Speech Enhancement With Robust Beamforming for Spatially Overlapped and Distributed Sources

Xiong Wenmeng, Bao Changchun, Jia Maoshen, José Picheral

Speech enhancement with robust Beamforming for spatially overlapped and distributed sources

XIONG Wenmeng, BAO Changchun, Senior Member, IEEE, JIA Maoshen, Senior Member, IEEE, José PICHERAL

***Abstract*—Most of the existing Beamforming methods are based on the assumptions that the sources are all point sources and the angular separation between the direction of arrival (DOA) of the source and the interference is large enough to assure good performance. In this paper, we consider a tough scenario where the target source and the interference are simultaneously spatially distributed and overlapped. To improve the performance of Beamforming in this scenario, we propose two approaches: the first approach exploits the non-Gaussianity as well as the spectrogram sparsity of the output of the microphone array; the second approach exploits the generalized sparsity with overlapped groups of the Beampattern. The proposed criteria are solved by methods based on linearized preconditioned alternating direction method of multipliers (LPADMM) with high accuracy and high computational efficiency. Numerical simulations and real data experiments show the advantages of the proposed approaches compared to previously proposed Beamforming methods for signal enhancement.**

***Index Terms*—speech enhancement, microphone array, MVDR, DOA**

I. Introduction

Beamforming [1] is a famous signal enhancing technique which has been widely used in speech enhancement, acoustic imaging, wireless communications, and radar processing, etc. Early Beamformings are called "fixed" Beamforming, including delay-and-sum Beamforming, which are independent of the received signals at the microphone array and have difficulty in the interference cancellation. The adaptive Beamforming algorithms are proposed latter and are more flexible, the coefficients of the Beamformers can be changed according to the received signals, in order to conserve the signal of interest while reducing the noise and the interference. Among the adaptive Beamforming algorithms, the MVDR Beamforming [2] with one linear constraint is famous due to its ease of implementation. The MVDR Beamforming minimizes the power of the output signals or the power of the residual noise of the microphone array[3], the latter formulation can achieve a better output Signal to Interference plus Noise Ratio (SINR) when the Input Signal to Noise Ratio (SNR) is high, but the reconstructed interference plus noise covariance matrix is difficult to obtain. The Linearly Constrained Minimum Variance (LCMV) Beamforming is an extension of the MVDR Beamforming with multiple linear constraints which can not only keep the target signals but also cancel the interferences.

The conventional MVDR and LCMV Beamforming techniques suffers from two drawbacks: firstly, the performance is sensitive to the steering vector mismatch, because the linear constraints in the Beamforming criteria depend on the steering vector of the sources or interferences; secondly, the objective functions are mostly based on the second order statistics of the array output, which are optimal for Gaussian signals and noise, but suboptimal for the most real world signals which are not Gaussian.

Many methods have been developed to solve the first problem: the subspace based Beamforming [4] which has been proved to be robust to various types of steering vector mismatches, but the performance degrades when the SNR is low. The other important class of methods to improve the robustness of Beamforming concerns relaxing the linear constraints to nonlinear constraints and the criteria are solved by convex optimization methods, such as the diagonal loading method [5] and the worst-case performance optimization method [6]. More recently, in [7] a new Beamforming formulation has been proposed with three constraints: a relaxed double-sided norm constraint, an additional similarity constraint, and a constraint enforcing the desired signal DOA to be far away from the DOA interval of all linear combinations of the interference steering vectors.

To solve the problem of non Gaussian source, various objective functions have been proposed: in [8] and [9], the l_p norm of the array output as the objective function is proposed, better performance is achieved with $p \leq 2$ for super-Gaussian signals and $p \geq 2$ for sub-Gaussian signals. In [10] the Gaussian entropy of the array output is minimized for the reception of second order non-circular/circular signal of interest. In [11] the data is transformed to the probability distribution for the Beamforming in the presence of non-Gaussian impulsive noise. In [12] the third-order Volterra MVDR Beamforming for non-Gaussian and non-circular interference cancellation is proposed. In [13] a unit-modulus constraint is imposed on the spatial filter and only the phase shifting at each sensor is considered to reduce the computational burden. Note that in some literature such as [14], the nonlinear constraint scenario and non-Gaussian signal scenario are considered jointly for new Beamformers.

Most of Beamforming techniques mentioned above as well as other array processing techniques [15] are developed in the scenario of point source. However, for many applications, the source can no longer be considered as point source and

This work was supported by the National Natural Science Foundation of China (Grant No.61831019, No.61971015, and No.62101013).

XIONG Wenmeng, Bao Changchun (corresponding author) and JIA Maoshen are with the Faculty of Information Technology, Beijing University of Technology, Beijing 100124, China (e-mail: wenmeng.xiong@bjut.edu.cn; chchbao@bjut.edu.cn; jiamaoshen@bjut.edu.cn)

PICHERAL José is with Laboratoire des signaux et systèmes (Signal and System Laboratory), Université Paris-Saclay, CNRS, CentraleSupélec, 91190, Gif-sur-Yvette, France (email: Jose.Picheral@centralesupelec.fr)

angular spreads should be taken into consideration. According to the statistical features of the emitted signal, the model of spatially distributed source can be classified into two types: incoherently distributed (ID) source and coherently distributed (CD) source [16]. On one hand, for ID source, signals coming from different points of the same distributed source can be considered uncorrelated. On the other hand, in the scenario of CD source, the emitted signal components are delayed and scaled replicas from different points of the same signal. Plenty methods have been investigated for the localization of ID and CD source such as in [16][17][18]. However, few works have focused on the signal enhancement problem in the scenario of distributed source. More recently, in [19] the correlation matrix of the received signals impinging from the scattered source is imposed in the linear constraint instead of the steering vector of the source and three Beamformers are proposed with the aim to maximize the white noise gain (WNG), the directivity factor (DF) and a compromise between the WNG and the DF, respectively. However, it is difficult to estimate accurately the correlation matrix of only the received signals on the microphone array in practice.

In this paper, we focus on the Beamforming technique for the application of speech enhancement in the scenario of overlapped distributed sources. The existing Beamforming techniques can easily deal with the cases that the overlapped distributed sources are all target sources or interferences, while the performance degrades severely when the target sources and the interferences are overlapped. To handle this problem, we propose two methods for both CD sources and ID sources: in the first approach we minimize jointly the l_1 norm and the entropy of the output of the array, for exploiting jointly the sparsity and the non-Gaussianity of the spectrogram of the speech signal; in the second approach we consider the generalized sparsity with overlapped groups of the Beampattern, for keeping the signal components impinging from the directions within the distributed sources and minimizing the signal components from other directions. We solve the proposed criteria by LPADMM [20] based methods instead of CVX toolbox, which are faster and provide more accurate solutions. The projection gradient descent method is used in each iteration of LPADMM. By introducing extra variables, the problem formulation of the projection gradient descent is transformed into a norm adaptation problem which is easy to solve. Note that as a point source can be considered as a special case of CD source when its angular spread becomes 0, the proposed algorithms can be used also in the point source scenario to improve the performance when the source and the interference are too close to each other.

This paper is organized as follows: in section II we briefly review the models of CD and ID sources, the conventional Beamforming technique, and the maximum negentropy criterion. In section III we propose two methods for Beamforming in the scenario of overlapped distributed signals. In section IV we present simulation results and real data experiments to show the advantages of the proposed criteria. Finally conclusions are given in V.

II. Signal model and LCMV Beamforming

A. CD source and ID source

Considering Q far-field wideband distributed sources impinging from the direction $\theta_i, i = 1, ..., Q$ on M microphones. The signals received on the microphone array at frequency ω can be given as:

$$\mathbf{y}(\omega) = \sum_{i=1}^{Q} \mathbf{x}_i(\omega) + \mathbf{n}(\omega), \tag{1}$$

$$\text{and} \qquad \mathbf{x}_i(\omega) = \int \mathbf{a}(\omega, \theta) p_i(\omega, \theta) d\theta, \tag{2}$$

where $\mathbf{x}_i$ is the signal received by the microphone array from source i, $p_i(\omega, \theta)$ is the signal component from the DOA θ in the distributed source i, $\mathbf{n}(\omega)$ is the noise which is assumed to be uncorrelated, $\mathbf{a}(\omega, \theta)$ is the steering vector of a point source from direction θ such that $\mathbf{a}(\omega, \theta) = \left[1, e^{-j\omega\Delta t_1(\theta)}, ..., e^{-j\omega\Delta t_{M-1}(\theta)}\right] \in \mathbb{C}^{M\times 1}$, when the first microphone is taken as the reference, $\Delta t_m(\theta)$ is the time delay from the m^{th} microphone to the reference microphone.

The distributed source can be classified into CD source and ID source according to its statistical features of the distribution [16]. If the i^{th} source is CD, $p_i(\omega, \theta)$ can be given as:

$$p_i(\omega, \theta) = h_i(\omega, \theta) s_i(\omega), \tag{3}$$

where $s_i(\omega)$ is the signal emitted by the i^{th} source and independent of θ. Using (2) and (3), the signal $\mathbf{x}_i$ can be given as:

$$\mathbf{x}_i(\omega) = \int \mathbf{a}_i(\omega, \theta) h_i(\omega, \theta) d\theta s_i(\omega), \tag{4}$$

where $h_i(\omega, \theta)$ is the spatial distribution function of the source i. In the following, we note $\mathbf{c}_i(\omega, h_i) = \int \mathbf{a}_i(\omega, \theta) h_i(\omega, \theta) d\theta \in \mathbb{C}^{M\times 1}$ as the steering vector for the CD source. The covariance matrix of the CD source received on the microphone array can be given as:

$$\mathbf{R}_{x_i} = E\left[\mathbf{x}_i(\omega)\mathbf{x}_i^H(\omega)\right] = \mathbf{c}_i(\omega, h_i)\mathbf{c}_i^H(\omega, h_i)\sigma_s^2, \tag{5}$$

where $\cdot^H$ is the conjugate transposition operator, $E[\cdot]$ is the expectation operator, σ_s^2 is the power of the signal.

If the i^{th} source is ID, the model (4) is no longer valid, and the covariance matrix of a ID source received on the microphone array can be given as [16]:

$$\begin{aligned}\mathbf{R}_{x_i} &= \int\int \rho_i(\omega, \theta, \theta') \mathbf{a}_i(\omega, \theta) \mathbf{a}_i^H(\omega, \theta') d\theta d\theta' \\ &= \sigma_s^2 \int \mathbf{a}_i(\omega, \theta) \mathbf{a}_i^H(\omega, \theta) d\theta,\end{aligned} \tag{6}$$

where $\rho_i(\omega, \theta, \theta')$ is the signal angular cross correlation kernel such that:

$$\rho_i(\omega, \theta, \theta') = E\left[p_i(\omega, \theta) p_i^\star(\omega, \theta')\right] = \begin{cases} \sigma_s^2, & \text{if } \theta = \theta'. \\ 0, & \text{otherwise,} \end{cases} \tag{7}$$

where $\cdot^\star$ is the conjugate operator.

Note that as stated by Valaee [16], for a purely CD source, the rank of the covariance matrix $\mathbf{R}_{x_i}$ is 1 from (5); while for a purely ID source, the rank of $\mathbf{R}_{x_i}$ is M from (6). In the case of partially coherently distributed source the rank of $\mathbf{R}_{x_i}$ is between 1 and M.

