An ML Algorithm for Outliers Detection and Source Localization

Victor A. N. Barroso † *José M. F. Moura* ‡

† CAPS, Dep. de Eng. Elect. e de Comp., Instituto Superior Técnico, 1096 Lisboa Codex, Portugal
‡ LASIP, Dep. of Elect. and Comp. Eng., Carnegie Mellon University, Pittsburgh, PA 15213, USA

Abstract

This paper addresses the problem of simultaneous detection of outliers and localization of multiple sources. This is motivated by the performance degradation observed when quadratic beamformers operate under those conditions. Our approach relies on maximum likelihood methods where outliers are modeled as a space/time impulsive noise process with unknown statistics. The maximization algorithm follows a strategy based on sequencial estimation and detection schemes, and it is initialized by an l_1 beamformer, yielding efficient detection of spikes and accurate estimates of their statistics. This enables the design of a model based beamformer for bearing estimation. The paper presents the derivation of the algorithm and discusses its efficiency using the results obtained from computer simulations.

1 Introduction

In this work, we consider the problem of localizing multiple independent sources when outliers are present, namely, due to the existence of failling sensors or to the occurrence of spikes. In these situations, the performance of beamformers based on the Gaussian noise assumption is seriously degraded, [2]. Our approach consists on the maximization of the likelihood of the observations set, where the outliers are modeled as impulsive noise. This is described by a high variance Gaussian random variable with samples governed by a space/time Bernoulli sequence. Obviously, the maximization of the likelihood function involves an optimal estimation of that space/time sequence, thus providing knowledge about the sensors and the time instants at which the impulses have occurred. However, the computational complexity of this approach increases exponentially with the number of array sensors and time samples, precluding its practical application. To combat the drawbacks of the direct solution, different detection algorithms have been used, [5, 4]. Our method relies on the robustness of the l_1 beamformer in the presence of spiky noise, [1]. In fact, the l_1 beamformer has the capability of self adjusting its gains in order to attenuate severely the effects of input outliers, thus yielding output residues that evidence the presence of spiky impulses. After robust detection of these space/time events, the statistics of the impulsive noise field are estimated, enabling the design of an optimal impulse detector and of a maximum likelihood beamformer for bearing estimation. These are used iteratively to improve the initial estimates provided by the l_1 beamformer. The simulation results show that very few iterations are needed to achieve accurate estimates of both the impulsive noise statistics and the bearing angles.

The paper is organized as follows. In section 2, we formulate the problem and we introduce the likelihood function of the data set. The initialization step, based on the l_1 beamformer, is also presented and discussed. In section 3, we describe the iterative algorithm, focusing on both the optimal impulse detector and the model based beamformer. Finally, in section 4, we report and discuss the results obtained by computer simulations, and we present the main conclusions of this work.

2 Problem Formulation

Let

$$\mathbf{z}(k) = \mathbf{a}(\theta)x(k) + \mathbf{w}(k) + \mathbf{s}(k) \tag{1}$$

be the complex envelope of the $(N \times 1)$ vector of array measurements at time instant k, where $x(\cdot)$ and $\mathbf{w}(\cdot)$ are the complex envelopes of the desired signal and of the background noise, respectively. The latter is the superposition of unknown directional interferences with a complex Gaussian white noise field with zero mean and known covariance matrix $R_0 = \sigma_0^2 I$. The covariance matrix of the total background noise field is then

$$R_w = R_i + \sigma_0^2 I, \tag{2}$$

where R_i is the covariance matrix of the interfering wave fields. The complex $(N \times 1)$ steering vector $\mathbf{a}(\theta)$ is specified by the arrival angle θ. In (1), $\mathbf{s}(\cdot)$ is the impulsive noise vector field, which is assumed to be independent of $\mathbf{w}(\cdot)$. At each sensor, the impulsive noise sample is defined by the product of three jointly independent random variables (r.v.): a time indexed Bernoulli r.v. $b(\cdot)$, a space/time indexed Bernoulli r.v. $d_n(\cdot)$, and a complex valued space/time indexed zero mean Gaussian r.v. $u_n(\cdot)$ with variance σ_u^2. In general, the power of the impulsive noise field is stronger than that of the background white noise, i.e., $\sigma_u^2 \gg \sigma_0^2$. Formally, we write

$$\mathbf{s}(k) = b(k)D(k)\mathbf{u}(k), \tag{3}$$

where $D(k) = \mathrm{diag}[d_n(k)]$. The samples of the space/time sequence $s_n(k)$ are assumed independent, and the probability of the space/time events is $p_b p_d$ where $p_b = \Pr[b(k) = 1]$ and $p_d = \Pr[d_n(k) = 1]$.

