\left(\mathbf{\Theta}^H\widetilde{\mathbf{C}}_{n,\mathcal{A}}^{-1}\mathbf{\Theta}\right)^{-1}\mathbf{\Theta}^H\widetilde{\mathbf{C}}_{n,\mathcal{A}}^{-1},$$

$$\begin{aligned}\Psi_{\mathcal{A}}(j) &= \frac{\left[\bar{\mathbf{C}}_n\right]_{jj} - \widetilde{\sigma}_{\mathcal{A}}^2}{|\mathcal{A}|}\mathbf{P}_{\mathcal{A}} + \frac{\widetilde{\sigma}_{w,\mathcal{A}}^2 - \left[\bar{\mathbf{C}}_n\right]_{jj}}{N - |\mathcal{A}|}\mathbf{P}_{\mathcal{A}}^{\perp} \\ &+ \left(\widetilde{\sigma}_{\mathcal{A}}^2 - \widetilde{\sigma}_{w,\mathcal{A}}^2\right)\frac{\widetilde{\sigma}_{w,\mathcal{A}}^2}{\widetilde{\sigma}_{\mathcal{A}}^2}\mathbf{u}_j\mathbf{u}_j^T.\end{aligned}$$

Proof: The derivations can be found in [21]. ■

The above proposition provides a mean to fix the threshold levels α_j, $j \notin \mathcal{A}$, to guarantee a certain asymptotic false alarm probability for each binary hypothesis test. To ensure that, in [21], it is shown that under the null hypothesis $H_0(j)$ we have $\mathbf{y} \sim \mathcal{CN}(\sqrt{N}\mathbf{\Theta}\mathbf{h}, \sigma^2\mathbf{P}_{\mathcal{A}} + \sigma_w^2\mathbf{P}_{\mathcal{A}}^{\perp})$, which means that $\widetilde{\sigma}_{\mathcal{A}}^2 = \sigma^2$, $\widetilde{\sigma}_{w,\mathcal{A}}^2 = \left[\bar{\mathbf{C}}_n\right]_{jj} = \sigma_w^2$ and

$$\begin{aligned}\zeta_{RML}^{(\mathcal{A})} - \zeta_{RML}^{(\mathcal{A}\cup\{j\})} &= \frac{1}{N}\log\left(\frac{\sigma_w^2}{\sigma^2}\right) + \frac{1}{N}\left(1 - \frac{\sigma_w^2}{\sigma^2}\right) \\ -\frac{1}{N}\left(1 - \frac{\sigma_w^2}{\sigma^2}\right)&\left[\frac{\mathbf{n}^H\mathbf{P}_{\mathcal{A}}\mathbf{n}}{\sigma^2|\mathcal{A}|} - \frac{\mathbf{n}^H\mathbf{u}_j\mathbf{u}_j^T\mathbf{n}}{\sigma_w^2}\right] + o_p\left(\frac{1}{N}\right).\end{aligned}$$

Now, clearly $\mathbf{u}_j^T\mathbf{n} \sim \mathcal{CN}\left(0, \sigma_w^2\right)$ and $\mathbf{U}_{\mathcal{A}}^H\mathbf{n} \sim \mathcal{CN}\left(0, \sigma^2\mathbf{I}_{|\mathcal{A}|}\right)$ and these two random variables are independent. Thus, it follows that

$$\frac{\mathbf{n}^H\mathbf{P}_{\mathcal{A}}\mathbf{n}}{\sigma^2|\mathcal{A}|} - \frac{\mathbf{n}^H\mathbf{u}_j\mathbf{u}_j^T\mathbf{n}}{\sigma_w^2} \sim \frac{1}{2}\left(\frac{1}{|\mathcal{A}|}\chi_{2|\mathcal{A}|}^2 - \chi_2^2\right),$$

where χ_2^2 and $\chi_{2|\mathcal{A}|}^2$ are two independent Chi-Square variables with 2 and $2|\mathcal{A}|$ degrees of freedom respectively. Let us write $\chi = \frac{1}{|\mathcal{A}|}\chi_{2|\mathcal{A}|}^2 - \chi_2^2$. The cumulative density function of this random variable is well known to be [22], [23]:

$$F_\chi(y) = \mathbb{P}(\chi \leq y) = \begin{cases}\left(\frac{|\mathcal{A}|}{|\mathcal{A}|+1}\right)^{|\mathcal{A}|}\mathrm{e}^{\frac{y}{2}} & y < 0, \\ 1 - \frac{\mathrm{e}^{-\frac{y|\mathcal{A}|}{2}}}{|\mathcal{A}|+1}\sum_{i=0}^{|\mathcal{A}|-1}\Upsilon & y \geq 0,\end{cases} \tag{8}$$

with $\Upsilon = \sum_{\ell=0}^{i}\frac{1}{(i-l)!}\left(\frac{|\mathcal{A}|}{|\mathcal{A}|+1}\right)^{|\mathcal{A}|-1-i}\left(\frac{y|\mathcal{A}|}{2}\right)^{i-\ell}$.