B. The LCMV Beamforming

The conventional problem formulation of the LCMV Beamforming [1] for point source is as follows:

$$\min_{\mathbf{w}(\omega)} \mathbf{w}^H(\omega)\mathbf{R}_y\mathbf{w}(\omega) \qquad \text{subject to} \quad \mathbf{w}^H(\omega)\mathbf{a}_i(\omega, h_i) = \mathbf{e}_i, \quad i = 1, ..., Q, \tag{8}$$

where $\mathbf{w}(\omega)$ can be considered as an finite impulse response filter applied to each microphone output at angular frequency ω, h_i is the spatial distribution function of source i as defined in (3), if the i^{th} source is useful and needed to be enhanced, $\mathbf{e}_i = 1$, otherwise $\mathbf{e}_i = 0$, $\mathbf{R}_y$ is the covariance matrix of the measured signals received on the microphone array such that:

$$\mathbf{R}_y = E\left[\mathbf{y}(\omega)\mathbf{y}^H(\omega)\right]. \tag{9}$$

The output SINR is a noise reduction performance measure of Beamforming methods, which is defined as:

$$\text{SINR} = \frac{\mathbf{w}^H\mathbf{R}_x\mathbf{w}}{\mathbf{w}^H\mathbf{R}_{i+n}\mathbf{w}}, \tag{10}$$

where $\mathbf{R}_x = E[\mathbf{x}(\omega)\mathbf{x}^H(\omega)]$ is the covariance matrix of the target signal, and $\mathbf{R}_{i+n} = E[\mathbf{i}(\omega)\mathbf{i}^H(\omega) + \mathbf{n}(\omega)\mathbf{n}^H(\omega)]$ is the covariance matrix of interference plus noise, $\mathbf{i}(\omega)$ is the sum of interferences received on the microphone array.

C. Maximum negentropy criterion

According to the *central limit theorem*, the probability density function (PDF) of the sum of independent random variables of any PDFs converges to Gaussian. In addition, it is well known that the speech signals are always super-Gaussian distributed, while noise is more nearly Gaussian distributed. Based on this property, various source separation methods such as independent component analysis (ICA) aims to find a signal which is the least Gaussian.

The negentropy is a popular criterion for measuring the non-Gaussianity of a signal, for a signal B, the negentropy can be given as:

$$J(B) \triangleq S(B_{\text{Gauss}}) - S(B), \tag{11}$$

where $S(B) = -E\left[\log(p(B))\right]$ is the entropy of B, $p(B)$ is the PDF of B, $S(B_{\text{Gauss}})$ is the entropy of a Gaussian variable which has the same variance as B. As $S(B_{\text{Gauss}})$ is constant, the value of $J(B)$ depends on that of $S(B)$. A larger value of $J(B)$ indicates that the signal B is less Gaussian. Therefore, by maximizing $J(B)$ or by minimizing $S(B)$, a least Gaussian signal as the signal B can be found. However, the function $p(B)$ is always very hard to obtained in reality, which causes difficulties to obtain closed form expressions of $J(B)$. To solve this problem, one can maximize the contrast function $E[G(|B|^2)]$ instead of minimizing the entropy, where $|\cdot|$ is the absolute value operator [21], and $G(\cdot)$ is a smooth even function. The contrast function is computationally simple and its extrema coincides with that of $S(B)$. Several choices of $G(|B|^2)$ have been proposed in the previous literature [21] such as:

$$G_1(|B|^2) = \sqrt{b_1 + |B|^2}, \tag{12}$$

$$G_2(|B|^2) = \log(b_2 + |B|^2), \tag{13}$$

$$G_3(|B|^2) = \frac{1}{2}|B|^4, \tag{14}$$

where b_1 and b_2 are arbitrary constants, in our work, $b_1 \approx 0.1$ and $b_2 \approx 0.1$. Note that G_1 and G_2 grow more slowly than G_3 and they give estimators more robust against outliers.

The Beamforming algorithm reduces the noise by minimizing the power of the output of the microphone array while adding some constraints for preserving the signal components coming from the direction of the target source. However, when a target distributed source is partially or totally overlapped with a distributed interference, the constraint for preserving the target source signal components will preserve also part of the interference signal components. To handle this problem, we introduce the maximization of the negentropy into the Beamforming criterion to further separate the target source and the interference.

III. Beamforming for overlapped distributed sources

In this section, we study the Beamforming algorithm in the case where the distributed sources are overlapped. In our simulations, we have found that the LCMV works well when the overlapped sources are all target sources or all interferences. However, when the target sources are overlapped with the interferences, the performance of the LCMV algorithm degrades severely. Thus, in the following, for the sake of simplicity, we consider the case that one target distributed source and one distributed interference are overlapped. Criteria based on the sparsity property of the spectrogram of the output of the microphone array and the Beampattern are proposed and solved with LPADMM based methods. Note that the proposed criteria can be easily generalized into the case of multiple sources and multiple interferences and solved by multi-block ADMM methods [22].

A. Sparsity of the microphone array output spectrogram

1) Target CD source and CD interference: In this subsection, we discuss the scenario that both the target source and interference are CD. According to [8], the minimization of l_p-norm instead of the second order of the output of the microphone array can improve the performance significantly, where $0 < p < 2$ for the super-Gaussian distributed information bearing signals such as speech. Here, we take $p = 1$ for exploiting also the sparsity property of the spectrogram of the speech, and add the entropy function defined in (11) of the output of the microphone array for extracting the signal component which is the least Gaussian. We propose to define the Beamforming filter as the solution:

$$\begin{aligned} &\min_{\mathbf{w}} S(\mathbf{w}^H(\omega)\mathbf{y}_d(\omega)) + \lambda_1\|\mathbf{w}^H(\omega)\mathbf{Y}(\omega)\|_1 \\ &\text{s.t. } \mathbf{w}^H(\omega)\mathbf{w}(\omega) \le \delta_{\mathbf{w}}, \mathbf{w}^H(\omega)\mathbf{c}_1 = 1, \mathbf{w}^H(\omega)\mathbf{c}_2 = 0, \end{aligned} \tag{15}$$

where $\mathbf{Y}(\omega) \in \mathbb{C}^{M\times D}$ are the samples of the signals received at the microphone array at frequency band ω after the short time Fourier transform (STFT), $\mathbf{y}_d$ is the d^{th} column of the matrix $\mathbf{Y}(\omega)$, ω is omitted in the following as all the algorithms are studied in the same sub-frequency band, D is the number of samples at each sub-frequency band, λ_1 is the sparsity penalization weighting parameter, $\delta_{\mathbf{w}}$ is the constraint that we impose on the norm of the Beamforming filter, $\mathbf{c}_1$ and $\mathbf{c}_2$ are assumed to be the steering vectors corresponding to the target source and the interference, respectively. The problem (15) can be solved by the CVX toolbox directly. However, the CVX toolbox solves the entropy family functions such as (12)-(14) using an experimental successive approximation method which is slower and less reliable than the method employed for other problems. To solve this problem, we reformulate (15) with only linear constraints as the problem (16), where the entropy replaces the contrast function. Note that the problem (16) is non-convex and non-continuous, we solve this problem by the LPADMM which is faster and more accurate.

Introducing an extra variable $\mathbf{z}$, the problem (15) can be reformulated as:

$$\begin{aligned}\min_{\mathbf{w},\mathbf{z}} \; & -E\left[G(|\mathbf{w}^H\mathbf{y}_d|^2)\right] + \lambda_1\|\mathbf{z}\|_1 + \iota(\mathbf{w}) \\ \text{s.t. } & \mathbf{w}^H\mathbf{c}_1 = 1, \mathbf{w}^H\mathbf{c}_2 = 0, \mathbf{Y}^H\mathbf{w} = \mathbf{z},\end{aligned} \tag{16}$$

where $E\left[G(|\mathbf{w}^H\mathbf{y}_d|^2)\right]$ can be approximated by $\frac{1}{D}\sum_{d=1}^{D} G(|\mathbf{w}^H\mathbf{y}_d|^2)$ due to the ergodicity of the signals, $\iota(\mathbf{w})$ is an indicator function such that:

$$\iota(\mathbf{w}) = \begin{cases} 0, & \text{if } \mathbf{w}^H\mathbf{w} \le \delta_{\mathbf{w}}, \\ +\infty, & \text{otherwise.}\end{cases} \tag{17}$$

The augmented Lagrangian of (16) is:

$$\begin{aligned}\mathcal{L}(\mathbf{w},\mathbf{z},\boldsymbol{\eta}) = & -E\left[G(|\mathbf{w}^H\mathbf{Y}|^2)\right] + \lambda_1\|\mathbf{z}\|_1 + \iota(\mathbf{w}) \\ & + \mathcal{R}\{\boldsymbol{\eta}^H(\mathbf{Y}^H\mathbf{w} - \mathbf{z})\} + \frac{\rho_{\mathbf{z}}}{2}\|\mathbf{Y}^H\mathbf{w} - \mathbf{z}\|_2^2, \\ & \text{s.t. } \mathbf{w}^H\mathbf{c}_1 = 1, \mathbf{w}^H\mathbf{c}_2 = 0.\end{aligned} \tag{18}$$

According to [23], when the ADMM-type methods are used to solve (18), a stationary solution is achievable when $\rho_{\mathbf{z}}$ is large enough even if the term $-E\left[G(|\mathbf{w}^H\mathbf{Y}|^2)\right]$ is not always convex. The problem (18) can be solved via some iterative steps, in the $(k+1)^{th}$ iteration, the parameters can be determined by:

$$\begin{aligned}\mathbf{w}^{(k+1)} = & \arg\min_{\mathbf{w}} \mathcal{L}(\mathbf{w}^{(k)}, \mathbf{z}^{(k)}, \boldsymbol{\eta}^{(k)}), \\ & \text{s.t. } \mathbf{w}^{(k+1)^H}\mathbf{c}_1 = 1, \mathbf{w}^{(k+1)^H}\mathbf{c}_2 = 0,\end{aligned} \tag{19a}$$

$$\mathbf{z}^{(k+1)} = \arg\min_{\mathbf{z}} \mathcal{L}(\mathbf{w}^{(k+1)}, \mathbf{z}^{(k)}, \boldsymbol{\eta}^{(k)}), \tag{19b}$$

$$\boldsymbol{\eta}^{(k+1)} = \boldsymbol{\eta}^{(k)} + \gamma(\mathbf{Y}^H\mathbf{w}^{(k+1)} - \mathbf{z}^{(k+1)}), \tag{19c}$$

where γ is the step size parameter.

The details for solving the problem (19) can be found in Appendix A.

In addition, backtracking steps are given in [20] for the selection of dynamic Lipschitz constants to guarantee the convergence when the objective functions are continuously differentiable. In this paper, we utilise the same method for the Lipschitz constant $L_{\mathbf{w}}$ in the steps to solve the objective function $G(\cdot)$, and extend this method for the Lipschitz constant $L_{\mathbf{z}}$ in the steps to solve the non continuously differentiable objective function such as l_1 norm with subgradient, where $\partial\cdot$ is the subgradient operator. The algorithm for solving the problem (16) is summarized in Algorithm 1. In our simulations, we have found that large enough values for the initialization of the Lipschitz constants are important for the efficiency of the algorithm.

Algorithm 1 LPADMM based Beamforming for CD sources.