Let $\mathbf{Z} = \{\mathbf{z}(k)\}_{k=1}^{K}$ and $\mathbf{X} = \{x(k)\}_{k=1}^{K}$ be the observations set and the time samples of the desired signal, respectively. Define also the sets of space/time descrete events $\mathbf{B} = \{b(k)\}_{k=1}^{K}$ and $\mathbf{D} = \{D(k)\}_{k=1}^{K}$. According to the model here described, $\mathbf{B}$ and $\mathbf{D}$ are random, while $\mathbf{X}$ and the unknown parameters θ, p_b, p_d and σ_u^2 are deterministic. For this problem, the *unconditional loglikelihood function* based on the data set $\mathbf{Z}$, see [3], is given by

$$\begin{aligned}\mathcal{L}(\mathbf{B},\mathbf{D},\mathbf{X},\theta,p_b,p_d,\sigma_u^2) &= \mathcal{M}_l(\mathbf{B},\mathbf{D},\mathbf{X},\theta,\sigma_u^2) \\ &+ \mathcal{M}_g(\mathbf{B},\mathbf{D},p_b,p_d),\end{aligned} \tag{4}$$

where

$$\begin{aligned}\mathcal{M}_l = &- \sum_{k=1}^{K} \ln\det R(k) \\ &- \sum_{k=1}^{K} \| \mathbf{z}(k) - \mathbf{a}(\theta)x(k) \|^2_{R^{-1}(k)}\end{aligned} \tag{5}$$

depends explicitly on the local realizations of the involved space/time processes, and

$$\begin{aligned}\mathcal{M}_g &= \bar{d}\ln p_d + [\bar{b}N - \bar{d}]\ln(1-p_d) \\ &+ \bar{b}\ln p_b + [K - \bar{b}]\ln(1-p_b)\end{aligned} \tag{6}$$

depends on global parameters. In (5),

$$R(k) = b(k)D(k)\sigma_u^2 + R_w \tag{7}$$

is the time variant covariance matrix of the total noise field and, in (6),

$$\bar{b} = \sum_{k=1}^{K} b(k) \ \text{ and } \ \bar{d} = \sum_{k=1}^{K}\sum_{n=1}^{N} d_n(k) \tag{8}$$

are the number of array samples disturbed by spiky impulses and the total number of realizations of the impulsive noise field, respectively. Notice that (5) and (6) do not correspond to a perfect decoupling of (4) into local and global variables. In fact, $\bar{b}$ and $\bar{d}$ depend on the particular realizations of the Bernoulli sequences to be estimated. Nevertheless, the computational complexity of this estimation problem can be effectively attenuated if that decoupling is considered. This is the strategy followed for the solution presented in the paper.

2.1 Robust Initialization

To initialize the estimation/detection algorithm that maximizes (4), we use the l_1 beamformer. For a given θ, the output of the l_1 beamformer is

$$\hat{x}_{l_1}(k,\theta) = \frac{\mathbf{a}^H(\theta)G^{-1}(k)\mathbf{z}(k)}{\mathbf{a}^H(\theta)G^{-1}(k)\mathbf{a}(\theta)}, \tag{9}$$

where $G(k) = \mathrm{diag}[|r_n(k|k-1)|]$, $r_n(k|k-1) = z_n(k) - a_n(\theta)\hat{x}_{l_1}(k|k-1)$ and $\hat{x}_{l_1}(k|k-1)$ is the predicted estimate of $x(k)$. Here, we assume that $\hat{x}_{l_1}(k|k-1) = \hat{x}_{l_1}(k-1)$. We can see that, whenever large impulses occur, the corresponding observations in (9) are strongly attenuated. As a consequence, the estimate $\hat{x}_{l_1}(k,\theta)$ is almost free of noisy spikes. The estimate of the arrival angle $\hat{\theta}_0$ is initialized by the value of θ that maximizes the sample covariance of $\hat{x}_{l_1}(k,\theta)$. Notice that, unless they are well separated, the l_1 beamformer will not resolve all the sources that are present. In this initialization step, if more than one source is detected, the residues are modified accordingly. In any case, (7) is assumed to be free of the effect of directional interferences, i.e., $R_i = 0$.