In order to achieve a certain false alarm probability we should choose α_j such that $\mathrm{FA}_j = \mathbb{P}_{H_0(j)}\left(\zeta_{RML}^{(\mathcal{A})} - \zeta_{RML}^{(\mathcal{A}\cup\{j\})} \geq \alpha_j\right)$. This probability can be asymptotically approximated by

$$\mathrm{FA}_j = \mathbb{P}\left(\chi \leq \left(\frac{2\sigma^2}{\sigma^2 - \sigma_w^2}\right)\log\left(\frac{\sigma_w^2}{\sigma^2}\right) + 2 - N\alpha_j\left(\frac{2\sigma^2}{\sigma^2 - \sigma_w^2}\right)\right)$$

and, therefore, we should fix α_j such that

$$\alpha_j = \frac{1}{N}\log\left(\frac{\sigma_w^2}{\sigma^2}\right) + \frac{\sigma^2 - \sigma_w^2}{2N\sigma^2}\left(2 - F_\chi^{-1}(\mathrm{FA}_j)\right),$$

where $F_\chi^{-1}(p)$ is the functional inverse of $F_\chi(y)$.

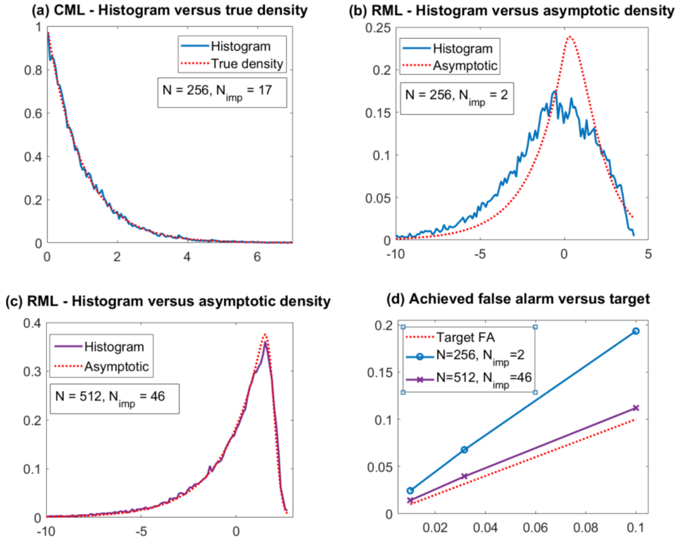

Fig. 1. (a,b,c) Histogram of the statistic versus true/asymptotic density for different values of N, N_{imp}. (d) Achieved false alarm rate versus the predefined target $\text{FA}_j \in [10^{-2}, 10^{-1}]$.

V. Numerical validation and conclusion

In this section, we numerically validate the theoretical results derived above by evaluating the accuracy of the two statistics derived under the deterministic and the random models. Then, we evaluate the validity of the multi-hypothesis tests for the false alarm probability in the RML case only, since the result we obtained in this case in Proposition 1 is asymptotic (thus, an approximation). For the DML case, we do not perform this evaluation as the statistic derived for (2) is exact. Finally, for the sake of space, we do not provide performance results of the channel parameter estimation (CIR, noise covariances), as these results are available in [19].

We consider here a scenario with a variable number of impulses, where the SNR is fixed to 10 dB and the INR to 20 dB. The CIR is randomly selected with an exponentially decaying power delay profile of duration equal to 20 samples, and the receiver assumed a total channel length of $L = 30$ samples. For each scenario, a total of 100 realizations of the input signal is generated, and for each realization we compute the $N - N_{\text{imp}}$ statistics corresponding to (2) and (6) for all j outside the support of the IN. The resulting values are then conveniently transformed and compared to the two densities in (3) –exact DML– and (8) –asymptotic RML– respectively.

In Fig.1(a,b,c) we represent the histograms and exact/asymptotic probability density functions corresponding to the DML and RML statistics for different values of N and N_{imp}. Observe that there is a perfect match in the DML case and that a higher accuracy is observed for large values N, N_{imp} in the RML case, as expected. In Fig.1(d), for the RML case, we represent the achieved false alarm rate versus the FA_j target that ranges from 10^{-2} to 10^{-1}. As it can be seen, the achieved performance is very close to the target, especially as N and N_{imp} become larger.

References

[1] M. Ghosh, "Analysis of the effect of impulse noise on multicarrier and single carrier QAM systems," *IEEE Transactions on Communications*, vol. 44, no. 2, pp. 145–147, Feb 1996.

[2] T. N. Zogakis, P. S. Chow, J. T. Aslanis, and J. M. Cioffi, "Impulse noise mitigation strategies for multicarrier modulation," in *Proc. of IEEE International Conference on Communications, 1993 (ICC '93) Geneva.*, May 1993, vol. 2, pp. 784–788.