1: Initialization: $k = j = 0$, $\mu_{\mathbf{w}}$, $L_{\mathbf{w}}$, $\rho_{\mathbf{w}}$, $\mu_{\mathbf{z}}$, $L_{\mathbf{z}}$, $\rho_{\mathbf{z}}$;
2: **repeat**
3: $j = 0$;
4: **repeat**
5: $\mu_{\mathbf{w}}^{(j)} = L_{\mathbf{w}}^{(j)} + \rho_{\mathbf{w}}\|\mathbf{Y}\|^2$,
6: Calculate (32a) and (32b);
7: **if** $G(\mathbf{w}^{(j+1)}) - G(\mathbf{w}^{(j)}) - \left\langle g(\mathbf{w}^{(j)}), \mathbf{w}^{(j+1)} - \mathbf{w}^{(j)}\right\rangle > \frac{L_{\mathbf{w}}}{2}\|\mathbf{w}^{(j+1)} - \mathbf{w}^{(j)}\|^2$ **then**
8: $L_{\mathbf{w}}^{(j+1)} = 2L_{\mathbf{w}}^{(j)}$, go to line 5;
9: **else**
10: $L_{\mathbf{w}}^{(j+1)} = L_{\mathbf{w}}^{(j)}$;
11: **end if**
12: $j = j + 1$;
13: **until** Convergence;
14: $\mathbf{w}^{(k+1)} = \mathbf{w}^{(j)}$;
15: $\mu_{\mathbf{z}}^{(k)} = L_{\mathbf{z}}^{(k)} + \rho_{\mathbf{z}}$;
16: Calculate (19b);
17: **if** $(\|\mathbf{z}^{(k+1)}\|_1 - \|\mathbf{z}^{(k)}\|_1) - \left\langle \partial\|\mathbf{z}^{(k)}\|_1, \mathbf{z}^{(k+1)} - \mathbf{z}^{(k)}\right\rangle > \frac{L_{\mathbf{z}}}{2}\|\mathbf{z}^{(k+1)} - \mathbf{z}^{(k)}\|^2$ **then**
18: $L_{\mathbf{z}}^{(k+1)} = 2L_{\mathbf{z}}^{(k)}$, go to line 15;
19: **else**
20: $L_{\mathbf{z}}^{(k+1)} = L_{\mathbf{z}}^{(k)}$;
21: **end if**
22: Calculate (19c);
23: $k = k + 1$;
24: **until** Convergence;

2) Target CD source and ID interference: In the applications such as noise reduction, the target source is always information bearing and CD. However, the interference can be either CD or directional ID with limited angular spread and definite DOA. The common scenarios of such directional ID source is the cluster of noise, such as the noise at the outlet of the air-condition, or the aero-acoustic source related to a turbulent flow. In this subsection, we discuss the scenario that the target source is CD and the interference is ID. The steering vector of the target source can be directly given by $\mathbf{c}_1$, while no constant steering vector can be directly associated to the ID interference. Assuming that the angular spread of the ID interference is small enough ($< 10^\circ$), we propose to use an steering vector of an equivalent CD source to model it. That is to say, we use the DOA and the angular spread of the ID interference in a CD source steering vector. In the proposed algorithm (20), we relax the linear constraints $\mathbf{w}^H\mathbf{c}_1 = 1$ and $\mathbf{w}^H\mathbf{c}_2 = 0$ in (15) to some extent for the reasons that: 1) A trade-off can be considered between the

linear constraint $\mathbf{w}^H\mathbf{c}_1 = 1$ and the output SINR. 2) As explained before, $\mathbf{c}_2$ is an approximate model for the ID interference. To compensate for the error introduced by this approximation, we relax the equality constraint on the steering vectors in (15) to an inequality constrained by δ_2. We propose to define the Beamforming as the solution:

$$\begin{aligned}&\min_{\mathbf{w}} S(\mathbf{w}^H\mathbf{Y}) + \lambda_1\|\mathbf{w}^H\mathbf{Y}\|_1\\ &\text{s.t. } \mathbf{w}^H\mathbf{w} \le \delta_{\mathbf{w}}, \|\mathbf{w}^H\mathbf{c}_1 - 1\|_2^2 \le \delta_1, \|\mathbf{w}^H\mathbf{c}_2\|_2^2 \le \delta_2, \quad (20)\end{aligned}$$

where $\delta_{\mathbf{w}}$, δ_1 and δ_2 can be given the minimum values as the convergence of the algorithm is guaranteed. Introducing an extra variable $\mathbf{z}$, the problem (20) can be reformulated as:

$$\begin{aligned}&\min_{\mathbf{w},\mathbf{z}} \; -E\left[G(|\mathbf{w}^H\mathbf{Y}|^2)\right] + \lambda_1\|\mathbf{z}\|_1 + \iota(\mathbf{w})\\ &\text{s.t. } \mathbf{Y}^H\mathbf{w} = \mathbf{z}, \|\mathbf{w}^H\mathbf{c}_1 - 1\|_2^2 \le \delta_1, \|\mathbf{w}^H\mathbf{c}_2\|_2^2 \le \delta_2. \quad (21)\end{aligned}$$

The augmented Lagrangian of (20) is:

$$\begin{aligned}\mathcal{L}(\mathbf{w},\mathbf{z},\boldsymbol{\eta}) = &-E\left[G(|\mathbf{w}^H\mathbf{Y}|^2)\right] + \lambda_1\|\mathbf{z}\|_1 + \iota(\mathbf{w})\\ &+ \mathcal{R}\{\boldsymbol{\eta}^H(\mathbf{Y}^H\mathbf{w} - \mathbf{z})\} + \frac{\rho_{\mathbf{z}}}{2}\|\mathbf{Y}^H\mathbf{w} - \mathbf{z}\|_2^2,\\ &\text{s.t. } \|\mathbf{w}^H\mathbf{c}_1 - 1\|^2 \le \delta_1, \|\mathbf{w}^H\mathbf{c}_2\|^2 \le \delta_2, \quad (22)\end{aligned}$$

The problem (22) can be solved via some iterative steps, in the $(k+1)^{th}$ iteration, the parameters can be determined by:

$$\begin{aligned}&\mathbf{w}^{(k+1)} = \arg\min_{\mathbf{w}} \mathcal{L}(\mathbf{w}^{(k)}, \mathbf{z}^{(k)}, \boldsymbol{\eta}^{(k)}),\\ &\text{s.t. } \|\mathbf{w}^{(k+1)^H}\mathbf{c}_1 - 1\|^2 \le \delta_1, \|\mathbf{w}^{(k+1)^H}\mathbf{c}_2\|^2 \le \delta_2, \quad (23a)\\ &\mathbf{z}^{(k+1)} = \arg\min_{\mathbf{z}} \mathcal{L}(\mathbf{w}^{(k+1)}, \mathbf{z}^{(k)}, \boldsymbol{\eta}^{(k)}), \quad (23b)\\ &\boldsymbol{\eta}^{(k+1)} = \boldsymbol{\eta}^{(k)} + \gamma(\mathbf{Y}^H\mathbf{w}^{(k+1)} - \mathbf{z}^{(k+1)}), \quad (23c)\end{aligned}$$

where γ is the step size parameter.

The details for solving the problem (23) can be found in Appendix B.

Note that the projection gradient is also used in [14], the advantage of our method is that by introducing extra variables, it is possible to avoid solving a fourth order equation for the optimal Lagrange multiplier and the computational complexity is further reduced.

The algorithm for solving the problem (20) with backtracking steps is given in Algorithm 2. Note that there is no need to utilise backtracking steps for estimate μ_{u_1} and μ_{u_2}, as the objective functions for μ_{u_1} and μ_{u_2} can be regarded as constants.

Algorithm 2 LPADMM based Beamforming for ID sources.

Initialization: $k = j = 0$, $\mu_{\mathbf{w}}$, $L_{\mathbf{w}}$, $\rho_{\mathbf{w}}$, $\mu_{\mathbf{z}}$, $L_{\mathbf{z}}$, $\rho_{\mathbf{z}}$, μ_{u_1}, L_{u_1}, ρ_{u_1}, μ_{u_2}, L_{u_2}, ρ_{u_2};
repeat
  $j = 0$;
  repeat
    $\mu_{\mathbf{w}}^{(j)} = L_{\mathbf{w}}^{(j)} + \rho_{\mathbf{w}}\|\mathbf{Y}\|^2$;
    Calculate (41a)-(41e);
    if $G(\mathbf{w}^{(j+1)}) - G(\mathbf{w}^{(j)}) - \langle g(\mathbf{w}^{(j)}), \mathbf{w}^{(j+1)} - \mathbf{w}^{(j)}\rangle > \frac{L_{\mathbf{w}}}{2}\|\mathbf{w}^{(j+1)} - \mathbf{w}^{(j)}\|^2$ **then**
      $L_{\mathbf{w}}^{(j+1)} = 2L_{\mathbf{w}}^{(j)}$, go to line 5;
    else
      $L_{\mathbf{w}}^{(j+1)} = L_{\mathbf{w}}^{(j)}$;
    end if
    $j = j + 1$;
  until Convergence;
  $\mathbf{w}^{(k+1)} = \mathbf{w}^{(j)}$;
  $\mu_{\mathbf{z}}^{(k)} = L_{\mathbf{z}}^{(k)} + \rho_{\mathbf{z}}$;
  Calculate (23b);
  if $(\|\mathbf{z}^{(k+1)}\|_1 - \|\mathbf{z}^{(k)}\|_1) - \langle \partial\|\mathbf{z}^{(k)}\|_1, \mathbf{z}^{(k+1)} - \mathbf{z}^{(k)}\rangle > \frac{L_{\mathbf{z}}}{2}\|\mathbf{z}^{(k+1)} - \mathbf{z}^{(k)}\|^2$ **then**
    $L_{\mathbf{z}}^{(k+1)} = 2L_{\mathbf{z}}^{(k)}$, go to line 15;
  else
    $L_{\mathbf{z}}^{(k+1)} = L_{\mathbf{z}}^{(k)}$;
  end if
  Calculate (23c);
  $k = k + 1$;
until Convergence;

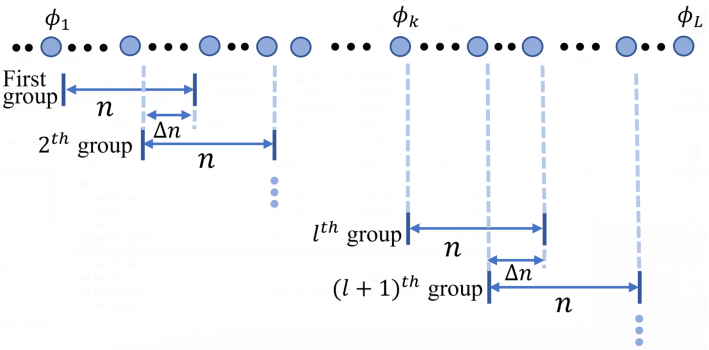


Fig. 1: The discretized angular sector of interest with overlapped groups

B. Generalized sparsity with overlapped groups for the Beampattern

It is well known that the Beampattern shape is related to the performance of the Beamforming, a Beampattern with a higher resolution implies better noise reduction performance. In our simulations, we have found when a distributed source and a distributed interference are too close to each other, the responses of the spatial filter to the signals coming from other directions different from the target source are amplified, which causes severe degradation of the Beampattern. Considering this fact, we propose a criteria for CD and ID sources based on the generalized sparsity of the Beampattern. That is to say, the responses of the spatial filter to the signal components within the sector of the distributed source are kept constant, while that to the signal components from other directions are minimized.

