Maximum Likelihood Beamforming in the Presence of Outliers

Victor A. N. Barroso

CAPS, Dep. Eng. Electr. e Comp.
Instituto Superior Técnico
1096 Lisboa Codex, Portugal

José M. F. Moura

LASIP, Depart. Elect. and Comp. Eng.
Carnegie Mellon University
Pittsburgh, PA, 15213

Abstract

We consider the problem of maximum likelihood beamforming in the presence of outliers. In practice, outliers occur due to malfunctioning of sensors or as a consequence of strong impulsive noise. The performance of beamformers based on maximum likelihood or minimum mean square error type criteria is seriously degraded by outliers, see [1]. One solution to combat this would be to optimally detect the failed sensors and the the presence of impulses. The complexity of this direct solution increases exponentially with the number of array sensors and time samples. In this paper, we purpose an alternative method that models outliers as impulsive noise and detects impulses by using the residues of the l_1 beamformer introduced in [1]. We develop this technique and discuss its efficiency.

1. Introduction

In this paper we are concerned with the design of the optimum beamformer for the case where impulsive noise is present. In practical situations, this may represent the existence of malfunctioning sensors or the occurrence of spikes. Our approach consists on the maximization of the likelihood of the data set. This is given by K independent time samples of vector observations collected by an array of N sensors. The impulsive noise is modeled as a complex space/time Bernoulli-Gauss sequence which is defined as the product of a Bernoulli sequence by a complex Gaussian white sequence. Clearly, the maximization of the likelihood function involves the estimation of the Bernoulli sequence which provides information about the sensors and time instants at which the impulses have occurred. Assuming that all the data parameters are known, e.g., the direction of arrival, the solution of that estimation problem can be obtained by picking, among the 2^{KN} possible Bernoulli sequences of size KN, that one that maximizes the data likelihood. The computational time involved in the searching procedure increases exponentially with KN; hence, this is prohibitive for practical purposes. Detection algorithms provide us an approach that do not suffer from the drawbacks of the direct method. Here, we present a detection algorithm based on the properties of the l_1 beamformer [2]. It has been shown [1] that the l_1 beamformer has the capability of adjusting itself to unexpected noise conditions. In particular, the unexpected impulses are strongly attenuated at the beamformer's output. Hence, if an impulse occurs at any array sensor, the amplitude of the corresponding residue depends mainly on the magnitude of the impulse. On the contrary, if the impulse is not present, the residue is determined by the background noise. In general, the impulses are much stronger than the background noise. These facts suggest that an hypothesis test based on the residues at the output of the l_1 beamformer can be used to design an efficient sequential detection algorithm. In order to implement this algorithm, the estimate of the direction of arrival has to be initialized. The l_1 approach described in [3] provides a robust method for doing that. Using this initial estimate of the direction of arrival, we use our sequential detection algorithm to estimate the Bernoulli sequence. Finally, the estimate of the angle of arrival can be updated by searching for the angle that maximizes the likelihood of the data conditioned on the estimated discrete events sequence.

The paper is organized as follows. In section 2, we formulate the problem and derive the data likelihood function. In section 3, the sequential detection algorithm for estimating the space/time Bernoulli sequence is presented. Estimation of other parameters characterizing the impulsive noise, such as the probability of occurrence of impulses and their power, is also discussed. In section 4, we present the structure of the resulting maximum likelihood beamformer which is used to update the estimate of the angle of arrival.

2. Problem Formulation

Let

$$z(k) = a(\theta)x(k) + s(k) + w(k) \qquad (1)$$

be the complex envelope of the $(N \times 1)$ vector of observations at time k, where $x(\cdot)$ and $w(\cdot)$ are the complex envelopes of the desired signal and of the background noise, respectively. The latter is assumed to be complex Gaussian with zero mean and known covariance matrix $R_w = \sigma^2 I$ and temporally white. The complex $(N \times 1)$ steering vector $a(\theta)$ is specified by the direction of arrival θ. In (1), $s(\cdot)$ is the impulsive noise vector field. At each sensor, the impulsive noise sample $s_n(\cdot)$ 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 space/time indexed zero mean Gaussian r.v. $u_n(\cdot)$ with variance σ_u^2. Formally,

$$s(k) = b(k)D(k)u(k), \qquad (2)$$

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

For the presented model, the likelihood function of the data set $Z_K = \{z(k)\}_{k=1}^K$ conditioned on the unknown parameters is given by

$$\begin{aligned}\mathcal{L}(X_K, B_K, D_K, \theta, p_b, p_d, \sigma_u^2) = &-\sum_{k=1}^{K} \ln\det R(k) \\ &- \sum_{k=1}^{K} (z(k) - a(\theta)x(k))^H R^{-1}(k)(z(k) - a(\theta)x(k)) \\ &+ \mathcal{M}(B_K, D_K, p_b, p_d), \qquad (3)\end{aligned}$$

where $R(k) = b(k)D(k)\sigma_u^2 + \sigma^2 I$, $X_K = \{x(k)\}_{k=1}^K$, $B_K = \{b(k)\}_{k=1}^K$, $D_K = \{D(k)\}_{k=1}^K$,