To detect the spiky impulses, which is equivalent to estimate the Bernoulli sequences, we maximize (5) for each time instant $k = 1, 2, \ldots, K$. Since $\sigma_u^2 \gg \sigma_0^2$, we assume that the largest residues are likely those that evidence the presence of impulses. Under these assumptions, (5) is simplified to

$$\begin{aligned}\mathcal{M}_l^k(M(k),\sigma_u^2) &= \mathcal{M}_l^k(0,\cdot) + M(k)\ln\left(\frac{\sigma_0^2}{\sigma_0^2+\sigma_u^2}\right) \\ &+ \frac{\sigma_u^2\sum_{n=1}^{M(k)}|r_n(k)|^2}{\sigma_0^2(\sigma_0^2+\sigma_u^2)},\end{aligned} \tag{10}$$

where $M(k)$ is the number of impulses at time k and the residues are supposed to be ordered decreasingly. In (10),

$$\mathcal{M}_l^k(0,\cdot) = -N\ln\sigma_0^2 - \frac{\sum_{n=1}^{N}|r_n(k)|^2}{\sigma_0^2} \tag{11}$$

is the value of $\mathcal{M}_l^k(M(k),\sigma_u^2)$ when $M(k) = 0$. Differentiating (10) with respect to σ_u^2 and equating to zero, we can obtain a local estimate of this parameter which, when used in (10), yields

$$\begin{aligned}\mathcal{M}_l^k(M(k)) &= \mathcal{M}_l^k(0,\cdot) - \ln\left(\frac{\sum_{n=1}^{M(k)}|r_n(k)|^2}{M(k)\sigma_0^2}\right) \\ &+ M(k)\left(\frac{\sum_{n=1}^{M(k)}|r_n(k)|^2}{M(k)\sigma_0^2} - 1\right).\end{aligned} \tag{12}$$

The maximization of this function results from a searching scheme on the possible values of $M(k)$, i.e., $n = 1, 2, \ldots, N$. If $\widehat{M}(k) = 0$, then $\hat{b}(k) = 0$; if $\widehat{M}(k) \neq 0$, then $\hat{b}(k) = 1$ and $\hat{d}_n(k) = 1$ for n corresponding to the largest $\widehat{M}(k)$ residues. Using the values of $\widehat{M}(k), k = 1, 2, \ldots, K$, in (5) and maximizing it with respect to σ_u^2, we obtain

$$\widehat{\sigma_u^2} + \sigma_0^2 = \frac{\sum_{k=1}^{K} \sum_{n=1}^{\widehat{M}(k)} |r_n(k)|^2}{\widehat{\hat{d}}}, \tag{13}$$

where

$$\widehat{\hat{d}} = \sum_{k=1}^{K} \widehat{M}(k) \text{ and } \widehat{\hat{b}} = \sum_{k=1}^{K} \hat{b}(k). \tag{14}$$

To complete the initialization step, we need to estimate the probability parameters that specify the Bernoulli sequences. This is done maximizing (6) with respect to p_b and p_d:

$$\begin{aligned} \hat{p}_b &= \widehat{\hat{b}}/K \\ \hat{p}_d &= \widehat{\hat{d}}/\widehat{\hat{b}}N. \end{aligned} \tag{15}$$

At this point, previous estimates of the statistics of the impulsive noise field are available, enabling the design of an optimal impulse detector and of a model based beamformer for angle of arrival estimation.

3 Description of the Algorithm

In this section, we present the two basic steps involved in each iteration of the algorithm: the impulsive noise detector and the angle of arrival estimator.