To exploit the generalized sparsity with overlapped groups of the Beampattern, firstly, we divide the angular sector of interest into overlapped groups as illustrated in Figure 1. Assume that the whole angular sector of interest is discretized by L grid point such that $\phi_k = \phi_1 + (k-1)\delta, k = 1, ..., L$, where δ is the angular separation between two points of the grid. We assume the position and the spatial extension of

the source are known (they can be previously estimated with methods such as [24]). We assume that the target source is spread over n grid points and that Δn is the number of points within the overlapped sector of the target source and the interference as illustrated in Figure 1. The overlapped groups can be constructed such that: 1)There are n points in each group and Δn points in the overlapped sector of two neighbour groups. 2)The l^{th} group corresponds to the target source and the $(l-1)^{th}$ or the $(l+1)^{th}$ group corresponds to the interference, where $1 \leq l \leq L+1-n$. 3)To ensure that the l^{th} group is centered on the DOA of the target source, the minimum bound of the sector ϕ_1 has to been chosen correctly. In this way, $N = \lfloor \frac{L-n}{n-\Delta n} + 1 \rfloor$ groups are formed, where $\lfloor \cdot \rfloor$ is the floor operator.

In addition, we can define $\mathbf{A} = [\mathbf{a}(\phi_1), ..., \mathbf{a}(\phi_k), ..., \mathbf{a}(\phi_L)]$ as the matrix composed of the steering vectors corresponding to all the ϕ_k, thus, the vector $\mathbf{A}^H\mathbf{w}$ gives the Beampattern for the sector of interest. We introduce the matrix $\mathbf{K}$ such that the generalized sparsity with overlapped groups of the Beampattern can be considered as a generalized Lasso problem [25] and can be achieved by the minimization of $\|\mathbf{K}\mathbf{A}^H\mathbf{w}\|_1$. The $(N \times L)$ matrix $\mathbf{K}$ is defined as follow:

$$\mathbf{K} = \begin{bmatrix} 1, & ..., & 1, & 0, & 0, & ..., & 0, & 0, & 0 \\ & ... & & & & & & & \\ 0, & ..., & 0, & 1, & ..., & 1, & 0, & ..., & 0 \\ & ... & & & & & & & \\ 0, & 0, & 0, & ..., & 0, & 0, & 1, & ..., & 1 \end{bmatrix}, \quad (24)$$

where each line corresponds to a group with n elements equal to 1 and other elements are 0. More precisely, the $(p,q)^{th}$ component of $\mathbf{K}$ can be given as:

$$\mathbf{K}_{p,q} = \begin{cases} 1, & \text{if} \quad (n-\Delta n)(p-1)+1 \leq q \\ & \leq (n-\Delta n)(p-1)+n, \\ 0, & \text{otherwise}, \end{cases} \quad (25)$$

Note that in the case the angular spread of the interference is larger than that of the target source, without loss of generality, assuming that the interference locates in the $(l+1)^{th}$ line of $\mathbf{K}$ and there are n_1 points within the interference, where $n_1 > n$, the $(l+1)^{th}$ line of $\mathbf{K}$ can be given as:

$$\mathbf{K}_{l+1,q} = \begin{cases} 1, & \text{if} \quad (n-\Delta n)l+1 \leq q \\ & \leq (n-\Delta n)l+n_1, \\ 0, & \text{otherwise}. \end{cases} \quad (26)$$

In the following, we discuss the Beamforming based on the generalized sparsity with overlapped groups of the Beampattern in the scenario of CD and ID sources.

1) Target CD source and CD interference: In the scenario that both the target source and interference are CD, the proposed Beamforming algorithm can be given as:

$$\begin{aligned} &\min_{\mathbf{w}} \|\mathbf{K}\mathbf{A}^H\mathbf{w}\|_1 \\ &\text{s.t. } \mathbf{w}^H\mathbf{w} \leq \delta_{\mathbf{w}}, \mathbf{w}^H\mathbf{c}_1 = 1, \mathbf{w}^H\mathbf{c}_2 = 0, \end{aligned} \quad (27)$$

Similarly to section III-A1, we can introduce an extra parameter $\mathbf{t} = \mathbf{K}\mathbf{A}^H\mathbf{w}$ and the augmented Lagrangian for the new problem can be given as:

$$\begin{aligned} \mathcal{L}(\mathbf{w}, \mathbf{t}, \boldsymbol{\eta}_3) =& \lambda_2\|\mathbf{t}\|_1 + \iota(\mathbf{w}) \\ &+ \mathcal{R}\{\boldsymbol{\eta}_3^H(\mathbf{K}\mathbf{A}^H\mathbf{w} - \mathbf{t})\} + \frac{\rho_{\mathbf{t}}}{2}\|\mathbf{K}\mathbf{A}^H\mathbf{w} - \mathbf{t}\|_2^2, \\ &\text{s.t. } \mathbf{w}^H\mathbf{c}_1 = 1, \mathbf{w}^H\mathbf{c}_2 = 0. \end{aligned} \quad (28)$$

The problem (27) can be solve with similar steps in Algorithm 1 for the problem (18).

2) Target CD source and ID interference: In the scenario that the target source is CD and the interference is ID, the main idea of the extension of the linear constraints on $\mathbf{c}_1$ and $\mathbf{c}_2$ to non-linear constraints is similar to that in III-A2, the proposed Beamforming algorithm can be given as:

$$\begin{aligned} &\min_{\mathbf{w}} \|\mathbf{K}\mathbf{A}^H\mathbf{w}\|_1 \\ &\text{s.t. } \mathbf{w}^H\mathbf{w} \leq \delta_{\mathbf{w}}, \|\mathbf{w}^H\mathbf{c}_1 - 1\|_2^2 \leq \delta_1, \|\mathbf{w}^H\mathbf{c}_2\|_2^2 \leq \delta_2. \end{aligned} \quad (29)$$

The augmented Lagrange for problem (29) can be given as:

$$\begin{aligned} \mathcal{L}(\mathbf{w}, \mathbf{t}, \boldsymbol{\eta}_4) =& \lambda_2\|\mathbf{t}\|_1 + \iota(\mathbf{w}) \\ &+ \mathcal{R}\{\boldsymbol{\eta}_4^H(\mathbf{K}\mathbf{A}^H\mathbf{w} - \mathbf{t})\} + \frac{\rho_{\mathbf{t}}}{2}\|\mathbf{K}\mathbf{A}^H\mathbf{w} - \mathbf{t}\|_2^2, \\ &\text{s.t. } \|\mathbf{w}^H\mathbf{c}_1 - 1\|_2^2 \leq \delta_1, \|\mathbf{w}^H\mathbf{c}_2\|_2^2 \leq \delta_2, \end{aligned} \quad (30)$$

which can be solved by similar steps based on LPADMM in Algorithm 2 with the introduction of extra variables $u_1 = \mathbf{w}^H\mathbf{c}_1 - 1$ and $u_2 = \mathbf{w}^H\mathbf{c}_2$.

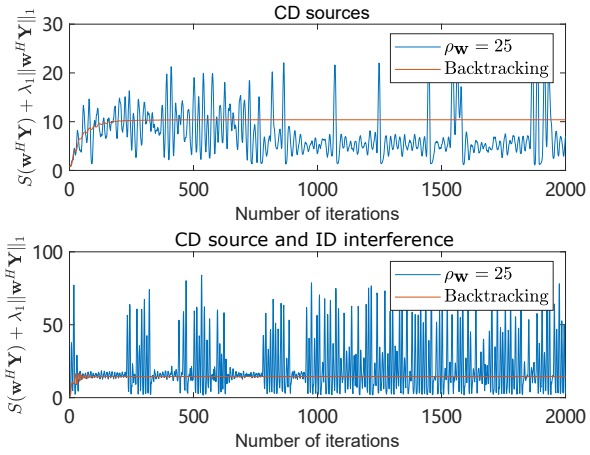
IV. Simulations

Expected if specified, we consider the following scenarios: the target source is uniformly distributed impinging from $60°$ and the interference is uniformly distributed impinging from the sector within $66°$. The angular spreads of both target source and interference are assumed to be $8°$. That is to say, the angular overlap of the target source and interference is $62° \sim 64°$. We set SIR $= 0$dB and SNR $= 10$dB. According to section II, the steering vector of a CD source indexed by i can be modelled by $\mathbf{c}_i$ as the integration of the steering vector $\mathbf{a}_i$ over the whole angular spread of the signal. In our simulations, the steering vector of an ID source is approximated by continuous independent CD sources whose angular spreads are all $1°$. The source is a female speaker reading a book from the dataset Librispeech [26], the total snapshots number is 2^{17}, the sampling rate of the source is 16 kHz, the STFT is computed with half-overlapping Hann window of 256 samples. The interference is a Gaussian random signal. A uniform linear array (ULA) of $M = 20$ omnidirectional microphones spaced by 5 cm is considered. Sources are assumed to be in far field.

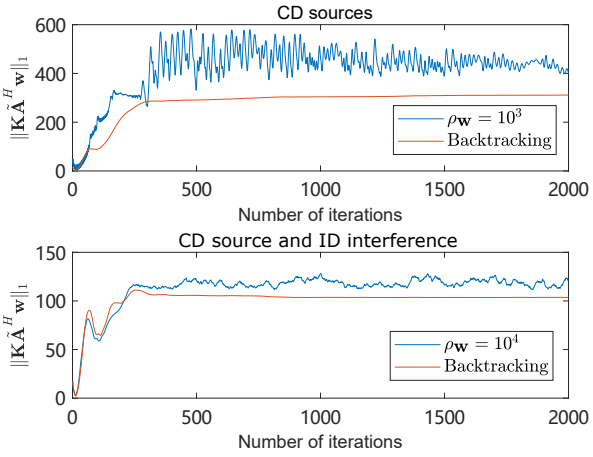
The Beampatterns and SINRs of the conventional LCMV Beamforming (legend: "LCMV"), the l_1 norm Beamforming proposed in [27] which is solved by the CVX toolbox (legend: "CVX"), and the two proposed Beamforming criteria (legend: "Entropy" for (15) and (20), "GeneLasso" for (27) and (29)) are compared in different configurations. The optimal value of the output SINR is the maximum eigenvalue of the matrix $(\mathbf{R}_{i+n})^{-1}\mathbf{R}_x$, where $\mathbf{R}_x$ and $\mathbf{R}_{i+n}$ are the covariance matrix of

source and the covariance matrix of interference plus noise, respectively. The log is used in the contrast function as in (13). 300 Monte Carlo trials are performed for the computation of the RMSE ans SINR. To illustrate the robustness, in all the simulations $\delta_{\mathbf{w}} = 0.35$ and 0.25 for CD sources and ID sources, respectively, $\delta_1 = \delta_2 = 0.1$.