$$\begin{aligned}\mathcal{M}(B_K, D_K, p_b, p_d) &= \bar{d}\ln p_d + [KN - \bar{d}]\ln(1 - p_d) \\ &+ \bar{b}\ln p_b + [K - \bar{b}]\ln(1 - p_b) \qquad (4)\end{aligned}$$

and

$$\bar{b} = \sum_{k=1}^{K} b(k), \qquad (5)$$

$$\bar{d} = \sum_{k=1}^{K}\sum_{n=1}^{N} d_n(k). \qquad (6)$$

The maximization of the likelihood function $\mathcal{L}$ is a complicated problem. Its complexity is associated with the large number of parameters to be estimated and with the discrete nature of B_K and D_K. However, these difficulties can be attenuated if we recursively use detection and block optimization algorithms [4]. Notice that, if initial guesses of θ, X_K, B_K and D_K are available, then maximization of $\mathcal{M}$ provides estimates of the a priori probabilities

$$\hat{p}_b = \bar{b}/K \qquad (7)$$

$$\hat{p}_d = \bar{d}/KN. \qquad (8)$$

On the other hand, maximization of the remaining terms of $\mathcal{L}$ yields $\widehat{\sigma_u^2}$. These can then be used to construct a maximum a posteriori detector in order to update the estimates of B_K and D_K. If the initial guesses are reliable, i.e., if they are close to their actual values, then the latter estimates are the best that we can obtain from the available data. If not, we have to continue with this iterative procedure until convergence is achieved. Hence, reliability of the initial estimates of θ, X_K, B_K and D_K is of utmost importance to speed up the convergence of the iterative block search method. Our point is that, for the problem that we are considering, the l_1 approach provides solutions whose properties match those we are looking for. This will be discussed in the following section.

3. l_1 Sequential Detector

In this section we derive a sequential detector which is based on the values of the residues at the output of the l_1 beamformer [1]. At each time instant, this detector uses all the array data to estimate $b(k)$ and $d_n(k), n = 1,2,\ldots,N$. Clearly, if $\hat{d}_n(k) = 0, n = 1,2,\ldots,N$, then $\hat{b}(k) = 0$. For a given θ, let the output of the l_1 beamformer be

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

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, assuming that $x(\cdot)$ is slowly time varying, we make $\hat{x}_{l_1}(k|k-1) = \hat{x}_{l_1}(k-1)$. From (9), we see that those sensors responsible for large residues are practically discarded in the processing. This happens whenever large impulses occur. Hence, $\hat{x}_{l_1}(k,\theta)$ is free of noisy spikes. To initialize the estimate of θ, we search for the value $\hat{\theta}_0$ that maximizes the sample covariance of $\hat{x}_{l_1}(k,\theta)$, i.e., $\sum_{k=1}^{K} |\hat{x}_{l_1}(k,\theta)|^2$. As it was shown in [3], this method gives an accurate estimate of the angle of arrival. Let

$$\hat{x}_0(k) = \hat{x}_{l_1}(k,\hat{\theta}_0) \tag{10}$$

and compute the residues

$$r_n(k) = z_n(k) - a_n(\hat{\theta}_0)\hat{x}_0(k), n = 1,2,\ldots,N. \tag{11}$$

In general, there will exist a number $M = 0,1,\ldots,N$ of spiky residues. These can be approximately modeled as zero mean complex r.v. with variance $\sigma_u^2 + \sigma^2$. The remaining $N - M$ residues are then zero mean complex r.v. with variance σ^2. In the case of interest, $\sigma_u^2 \gg \sigma^2$, the M spiky sensors are likely those that are responsible for the M largest residues. Suppose that the set $\{|r_n(k)|\}_{n=1}^{N}$ is ordered decreasingly. Then, under our assumptions, the likelihood of the residues set is approximately given by

$$\begin{aligned}\mathcal{L}(M,\sigma_u^2) &= -M\ln(\sigma_u^2+\sigma^2) - \sum_{n=1}^{M}\frac{|r_n(k)|^2}{\sigma_u^2+\sigma^2} \\ &+ M\ln(\sigma^2) - \sum_{n=1}^{M}\frac{|r_n(k)|^2}{\sigma_u^2}.\end{aligned} \tag{12}$$

Differentiating with respect to σ_u^2 and equating to zero, we obtain

$$\widehat{\sigma_u^2}(k) + \sigma^2 = \frac{1}{M}\sum_{n=1}^{M}|r_n(k)|^2 \tag{13}$$

which, when back substituted in (12), yields

$$\begin{aligned}\mathcal{L}(M) &= M\ln\left(\frac{M\sigma^2}{\sum_{n=1}^{M}|r_n(k)|^2}\right) \\ &- M - \sum_{n=M+1}^{N}\frac{|r_n(k)|^2}{\sigma^2}.\end{aligned} \tag{14}$$

A search method can now be used to maximize $\mathcal{L}(M)$ at time k. 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.