3.1 Optimum Impulse Detector

Here, the objective is to detect Gaussian impulses with known variance and probability of occurrence in independent white Gaussian noise, given the residues $r_n(k), = 1, 2, \ldots, N$. Making $\hat{p} = \hat{p}_b \hat{p}_d$, the minimum probability of error impulse detector is specified by the optimal threshold

$$\gamma = \frac{\sigma_0^2(\sigma_u^2 + \sigma_0^2)}{\sigma_u^2} \ln\left(\frac{\sigma_u^2 + \sigma_0^2}{\sigma_0^2} \frac{1 - \hat{p}}{\hat{p}}\right). \tag{16}$$

The estimates of the Bernoulli sequences are updated according to the following rule:

$$\begin{aligned} &\text{make } \hat{d}_n(k) = 1 \quad \text{if} \quad |r_n(k)|^2 > \gamma, \\ &\text{make } \hat{d}_n(k) = 0 \quad \text{if} \quad |r_n(k)|^2 < \gamma, \end{aligned}$$

Using the results obtained with this optimal detector, we can compute new estimates of the parameters that specify the impulsive noise field. These are given by (13), (14) and (15), where $\widehat{M}(k) = \sum_{n=1}^{N} \hat{d}_n(k)$, and $\hat{b}(k) = 0$ if $\widehat{M}(k) = 0$ or $\hat{b}(k) = 1$ if $\widehat{M}(k) \neq 0$.

3.2 Optimum Beamformer

The design of the optimum beamformer for bearing estimation relies on the estimates of the Bernoulli sequences $\widehat{\mathbf{B}}$ and $\widehat{\mathbf{D}}$ obtained with the optimum detector described in the previous subsection. We notice that only the second term of the right hand side of (5) depends on both θ and $\mathbf{X}$. Differentiating this function with respect to $x(k)$ and equating to zero, we obtain

$$\hat{x}(k, \theta) = \frac{\mathbf{a}^H(\theta) R^{-1}(k) \mathbf{z}(k)}{\mathbf{a}^H(\theta) R^{-1}(k) \mathbf{a}(\theta)}, \tag{17}$$

which is equivalent to

$$\hat{x}(k, \theta) = \frac{\mathbf{a}^H(\theta) R_z^{-1}(k) \mathbf{z}(k)}{\mathbf{a}^H(\theta) R_z^{-1}(k) \mathbf{a}(\theta)}, \tag{18}$$

where, according to (1) and (7),

$$\begin{aligned} R_z(k) &= \mathbf{a}(\theta) S \mathbf{a}^H(\theta) + R_w(k) + b(k) D(k) \sigma_u^2 \\ &= \mathbf{a}(\theta) S \mathbf{a}^H(\theta) + R(k) \end{aligned} \tag{19}$$

is the covariance matrix of the observations. In this formula we can divise a time variant term, due to the presence of the impulsive noise, and an invariant term

$$R_z^i = \mathbf{a}(\theta) S \mathbf{a}^H(\theta) + R_w(k), \tag{20}$$

where S is the signal power. The former term can be computed using the estimates given by the optimum detector and $\widehat{\sigma_u^2}$, and the latter is estimated by the sample covariance matrix of the observations vector taken in the time instants when impulses do not occur, i.e.,

$$\widehat{R_z^i} = \frac{1}{K - \widehat{\hat{b}}} \sum_{k : \hat{b}(k) = 0} \mathbf{z}(k) \mathbf{z}^H(k). \tag{21}$$

To find the angles of arrival, we scan for the values of θ that maximize the power of the beamformer output (18):

$$P(\theta) = \sum_{k=1}^{K} \| \hat{x}(k, \theta) \|^2, \tag{22}$$

where

$$\widehat{R_z}(k) = \widehat{R_z^i} + \hat{b}(k) \hat{D}(k) \widehat{\sigma_u^2} \tag{23}$$

is used for $R_z(k)$.

The detection and estimation schemes described in this section are used iteratively until some prespecified condition, involving the estimated parameters, is reached.