A. Simulated data experiments



(a) Spectrogram sparsity based methods

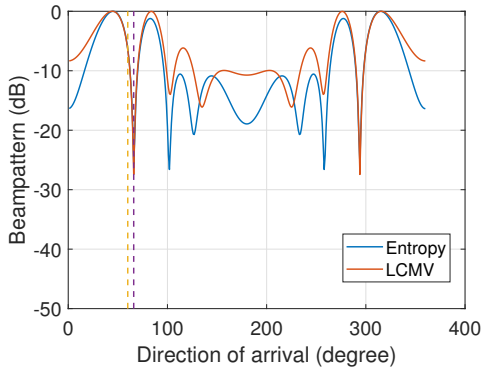


(b) Beampattern sparsity based methods

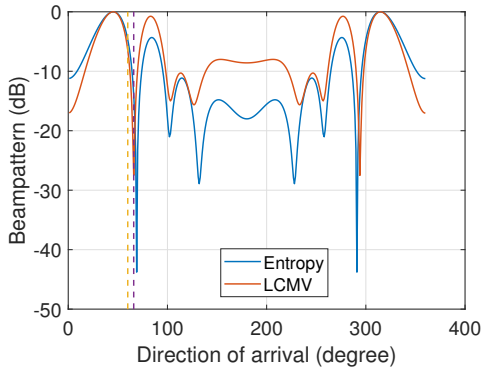
Fig. 2: Objective functions vs. number of iterations ($\theta_1 = 60°$, $\theta_2 = 66°$, uniformly distributed sources, $M = 20$)

1) Performance without steering vector mismatches: In figure 2, we first investigate the convergence behaviour of the proposed criteria with backtracking steps for the selection of the dynamic Lipschitz constants and with arbitrary Lipschitz constants, respectively. The objective functions versus the number of iterations for (15) and (20) are plotted in figure 2a, and that for (27) and (29) are plotted in figure 2b. It can be seen that in all the scenarios the LPADMM based methods can achieve convergence within 500 iterations with the backtracking steps, while it is hard to achieve convergence if improper values are selected for Lipschitz constants.

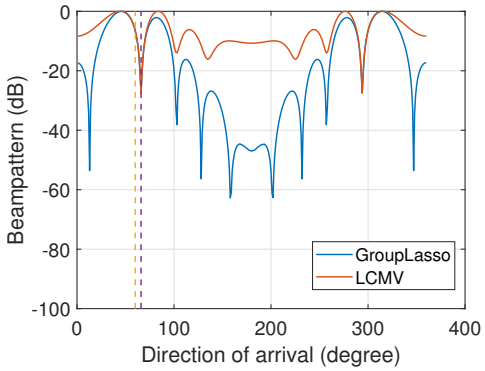
In figure 3 we show the Beampatterns of the conventional LCMV and the proposed criteria. It can be seen that the



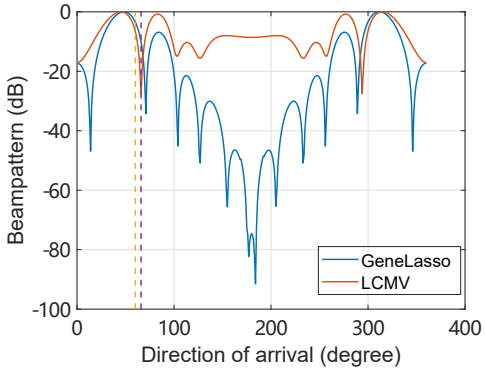
(a) Target CD source and CD interference for (15)



(b) Target CD source and ID interference for (20)



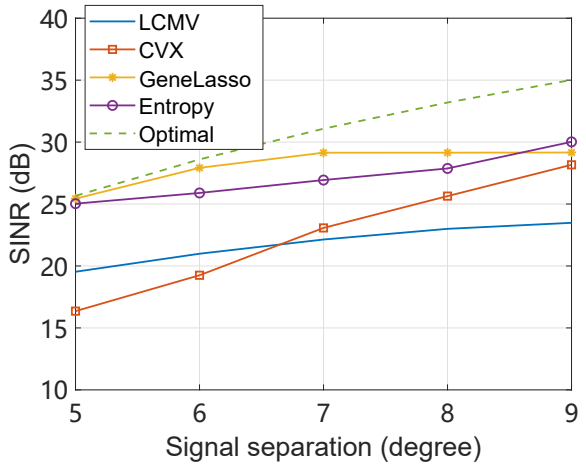
(c) Target CD source and CD interference for (27)



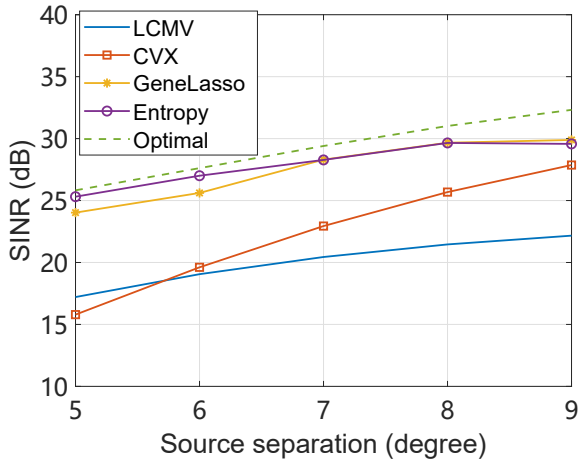
(d) Target CD source and ID interference for (29)

Fig. 3: Beampatterns ($\theta_1 = 60°$, $\theta_2 = 66°$, uniformly distributed sources, $M = 20$)

Beampatterns of the LCMV method and that of the proposed criteria are similar in the region of the overlapped sources, while in other regions the Beampatterns of the proposed criteria are much lower than that of the LCMV method. The results illustrate the advantage in the reduction of the overall noise of the proposed criteria. By comparing figure 3c and figure 3d to figure 3a and figure 3b we can see that due to the direct minimization of the Beampattern in the objective functions, the criteria proposed in section III-B has better overall noise reduction performance than that proposed in III-A.



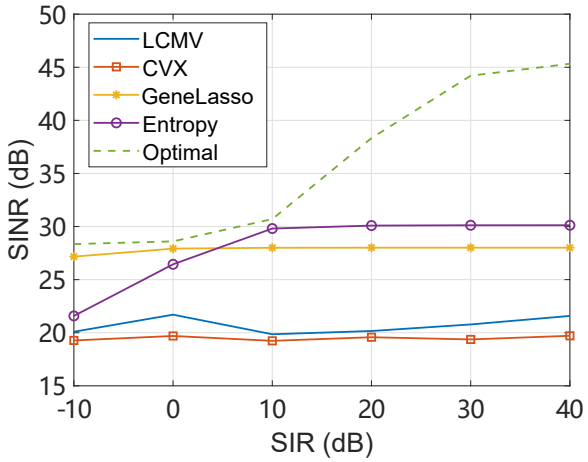
(a) Target CD source and CD interference



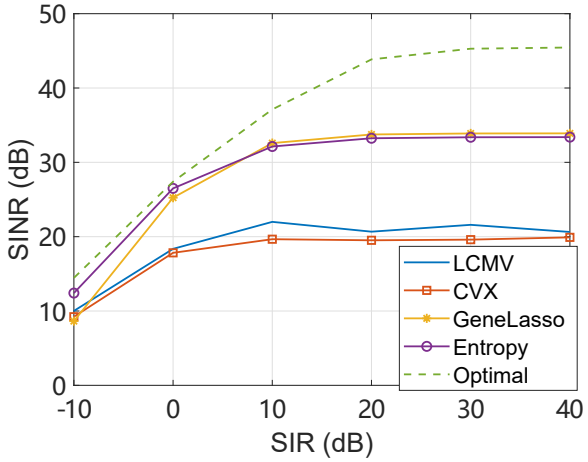
(b) Target CD source and ID interference

Fig. 4: Simulation data: Output SINR vs. signal angular separation ($\theta_1 = 60°$, $\theta_2 = 65° \sim 69°$, SIR $= 0$dB, SNR $= 10$dB, uniformly distributed signals, $M = 20$)

In figure 4, we plot the output SINR as a function of the angular separation between central DOAs of the two distributed sources for both CD and ID sources. The central DOA of the target distributed source is $\theta_1 = 60°$, while the central DOA of the distributed interference is varying from $65° \sim 69°$. In general, the output SINR increases as the angular separation between the source and the interference increases for all the Beamforming techniques. The advantages of the proposed criteria are more obvious when the angular separation is smaller.



(a) Target CD source and CD interference

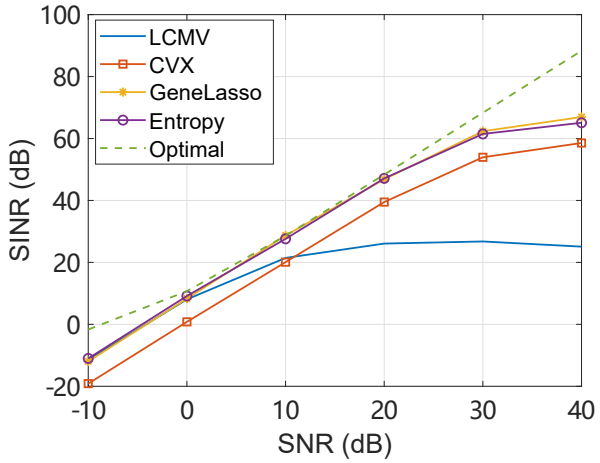


(b) Target CD source and ID interference

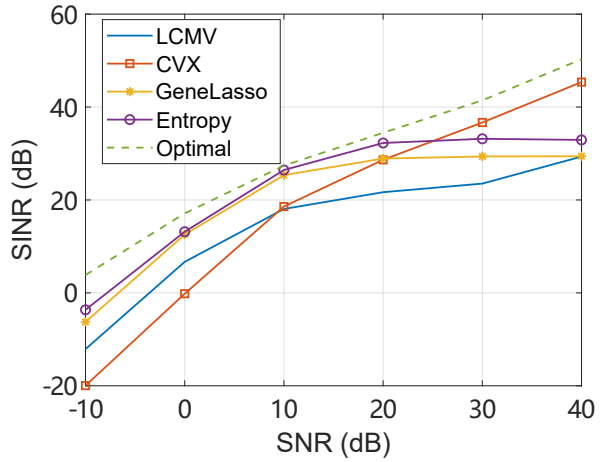
Fig. 5: Simulation data: Output SINR vs. SIR ($\theta_1 = 60°$, $\theta_2 = 66°$, uniformly distributed signals, $M = 20$)

In figure 5 the output SINR as a function of SIR is presented. In general, the output SINRs of the proposed criteria outperform other methods for all the SIRs. From figure 5a we can see that the conventional LCMV, the l_1 norm Beamforming, and the Beampattern sparsity based criterion proposed in section III-B are robust to SIR due to the linear constraint $\mathbf{w}^H\mathbf{c} = 0$, while in figure 5b the performance improves as the SIR increases for all methods.

The output SINR as a function of SNR is shown in figure 6. In general the output SINRs of all the Beamforming techniques increase as the SNR increases. We can see that the proposed criteria outperform the other methods in most cases except when the SNR is higher than 30 dB in the CD source and ID interference scenario. In this case the values of the parameters $\delta_{\mathbf{w}}$, δ_1 and δ_2 given at the beginning of this section are no longer optimal for the proposed criteria. The methods for choosing the optimal values automatically for the parameters in different configurations will be investigated in the future work.