[3] Y. R. Chien, "Iterative channel estimation and impulsive noise mitigation algorithm for OFDM-based receivers with application to power-line communications," *IEEE Trans. on Power Delivery*, vol. 30, Dec 2015.

[4] H. A. Suraweera, C. Chai, J. Shentu, and J. Armstrong, "Analysis of impulse noise mitigation techniques for digital television systems," in *8th International OFDM Workshop*, Sept. 2003, pp. 172–176.

[5] G. Ndo, P. Siohan, and M. H. Hamon, "Adaptive noise mitigation in impulsive environment: Application to power-line communications," *IEEE Trans. on Power Delivery*, vol. 25, no. 2, pp. 647–656, April 2010.

[6] B. Adebisi, K. Anoh, K. M. Rabie, A. Ikpehai, M. Fernando, and A. Wells, "A new approach to peak threshold estimation for impulsive noise reduction over power line fading channels," *IEEE Syst. Jrnl.*, 2018.

[7] T. Bai, C. Xu, R. Zhang, A. F. Al Rawi, and L. Hanzo, "Joint impulsive noise estimation and data detection conceived for LDPC-coded DMT-based DSL systems," *IEEE Access*, vol. 5, pp. 23133–23145, 2017.

[8] L. Lampe, "Bursty impulse noise detection by compressed sensing," in *Proc. of IEEE International Symposium on Power Line Communications and Its Applications*, April 2011, pp. 29–34.

[9] J. Zhang, Z. He, P. Chen, and Y. Rong, "A compressive sensing based iterative algorithm for channel and impulsive noise estimation in underwater acoustic ofdm systems," in *OCEANS 2017*, Sept 2017.

[10] H. Zhang, L. L. Yang, and L. Hanzo, "Compressed impairment sensing-assisted and interleaved-double-FFT-aided modulation improves broadband power line communications subjected to asynchronous impulsive noise," *IEEE Access*, vol. 4, pp. 81–96, 2016.

[11] T. Y. Al-Naffouri, F. F. Al-Shaalan, A. A. Quadeer, and H. Hmida, "Impulsive noise estimation and cancellation in DSL using compressive sampling," in *Proc. of IEEE International Symposium of Circuits and Systems (ISCAS'11)*, May 2011, pp. 2133–2136.

[12] D. L. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, April 2006.

[13] A. B. Ramirez, R. E. Carrillo, G. Arce, K. E. Barner, and B. Sadler, "An overview of robust compressive sensing of sparse signals in impulsive noise," in *Proc. of European Signal Processing Conference (EUSIPCO'15)*, Aug 2015, pp. 2859–2863.

[14] W. Dai and O. Milenkovic, "Subspace pursuit for compressive sensing: Closing the gap between performance and complexity," *Preprint*, March 2008. Available online: https://arxiv.org/abs/0803.0811.

[15] Needell and J. A. Tropp, "Cosamp: Iterative signal recovery from incomplete and inaccurate samples," *Preprint*, March 2008. Available online at: https://arxiv.org/abs/0803.2392. A later published version is available at https://doi.org/10.1016/j.acha.2008.07.002.

[16] T. T. Do, L. Gan, N. Nguyen, and T. D. Tran, "Sparsity adaptive matching pursuit algorithm for practical compressed sensing," in *Proc. of Asilomar Conference on Signals, Systems and Computers*, Oct 2008.

[17] S. Liu, F. Yang, W. Ding, and J. Song, "A priori aided compressive sensing approach for impulsive noise reconstruction," in *Proc. of International Wireless Communications and Mobile Computing Conference (IWCMC'15)*, Aug 2015, pp. 205–209.

[18] F. Abdelkefi and J. Ayadi, "Efficient techniques for impulsive noise cancellation in CGU/SD systems," *IEEE Transactions on Signal Processing*, vol. 65, no. 14, pp. 3749–3760, July 2017.

[19] D. Shrestha, X. Mestre, and M. Payaró, "On channel estimation for power line communication systems in the presence of impulsive noise," *ELSEVIER Computers & Electrical Engineering*, vol. 72, Nov 2018.

[20] G. Leus and A. J. van der Veen, "Channel estimation," in *Smart Antennas – State of the art*, chapter 5. Hindawi, 2006.

[21] D. Shrestha, X. Mestre, and M. Payaró, "On maximum-likelihood joint channel estimation and impulsive noise support detection," *submitted to IEEE Trans. on Signal Processing*, Mar 2019. Derivations and proofs are available at https://cloud.cttc.es/index.php/s/EijcyLGdf8tecgx.

[22] J.K. Omura and T. Kailath, "Some useful probability functions," *Tech. Rep*, vol. 7050-6, 1965.

[23] M.K. Simon, *Probability Distributions Involving Gaussian Random Variables: A Handbook for Engineers and Scientists*, Springer, 2006.