4 Simulations and Discussion

In the computer simulations here reported, we considered one directional waveform, with power $S = 1$, impinging from broadside, i.e., $\theta = 0°$, an uniform and linear array of $N = 15$ sensors. Tables 1 and 2 show the results obtained for two distinct experiments, where $K = 50$ time samples were used. The first experiment evidences the adequacy of the l_1 beamformer for initialization purposes. In fact, the results obtained at the end of the 1st iteration, specially the bearing estimate ($\hat{\theta} = -1°$ was obtained with the MVDR beamformer), are quite similar to the actual values of the parameters of interest. Those estimates are clearly improved in the 2nd iteration, where the optimum detector and the optimum beamformer were used. Figures 1 and 2 show the detection errors of the space/time indexed Bernoulli sequences obtained in each iteration of this experiment, evidencing the small number of errors provided by the l_1 beamformer (Fig.1) and the capability of the optimum detector (Fig.2) to correct them. For an expected value of 30 space/time occurrences of noisy impulses, 10% of detection errors were achieved in this run of the experiment. These results are confirmed by the second experiment. In this case, worst impulsive noise conditions were imposed. Nevertheless, accurate results were obtained with only two iterations. Notice that, in this experiment, the biased estimate of the bearing angle provided by the l_1 beamformer ($\hat{\theta} = -3°$ was obtained with the MVDR beamformer) was corrected by the optimum beamformer.

param.	σ_0^2	σ_u^2	p_b	p_d	θ
values	.5	100	.2	.2	0°
1st it.	•	92.74	.12	.28	0°
2nd it.	•	91.10	.18	.21	0°

Table 1: 1st experiment.

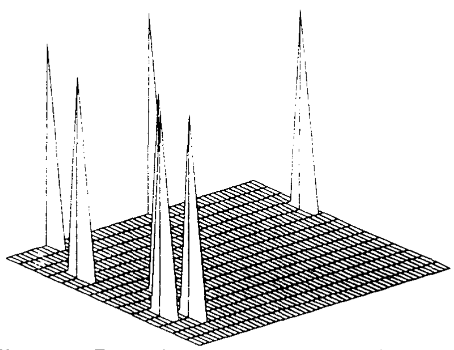

Figure 1: Detection errors - 1st iteration.

These simulations confirm the efficiency of the algorithm proposed in the paper in relation to both **(i)** computational complexity and **(ii)** quality of the resulting estimates. The former is a consequence of the initialization provided by the l_1 approach, and the latter relies on the optimality of the impulse detector.

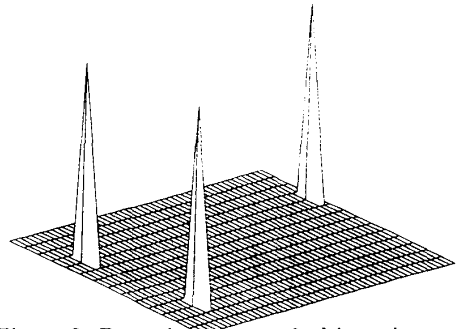

Figure 2: Detection errors - 2nd iteration.

param.	σ_0^2	σ_u^2	p_b	p_d	θ
values	1	100	.8	.6	0°
1st it.	•	•	•	•	$-1°$
2nd it.	•	104.76	.9	.54	0°

Table 2: 2nd experiment.

References

[1] Victor A.N. Barroso and José M.F. Moura. l_2 and l_1 beamformers: recursive implementation and performance analysis. *Submitted for publication*, 1990.

[2] Victor A.N. Barroso and José M.F. Moura. Detection performance of the l_1 beamformer in the presence of underwater burst noise. In *International Conference on Acoustics, Speech and Signal Processing, ICASSP90*, IEEE, Albuquerque, 1990.

[3] J.M. Mendel. Maximum likelihood deconvolution: a journey into model based signal processing. *Springer-Verlag*, 1990.

[4] Victor A.N. Barroso and José M.F. Moura. Maximum likelihood beamforming in the presence of outliers. In *International Conference on Acoustics, Speech and Signal Processing, ICASSP91*, IEEE, Toronto, 1991.

[5] J.M. Passerieux. Bearings only tracking with non Gaussian measurements errors. In *International Conference on Acoustics, Speech and Signal Processing, ICASSP91*, IEEE, Toronto, 1991.