(a) Target CD source and CD interference

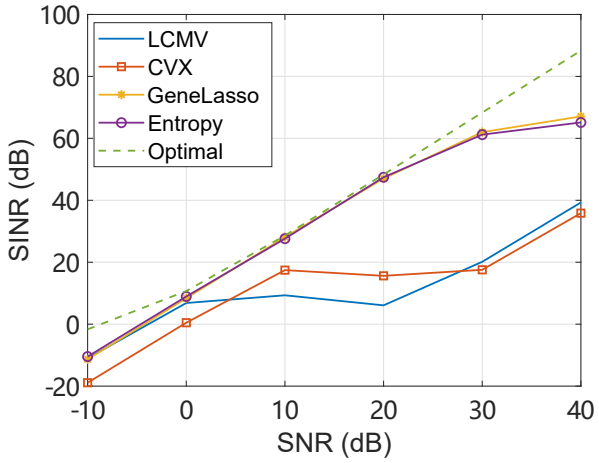


(b) Target CD source and ID interference

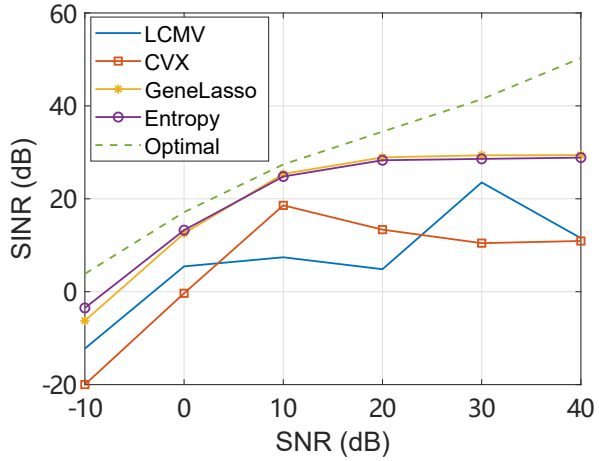
Fig. 6: Simulation data: Output SINR vs. SNR ($\theta_1 = 60°$, $\theta_2 = 66°$, uniformly distributed signals, $M = 20$)

2) Performance with steering vector mismatches: In figure 7 we consider the performance of the Beamforming techniques when steering vector mismatches are caused by the DOA estimation error: The central DOA of the target source in Beamforming is assumed to be $60°$ while the actual central DOA of the target source is $58°$. It can be seen that as the SNR increases, the performance of the LCMV and the l_1 norm Beamforming degrades severely due to the steering vector model mismatch, while the proposed criteria achieve almost the same performance as in the scenario without steering vector mismatch.

In figure 9 we investigate the performance of the Beamforming techniques when steering vector mismatches are caused by the microphone position error: The difference between the coordinate of actual microphone array and that in the steering vector model used in Beamforming is 3 cm. It can be seen that the microphone position mismatch influences little the performance of the Beamforming techniques due to the far field assumption, except for the LCMV in the high SNR scenario.



(a) Target CD source and CD interference



(b) Target CD source and ID interference

Fig. 7: Simulation data: Output SINR vs. SNR ($\theta_1 = 58°$, $\theta_2 = 66°$, uniformly distributed signals, $M = 20$)

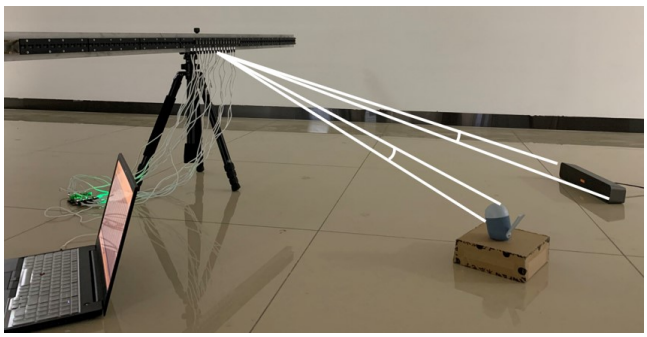
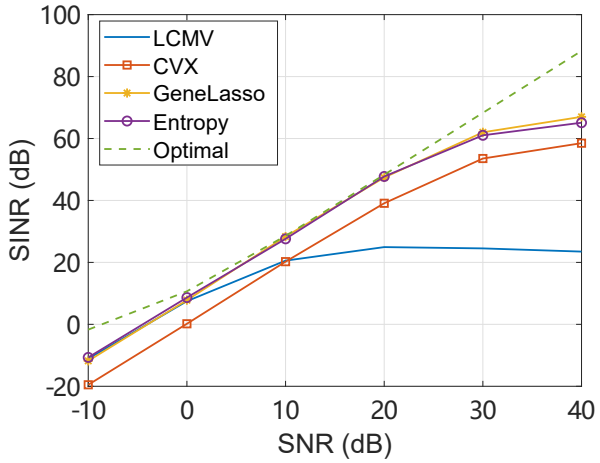
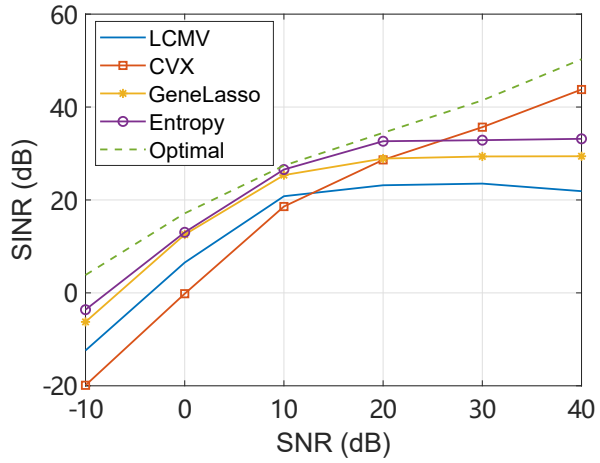


Fig. 8: Microphone array and soundboxes used in real data experiments (the larger soundbox width: 40 cm, the smaller soundbox width: 5 cm)

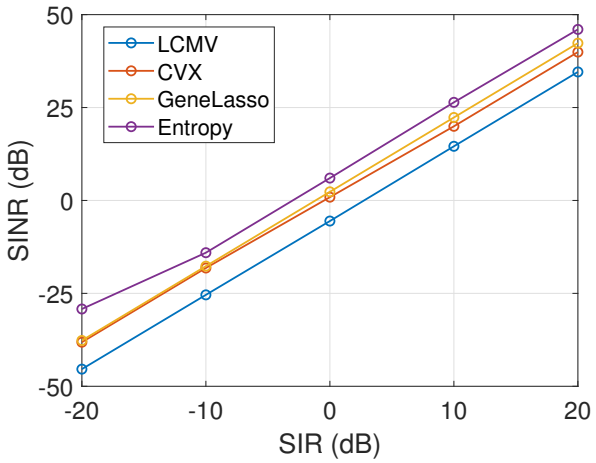


(a) Target CD source and CD interference

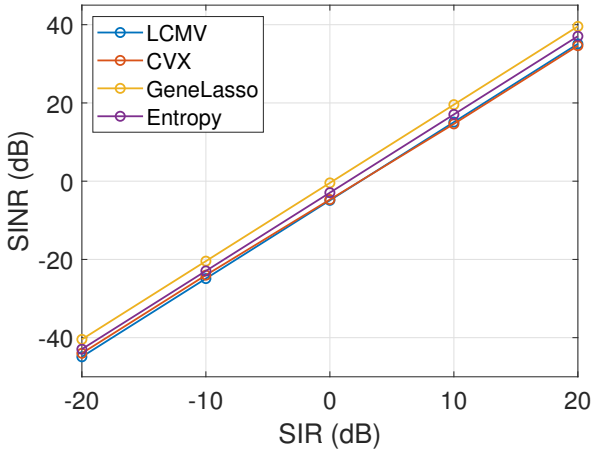


(b) Target CD source and ID interference

Fig. 9: Simulation data: Output SINR vs. SNR ($\theta_1 = 60°$, $\theta_2 = 66°$, microphone position mismatch 3 cm, uniformly distributed signals, $M = 20$)



(a) Target CD source and CD interference



(b) Target CD source and ID interference

Fig. 10: Real data: Output SINR vs. SIR ($\theta_1 = 90°$, $\theta_2 = 87.5°$, inter-element space of microphone 2 cm, uniformly distributed signals, $M = 10$)

B. Real data experiments

	Sources	DOA	Angular spread
fig.10a	CD target source	90.5°	7.6°
	CD interference	87.5°	7.6°
fig.10b	CD target source	90°	7.6°
	ID interference	87.5°	5°

TABLE I: Angular configurations of CD/ID sources in real data experiments.

The real data experiments are realized at Beijing Institute of Technology. As shown in figure 8, a linear uniform array consisting of ten omnidirectional microphones is used to pick up sound signals. Soundboxes are placed about 3 m far away from the microphone array and two types of soundboxes are used: the smaller one is 5 cm wide and its angular spread is 1°, the larger one is 40 cm wide and its angular spread is about 7.6°. Thus, the larger soundbox can be considered as a CD source, while an ID source can be constructed by placing the smaller soundbox at continuous DOA and then adding up all the received signals at the microphone array. It is true that the constructed source is not a real physical ID source, but the approximation is mathematically legitimate. Music segments are used as target sources and white Gaussian noise segments are used as interferences. The overall noise is estimated by noise-only frames. The SNR is about 20 dB. Here the soundboxes are chosen as an example of the distributed source because the parameters such as the distance between the soundbox and the microphone array are easy to control.

In figure 10 the performance of the noise reduction methods is compared, the DOAs and angular spread configurations are given in TABLE I. We can see that in the CD target source-CD interference scenario, the angular overlap of the target source and the interference is $86.7° \sim 91.3°$; in the CD target source-ID interference scenario, the angular overlap of the target source and the interference is $86.2° \sim 90°$. We can also see that due to the different SNRs, the reverberation, the DOA estimation error, the hardware limitation, etc, the output SINRs in the real data experiments are different compared to

the simulated data experiments (see figure 5). Nevertheless, the two proposed algorithms still outperform the baseline algorithms.

V. Conclusions

In this paper, we have proposed two Beamforming criteria for the case that the target source and the interference are spatially distributed and overlapped: in the first criterion we exploit the entropy/non-Gaussianity as well as the sparsity of the spectrogram of the output of the microphone array; in the second criterion we exploit the generalized sparsity with overlapped groups of the Beampattern. In the two proposed criteria, the constraints for keeping the target source and cancelling the interference are linear for CD sources and are relaxed to non linear ones for ID interference. The criteria are solved by LPADMM based methods including soft-thresholding and projection gradient descent with high accuracy and high computational efficiency. The simulations and real data experiments have shown the advantages of the proposed criteria compared to previously proposed Beamforming methods.