The results obtained with this time sequential detector can now be used to estimate the variance of the impulses. Using (11), the maximization of (3) yields

$$\widehat{\sigma_u^2} + \sigma^2 = \frac{\sum_{k=1}^{K}\sum_{n=1}^{\widehat{M}(k)}|r_n(k)|^2}{\widehat{\hat{d}}}, \tag{15}$$

where

$$\widehat{\hat{d}} = \sum_{k=1}^{K}\widehat{M}(k). \tag{16}$$

Defining

$$\widehat{\hat{b}} = \sum_{k=1}^{K}\hat{b}(k), \tag{17}$$

we can estimate the a priori probabilities using (17) and (16) in (7) and (8), respectively. With $\hat{p} = \hat{p}_b\hat{p}_d$ and $\widehat{\sigma_u^2}$ we can define a minimum probability of error impulse detector. Being

$$\gamma = \frac{\sigma^2(\widehat{\sigma_u^2}+\sigma^2)}{\widehat{\sigma_u^2}}\ln\left(\frac{\widehat{\sigma_u^2}+\sigma^2}{\sigma^2}\frac{1-\hat{p}}{\hat{p}}\right) \tag{18}$$

the optimum threshold for this detector, the estimate of the space/time Bernoulli sequence is updated according to:

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

Notice that, depending on the values of $\hat{p}$, $\widehat{\sigma_u^2}$ and σ^2, the optimum threshold γ can take negative values. This happens when, for large values of the ratio

$\widehat{\sigma_u^2}/\sigma^2$, $\hat{p}$ is close to one. In this case, the total noise field is approximately Gaussian with variance $\widehat{\sigma_u^2} + \sigma^2$ and the optimum detector decides for the occurrence of an impulse with probability one. Thus, $\widehat{M}(k) = N$ for all k, $\widehat{d} = NK$, and (15) gives an estimate of the variance of the total noise field. As it was said before, the l_1 approach provides reliable estimates of the signal and of its direction of arrival, the possible input impulses being accurately replicated in the residues. Hence, the l_1 sequential detector provides an efficient initialization of the space/time Bernoulli sequence, enabling the computation of accurate estimates of the remaining parameters characterizing the impulsive noise, which can then be used to perform a one step "optimum" updating of the discrete events sequence. Once this is done, we have available estimates of all the parameters necessary to the design of the maximum likelihood beamformer for this model. This is used in a search procedure yielding the maximum likelihood estimate of the direction of arrival.

4. Maximum Likelihood Beamformer

We observe that only the second term of the right hand side of equation (3) depends on the direction of arrival θ and on the signal sequence X_K. Hence, to estimate those quantities we have to minimize

$$\mathcal{L}(X_K, \theta) = \sum_{k=1}^{K} \left[(z(k) - a(\theta)x(k))^H \widehat{R}^{-1}(k) \right. \left. (z(k) - a(\theta)x(k)) \right], \tag{19}$$

where

$$\widehat{R}(k) = \hat{b}(k)\widehat{D}(k)\widehat{\sigma_u^2} + \sigma^2 I. \tag{20}$$

Differentiating (19) with respect to $x(k)$ and equating to zero, we get

$$\hat{x}(k,\theta) = \frac{a^H(\theta)\widehat{R}^{-1}z(k)}{a^H(\theta)\widehat{R}^{-1}a(\theta)}. \tag{21}$$

Notice that the l_1 beamformer (9) behaves much like the optimum beamformer. In both cases, the data at those sensors where strong impulses occur is strongly attenuated. Given the direction of arrival θ, this equation specifies the output of the maximum likelihood beamformer for the case where impulsive noise plus Gaussian background noise is present. Substitution of this result in (19) yields

$$\mathcal{L}(\theta) = \sum_{k=1}^{K} \left[z^H(k)\widehat{R}^{-1}(k)z(k) - |\hat{x}(k,\theta)|^2 a^H(\theta)\widehat{R}^{-1}(k)a(\theta) \right]. \tag{22}$$

The maximum likelihood estimate of θ is then

$$\hat{\theta} = \arg\max_{\theta} \sum_{k=1}^{K} |\hat{x}(k,\theta)|^2 a^H(\theta)\widehat{R}^{-1}(k)a(\theta). \tag{23}$$

5. Conclusions

In this paper, optimum beamforming design in the presence of strong impulsive noise was addressed. It was shown that the resulting structure enables computation of the maximum likelihood estimates of the signal direction of arrival, provided that accurate previous estimates of the space/time impulsive events are available. A technique that initializes those estimates was derived. This is based on the l_1 norm and, as discussed in the paper, its efficiency speeds up the convergence of the block search method used to maximize the likelihood function.

References

[1] V. A. N. Barroso. l_2 and l_1 beamformers: recursive implementation and performance analysis. *submitted for publication*, December 1990.

[2] V. A. N. Barroso and J. M. F. Moura. Adaptive beamforming as an inverse problem. In *Internacional Conference on Acoustics, Speech and Signal Processing, ICASSP89*, IEEE, Glasgow, 1989.

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

[4] J. M. Mendel. *Maximum-Likelihood Deconvolution. A Journey into Model-Based Signal Processing.* Springer-Verlag, 1990.