Appendix A

The subproblem (19a) is an optimization problem with linear constraints. To solve (19a), firstly, ignoring the constants terms, the augmented Lagrangian of (19a) can be given as:

$$\begin{aligned}\mathcal{L}_a(\mathbf{w},\boldsymbol{\eta}_1) =& -E\left[G(|\mathbf{Y}^H\mathbf{w}|^2)\right]+\mathcal{R}\{\boldsymbol{\eta}^H(\mathbf{Y}^H\mathbf{w}-\mathbf{z})\}\\ &+\frac{\rho_{\mathbf{z}}}{2}\|\mathbf{Y}^H\mathbf{w}-\mathbf{z}\|_2^2+\mathcal{R}\{\boldsymbol{\eta}_1^H(\mathbf{C}^H\mathbf{w}-\mathbf{p})\}\\ &+\frac{\rho_{\mathbf{w}}}{2}\|\mathbf{C}^H\mathbf{w}-\mathbf{p}\|_2^2+\iota(\mathbf{w}),\end{aligned} \tag{31}$$

where $\mathbf{C}=[\mathbf{c}_1,\mathbf{c}_2]$, $\mathbf{p}=[1;0]$. Similarly to (18), this subproblem can be solved via some iterative steps, for the $(j+1)^{th}$ iteration to solve (31) in the $(k+1)^{th}$ iteration in (19), the parameters can be determined as:

$$\mathbf{w}^{(j+1)}=\arg\min_{\mathbf{w}}\mathcal{L}_a(\mathbf{w}^{(j)},\boldsymbol{\eta}_1^{(j)},\mathbf{z}^{(k)},\boldsymbol{\eta}^{(k)}), \tag{32a}$$

$$\boldsymbol{\eta}_1^{(j+1)}=\boldsymbol{\eta}_1^{(j)}+\gamma_1(\mathbf{C}^H\mathbf{w}^{(j+1)}-\mathbf{p}), \tag{32b}$$

where γ_1 is the step size parameter.

Linearizing both the lower semi-continuous function and the linear constraint terms, and introducing the proximal term $\frac{\mu_{\mathbf{w}}}{2}\|\mathbf{w}-\mathbf{w}^{(j)}\|_2^2$, and ignoring the constants terms, the problem (32a) can be given as:

$$\begin{aligned}\mathbf{w}^{(j+1)}=&\arg\min_{\mathbf{w}}<\nabla_{\mathbf{w}}F(\mathbf{w}^{(j)},\boldsymbol{\eta}_1^{(j)},\mathbf{z}^{(k)},\boldsymbol{\eta}^{(k)}),\mathbf{w}-\mathbf{w}^{(j)}>\\ &+\frac{\mu_{\mathbf{w}}}{2}\|\mathbf{w}-\mathbf{w}^{(j)}\|_2^2+\iota(\mathbf{w}),\\ =&\arg\min_{\mathbf{w}}\frac{\mu_{\mathbf{w}}}{2}\|\mathbf{w}-(\mathbf{w}^{(j)}-\frac{1}{\mu_{\mathbf{w}}}\nabla_{\mathbf{w}}F(\mathbf{w}^{(j)},\boldsymbol{\eta}_1^{(j)},\mathbf{z}^{(k)},\boldsymbol{\eta}^{(k)}))\|_2^2\\ &+\iota(\mathbf{w}),\end{aligned} \tag{33}$$

where $\langle\cdot,\cdot\rangle$ is the inner product operator, F and $\nabla_{\mathbf{w}}F(\mathbf{w}^{(j)},\boldsymbol{\eta}_1^{(j)},\mathbf{z}^{(k)},\boldsymbol{\eta}^{(k)})$ are given in (34) and (35) at the top of next page, respectively, where $\nabla_{\mathbf{w}}F(\mathbf{w}^{(j)},\boldsymbol{\eta}_1^{(j)},\mathbf{z}^{(k)},\boldsymbol{\eta}^{(k)})$ is the gradient of F in terms of $\mathbf{w}$, and $g(x)$ is the gradient of $G(x)$ in terms of a scalar x.

The problem (33) can be solved by the projection gradient descent which can be given as:

$$\mathbf{w}^{(j+1)}=\mathcal{P}_c(\mathbf{w}^{(j)}-\frac{1}{\mu_{\mathbf{w}}}\nabla_{\mathbf{w}}F(\mathbf{w}^{(j)},\boldsymbol{\eta}_1^{(j)},\mathbf{z}^{(k)},\boldsymbol{\eta}^{(k)})), \tag{36}$$

where $\mathcal{P}_c(\mathbf{x})$ is the projection operator for a vector $\mathbf{x}$ such that:

$$\mathcal{P}_c(\mathbf{x})=\begin{cases}\mathbf{x}, & \text{if } \mathbf{x}^H\mathbf{x}\le\delta_{\mathbf{x}},\\ \frac{\sqrt{\delta_{\mathbf{x}}}\mathbf{x}}{\|\mathbf{x}\|_2}, & \text{otherwise.}\end{cases} \tag{37}$$

Similarly, to solve (19b), noting that $V(\mathbf{w}^{(k+1)},\mathbf{z}^{(k)},\boldsymbol{\eta}^{(k)})=\mathcal{R}\{\boldsymbol{\eta}^{(k)^H}(\mathbf{Y}^H\mathbf{w}^{(k+1)}-\mathbf{z}^{(k)})\}+\frac{\rho_{\mathbf{z}}}{2}\|\mathbf{Y}^H\mathbf{w}^{(k+1)}-\mathbf{z}^{(k)}\|_2^2$, and the gradient of V in terms of $\mathbf{z}$ as $\nabla_{\mathbf{z}}V(\mathbf{w}^{(k+1)},\mathbf{z}^{(k)},\boldsymbol{\eta}^{(k)})=-\boldsymbol{\eta}^{(k)}-\mathbf{Y}^H\mathbf{w}^{(k+1)}+\mathbf{z}^{(k)}$, introducing the proximal terms and ignoring the constants terms, (19b) can be given as:

$$\begin{aligned}\mathbf{z}^{(k+1)}=&\arg\min_{\mathbf{z}}<\nabla_{\mathbf{z}}V(\mathbf{w}^{(k+1)},\mathbf{z}^{(k)},\boldsymbol{\eta}^{(k)}),\mathbf{z}-\mathbf{z}^{(k)}>\\ &+\frac{\mu_{\mathbf{z}}}{2}\|\mathbf{z}-\mathbf{z}^{(k)}\|_2^2+\lambda_1\|\mathbf{z}\|_1\\ =&\arg\min_{\mathbf{z}}\frac{\mu_{\mathbf{z}}}{2}\|\mathbf{z}-(\mathbf{z}^{(k)}-\frac{1}{\mu_{\mathbf{z}}}\nabla_{\mathbf{z}}V(\mathbf{w}^{(k+1)},\mathbf{z}^{(k)},\boldsymbol{\eta}^{(k)}))\|_2^2\\ &+\lambda_1\|\mathbf{z}\|_1\\ =&\mathcal{S}_{\lambda_1/\mu_{\mathbf{z}}}\left(\mathbf{z}^{(k)}-\frac{1}{\mu_{\mathbf{z}}}\nabla_{\mathbf{z}}V(\mathbf{w}^{(k+1)},\mathbf{z}^{(k)},\boldsymbol{\eta}^{(k)})\right),\end{aligned} \tag{38}$$

where $\mathcal{S}_{\lambda_1/\mu_{\mathbf{z}}}(\mathbf{x})$ is the soft thresholding operator of the vector $\mathbf{x}$ such that:

$$\mathcal{S}_{\lambda_1/\mu_{\mathbf{z}}}(\mathbf{x})=\begin{cases}\mathbf{x}-\lambda_1/\mu_{\mathbf{z}}, & \text{if } \mathbf{x}\ge\lambda_1/\mu_{\mathbf{z}},\\ 0, & \text{otherwise,}\end{cases} \tag{39}$$

where the comparisons and subtractions are applied element-wise.

Appendix B

To solve (23a), introducing two extra variables $u_1=\mathbf{w}^H\mathbf{c}_1-1$ and $u_2=\mathbf{w}^H\mathbf{c}_2$, the augmented Lagrangian of this sub-problem can be given as:

$$\begin{aligned}&\mathcal{L}_a(\mathbf{w},u_1,u_2,\eta_1,\eta_2)=\\ &-E\left[G(|\mathbf{w}^H\mathbf{Y}|^2)\right]+\iota(\mathbf{w})+\iota(u_1)+\iota(u_2)+\\ &\mathcal{R}\{\boldsymbol{\eta}^H(\mathbf{Y}^H\mathbf{w}-\mathbf{z})+\eta_1^\star(\mathbf{w}^H\mathbf{c}_1-1-u_1)+\eta_2^\star(\mathbf{w}^H\mathbf{c}_2-u_2)\}+\\ &\frac{\rho_{\mathbf{z}}}{2}\|\mathbf{Y}^H\mathbf{w}-\mathbf{z}\|_2^2+\frac{\rho_{u_1}}{2}\|\mathbf{w}^H\mathbf{c}_1-1-u_1\|_2^2+\frac{\rho_{u_2}}{2}\|\mathbf{w}^H\mathbf{c}_2-u_2\|_2^2,\end{aligned} \tag{40}$$

where $\cdot^\star$ is the conjugate operator.

Similarly to the CD sources scenario, based on LPADMM, the parameters in (40) can be determined by several iterations, for the $(j+1)^{th}$ iteration to solve the problem (40) in the $(k+1)^{th}$ iteration in (23) , the parameters can be determined by:

$$\mathbf{w}^{(j+1)}=\mathcal{P}_c\left(\mathbf{w}^{(k)}-\frac{1}{\mu_{\mathbf{w}}}\nabla_{\mathbf{w}}F_1(\mathbf{w}^{(j)},u_1^{(j)},u_2^{(j)},\eta_1^{(j)},\eta_2^{(j)})\right), \tag{41a}$$

$$u_1^{(j+1)}=\mathcal{P}_c\left(u_1^{(j)}-\frac{1}{\mu_{u_1}}(-\eta_1^{(j)}-\mathbf{w}^{(j+1)^H}+1+u_1^{(j)})\right), \tag{41b}$$

$$F(\mathbf{w}^{(j)},\boldsymbol{\eta}_1^{(j)},\mathbf{z}^{(k)},\boldsymbol{\eta}^{(k)}) = -E\left[G(|\mathbf{Y}^H\mathbf{w}^{(j)}|^2)\right] + \mathcal{R}\{\boldsymbol{\eta}^{(k)^H}(\mathbf{Y}^H\mathbf{w}^{(j)}-\mathbf{z}^{(k)})\} + \frac{\rho_{\mathbf{z}}}{2}\|\mathbf{Y}^H\mathbf{w}^{(j)}-\mathbf{z}^{(k)}\|_2^2 + \mathcal{R}\{\boldsymbol{\eta}_1^{(j)^H}(\mathbf{C}^H\mathbf{w}^{(j)}-\mathbf{p})\} + \frac{\rho_{\mathbf{w}}}{2}\|\mathbf{C}^H\mathbf{w}^{(j)}-\mathbf{p}\|_2^2, \tag{34}$$

$$\nabla_{\mathbf{w}}F(\mathbf{w}^{(j)},\boldsymbol{\eta}_1^{(j)},\mathbf{z}^{(k)},\boldsymbol{\eta}^{(k)}) = -E\left[g(|\mathbf{Y}^H\mathbf{w}^{(j)}|^2)\mathbf{Y}\mathbf{Y}^H\mathbf{w}^{(j)}\right] + \mathbf{Y}\boldsymbol{\eta}^{(k)} + \frac{\rho_{\mathbf{z}}}{2}(\mathbf{Y}\mathbf{Y}^H\mathbf{w}^{(j)}-\mathbf{Y}\mathbf{z}^{(k)}) + \mathbf{C}\boldsymbol{\eta}_1^{(j)} + \frac{\rho_{\mathbf{w}}}{2}(\mathbf{C}\mathbf{C}^H\mathbf{w}^{(j)}-\mathbf{C}\mathbf{p}), \tag{35}$$

$$u_2^{(j+1)} = \mathcal{P}_c\left(u_2^{(j)} - \frac{1}{\mu_{u_2}}(-\eta_2^{(j)} - \mathbf{w}^{(j+1)^H} + u_2^{(j)})\right), \tag{41c}$$

$$\eta_1^{(j+1)} = \eta_1^{(j)} + \gamma_2(u_1^{(j+1)} - \mathbf{w}^{(j+1)^H}\mathbf{c}_1 + 1), \tag{41d}$$

$$\eta_2^{(j+1)} = \eta_2^{(j)} + \gamma_3(u_2^{(j+1)} - \mathbf{w}^{(j+1)^H}\mathbf{c}_2), \tag{41e}$$

where $\nabla_{\mathbf{w}}F_1(\mathbf{w}^{(j)},u_1^{(j)},u_2^{(j)},\eta_1^{(j)},\eta_2^{(j)})$ is the gradient of $F_1(\mathbf{w}^{(j)},u_1^{(j)},u_2^{(j)},\eta_1^{(j)},\eta_2^{(j)})$ in terms of $\mathbf{w}$, $F_1(\mathbf{w}^{(j)},u_1^{(j)},u_2^{(j)},\eta_1^{(j)},\eta_2^{(j)})$ can be given as:

$$\begin{aligned} F_1(\mathbf{w}^{(j)},u_1^{(j)},u_2^{(j)},\eta_1^{(j)},\eta_2^{(j)}) &= -E\left[G(|\mathbf{w}^{(j)^H}\mathbf{Y}|^2)\right] + \\ &\mathcal{R}\{\boldsymbol{\eta}^{(j)^H}(\mathbf{Y}^H\mathbf{w}^{(j)}-\mathbf{z}^{(k)})\} + \mathcal{R}\{\eta_1^{(j)^\star}(\mathbf{w}^{(j)^H}\mathbf{c}_1 - 1 - u_1^{(j)})\} \\ &+ \mathcal{R}\{\eta_2^{(j)^\star}(\mathbf{w}^{(j)^H}\mathbf{c}_2 - u_2^{(j)})\} + \frac{\rho_{\mathbf{z}}}{2}\|\mathbf{Y}^H\mathbf{w}^{(j)}-\mathbf{z}^{(k)}\|_2^2 \\ &+ \frac{\rho_{u_1}}{2}\|\mathbf{w}^{(j)^H}\mathbf{c}_1 - 1 - u_1^{(j)}\|_2^2 + \frac{\rho_{u_2}}{2}\|\mathbf{w}^{(j)^H}\mathbf{c}_2 - u_2^{(j)}\|_2^2. \end{aligned} \tag{42}$$

The problem (23b) can be solved with the same method for the problem (19b).

ACKNOWLEDGEMENT

The authors are grateful to professor SHEN Qing for his help in the real data experiments.

REFERENCES

[1] J. Benesty, J. Chen, Y. Huang, and B. Rafaely, "Microphone array signal processing," *Journal of the Acoustical Society of America*, vol. 125, no. 6, pp. 4097–4098, 2009.

[2] Y. H. Jacob Benesty, Jingdong Chen, *Microphone array signal processing*, S. T. in Signal Processing, Ed. Springer-Verlag Berlin Heidelberg, 2008.

[3] Y. Gu, N. A. Goodman, S. Hong, and Y. Li, "Robust adaptive beamforming based on interference covariance matrix sparse reconstruction," *Signal Processing*, vol. 96, pp. 375 – 381, 2014.

[4] Yi, Shanchao, Wu, Ying, Wang, and Yunlong, "Projection-based robust adaptive beamforming with quadratic constraint," *Signal Processing: The Official Publication of the European Association for Signal Processing (EURASIP)*, vol. 122, pp. 65–74, 2016.

[5] J. Li, P. Stoica, and Z. Wang, "On robust capon beamforming and diagonal loading," *IEEE Transactions on Signal Processing*, vol. 51, no. 7, pp. 1702–1715, 2003.

[6] S. Vorobyov, A. Gershman, and Z.-Q. Luo, "Robust adaptive beamforming using worst-case performance optimization: a solution to the signal mismatch problem," *IEEE Transactions on Signal Processing*, vol. 51, no. 2, pp. 313–324, 2003.

[7] Y. Huang, M. Zhou, and S. A. Vorobyov, "New designs on mvdr robust adaptive beamforming based on optimal steering vector estimation," *IEEE Transactions on Signal Processing*, vol. 67, no. 14, pp. 3624–3638, 2019.

[8] X. Jiang, W.-J. Zeng, A. Yasotharan, H. C. So, and T. Kirubarajan, "Minimum dispersion beamforming for non-gaussian signals," *IEEE Transactions on Signal Processing*, vol. 62, no. 7, pp. 1879–1893, 2014.

[9] X. Jiang, W.-J. Zeng, H. C. So, A. M. Zoubir, and T. Kirubarajan, "Beamforming via nonconvex linear regression," *IEEE Transactions on Signal Processing*, vol. 64, no. 7, pp. 1714–1728, 2016.

[10] G. Xia, W. Xia, and H. Yu, "Minimum gaussian entropy based distortionless adaptive beamforming algorithms," *IEEE Signal Processing Letters*, vol. 28, pp. 603–607, 2021.

[11] N. Yazdi and K. Todros, "Measure-transformed mvdr beamforming," *IEEE Signal Processing Letters*, vol. 27, pp. 1959–1963, 2020.

[12] P. Chevalier, J.-P. Delmas, and M. Sadok, "Third-order volterra mvdr beamforming for non-gaussian and potentially non-circular interference cancellation," *IEEE Transactions on Signal Processing*, vol. 66, no. 18, pp. 4766–4781, 2018.

[13] X. Jiang, J. Chen, X. Liu, A. M. Zoubir, and Z. Zhou, "Phase-only robust minimum dispersion beamforming," *IEEE Transactions on Signal Processing*, vol. 68, pp. 5664–5679, 2020.

[14] X. Jiang, W.-J. Zeng, A. Yasotharan, H. C. So, and T. Kirubarajan, "Quadratically constrained minimum dispersion beamforming via gradient projection," *IEEE Transactions on Signal Processing*, vol. 63, no. 1, pp. 192–205, 2015.

[15] Y. Liang, W. Cui, Q. Shen, W. Liu, and S. Wu, "Cramér-rao bound analysis of underdetermined wideband doa estimation under the subband model via frequency decomposition," *IEEE Transactions on Signal Processing*, vol. 69, pp. 4132–4148, 2021.

[16] S. Valaee, B. Champagne, and P. Kabal, "Parametric localization of distributed sources," *IEEE Transactions on Signal Processing*, vol. 43, no. 9, pp. 2144–2153, 1995.

[17] J. A. Chaaya, J. Picheral, and S. Marcos, "Localization of spatially distributed near-field sources with unknown angular spread shape," *Signal Processing*, vol. 106, no. jan., pp. 259–265, 2015.

[18] A. Zoubir, Y. Wang, and P. Chargé, "Efficient subspace-based estimator for localization of multiple incoherently distributed sources," *IEEE Transactions on Signal Processing*, vol. 56, no. 2, pp. 532–542, 2008.

[19] X. Wang, I. Cohen, J. Chen, and J. Benesty, "On robust and high directive beamforming with small-spacing microphone arrays for scattered sources," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 27, no. 4, pp. 842–852, 2019.

[20] Y. Ouyang, Y. Chen, G. Lan, and J. E. Pasiliao, "An accelerated linearized alternating direction method of multipliers," *SIAM Journal on Imaging Sciences*, vol. 8, no. 1, 2014.

[21] ELLA, BINGHAM, AAPO, and HYVÄRINEN, "A fast fixed-point algorithm for independent component analysis of complex valued signals," *International Journal of Neural Systems*, 2000.

[22] W. Deng, M. Lai, Z. Peng, and W. Yin, "Parallel multi-block admm with o(1 / k) convergence," *Journal of Scientific Computing*, 2016.

[23] M. Hong, Z. Luo, and M. Razaviyayn, "Convergence analysis of alternating direction method of multipliers for a family of nonconvex problems," *SIAM Journal on Optimization*, vol. 26, no. 1, pp. 337–364, 2016.

[24] W. Xiong, J. Picheral, G. Chardon, and S. Marcos, "Sparsity-based localization of spatially coherent distributed sources," in *IEEE International Conference on Acoustics*, 2016, pp. 3241–3245.

[25] T. J. Taylor, "The solution path of the generalized lasso," *Annals of Statistics*, vol. 39, no. 3, pp. 1335–1371, 2011.

[26] V. Panayotov, G. Chen, D. Povey, and S. Khudanpur, "Librispeech: An asr corpus based on public domain audio books," in *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2015, pp. 5206–5210.

[27] S. Emura, S. Araki, T. Nakatani, and N. Harada, "Distortionless beamforming optimized with l1 norm minimization," *IEEE Signal Processing Letters*, pp. 1–1, 2018.

XIONG Wenmeng received the B.E. degree in 2011 from Huazhong University of Science and Technology, China. She received the Ph.D. degree in Signal and Image processing in 2017 from CentraleSupelec, France. From 2017 to 2019, She was an algorithm engineer with Ericsson(China), Beijing, China. She is now an assistant professor with the Faculty of Information Technology at the Beijing University of Technology. Her current research interests lie in array signal processing, including source localization, speech enhancement and array geometry optimization.

BAO Changchun (IEEE Senior Member, CIE Fellow) is currently a Full Professor with the Faculty of Information Technology, Beijing University of Technology, Beijing, China. His research interests include speech and audio signal processing, speech coding, speech enhancement, speech transcoding, audio coding, audio enhancement, bandwidth extending for speech and audio signals, and 3D audio signal processing. He is the author or co-Author of more than 330 papers in journals and conferences and holds ten patents. He was ever an Associate Editor for the Journal of Communications, and currently the Editor of the Signal Processing and the Editor of the Journal of Data Acquisition & Processing. Prof. Bao is a Board Member of the Chinese Institute of Electronics (CIE), the Vice President of CIE Signal Processing Society Board of Governors, an Associate Editor and the Leader Guest Editor of EURASIP Journal on Audio, Speech, and Music Processing (JASMP), Member of International Speech Communication Association, the Chair of the APSIPASLA TC from 2015 to 2016, and the Chair of National Conference on Man-Machine Speech Communication-Standing Committee.

JIA Maoshen (M'13-SM'15) received the B.E. degree in Electronic Information Engineering in 2005 from Hebei University, China. He received the Ph.D. degree in Electronic Science and Technology in 2010 from Beijing University of Technology, China. Recently, he is a professor with the Faculty of Information Technology at the Beijing University of Technology. His current research interests include multichannel audio signal processing, audio coding and array signal processing.

José PICHERAL graduated from Supélec, Paris, France, and from the Politecnico di Milano, Milan, Italy, in 1999. He received is PhD from the Université Paris Sud in 2003 for his study about " High resolution methods for joint angle-delay estimation – application to UMTS and geophysics " and the HDR in 2017 for his work about " Parametric methods and inverse approach for coherently distributed sources localization ". He is currently associate Professor at CentraleSupelec and member of the Laboratory of Signals and Systems (L2S).