Tengan *et al.*
EURASIP Journal on Audio, Speech, and Music Processing (2023) 2023:38
https://doi.org/10.1186/s13636-023-00304-8

EURASIP Journal on Audio, Speech, and Music Processing

METHODOLOGY

Open Access

Direction-of-arrival and power spectral density estimation using a single directional microphone and group-sparse optimization

Elisa Tengan[1*], Thomas Dietzen[1], Filip Elvander[2] and Toon van Waterschoot[1]

Abstract

In this paper, two approaches are proposed for estimating the direction of arrival (DOA) and power spectral density (PSD) of stationary point sources by using a single, rotating, directional microphone. These approaches are based on a method previously presented by the authors, in which point source DOAs were estimated by using a broadband signal model and solving a group-sparse optimization problem, where the number of observations made by the rotating directional microphone can be lower than the number of candidate DOAs in an angular grid. The DOA estimation is followed by the estimation of the sources' PSDs through the solution of an overdetermined least squares problem. The first approach proposed in this paper includes the use of an additional nonnegativity constraint on the residual noise term when solving the group-sparse optimization problem and is referred to as the Group Lasso Least Squares (GL-LS) approach. The second proposed approach, in addition to the new nonnegativity constraint, employs a narrowband signal model when building the linear system of equations used for formulating the group-sparse optimization problem, where the DOAs and PSDs can be jointly estimated by iterative, group-wise reweighting. This is referred to as the Group-Lasso with l_1-reweighting (GL-L1) approach. Both proposed approaches are implemented using the alternating direction method of multipliers (ADMM), and their performance is evaluated through simulations in which different setup conditions are considered, ranging from different types of model mismatch to variations in the acoustic scene and microphone directivity pattern. The results obtained show that in a scenario involving a microphone response mismatch between observed data and the signal model used, having the additional nonnegativity constraint on the residual noise can improve the DOA estimation for the case of GL-LS and the PSD estimation for the case of GL-L1. Moreover, the GL-L1 approach can present an advantage over GL-LS in terms of DOA estimation performance in scenarios with low SNR or where multiple sources are closely located to each other. Finally, it is shown that having the least squares PSD re-estimation step is beneficial in most scenarios, such that GL-LS outperformed GL-L1 in terms of PSD estimation errors.

Keywords Direction-of-arrival estimation, Power spectral density estimation, Single-channel, Group-sparsity

1 Introduction

In the field of audio signal processing, the ability to exploit spectral and spatial information from the auditory scene plays an important role in developing speech enhancement, noise reduction, and scene analysis techniques [1–8]. The applications in which these tasks have great influence are numerous, with some examples being binaural processing for hearing aids, hands-free telephony and videoconferencing, acoustic surveillance,

*Correspondence:
Elisa Tengan
elisa.tengan@esat.kuleuven.be
[1] Department of Electrical Engineering (ESAT/STADIUS), KU Leuven, Leuven, Belgium
[2] Department of Information and Communications Engineering, Aalto University, Espoo, Finland

autonomous robots, and so on. By localizing target and interfering sound sources in space and estimating their power spectral densities (PSDs), one can distinguish them and process the recorded signals using the appropriate noise reduction and source classification methods.

Sound source localization is performed by estimating the direction of arrival (DOA) of the signals being recorded. Pioneering methods such as Capon's beamformer [9], the MUSIC algorithm [10], and generalized correlation-based methods [11, 12] are still used for the task of DOA estimation and keep being further modified for improved performance. Alternatively, different approaches have also been developed. Compressed sensing techniques have become more popular [13–16], with the focus on exploiting sparsity in the signal models considered. Moreover, with the growth of machine learning, data-driven methods have naturally gained more attention as well [17, 18].

In terms of PSD estimation, especially for noise and interfering sources, methods based on minimum statistics [19, 20], minimum mean squared error (MMSE) [21–23], and minima controlled recursive averaging (MCRA) [24–28] have set the baseline for allowing noise reduction and speech enhancement methods to be developed. Later, subspace-based methods [29–31], as well as deep learning methods [32–35], provided different perspectives on how to perform PSD estimation.

Although much has been achieved through the studies presented above, some challenges still remain present. When estimating DOAs, most methods rely on having recordings from multiple microphones, such that their spatial diversity is used for inferring a source's location [36]. However, it is well known that in practice, constraints on hardware design, computational complexity and simultaneous access to multiple microphones' data, often encountered in different devices, may limit the ideal advantages of microphone array processing techniques [37–39]. In this paper, we propose two alternative approaches for DOA and PSD estimation using a single, rotating, and directional microphone. By exploring the potential of a single-microphone setup, this study aims to not only provide a single-channel solution that can be more easily adapted to diverse applications, but to also establish a foundational framework for potential multi-channel extensions that exploit similar principles to those considered in the development of the proposed approaches.

In literature, single-channel DOA estimation methods have already been proposed, with some examples including the use of machine learning [40], a circularly moving microphone that exploits the Doppler effect [41] and a time delay-based subspace approach with a single hydrophone [42]. Our previous work [43] introduced the concept of using a single, rotating directional microphone for performing DOA and PSD estimation. The proposed method involved capturing spatially static and localized sound sources with a cardioid microphone, oriented towards different directions for different observation frames, so that changes in the microphone signal power could be analyzed for determining spatial information about the sources generating the observed sound field. By solving a group-sparsity constrained optimization problem while using a broadband signal model and an overcomplete angular dictionary of possible candidate DOAs for the point sources, DOA estimates could be obtained and used to estimate the localized point sources' PSDs, by solving an overdetermined least squares problem with a nonnegativity constraint for each frequency bin separately. The use of group-sparsity has been previously exploited in multi-channel DOA estimation methods, such as in [44], where a covariance matrix estimated from signals captured at multiple co-prime arrays is modeled based on the corresponding steering vectors and point source PSDs. Moreover, estimating PSDs through the solution of a least squares problem has been also performed in [45], in which multiple beamformers are applied to microphone array signals and their outputs are used as the proposed method's observed data. Nonetheless, when employing the mentioned multi-channel DOA and PSD estimation methods, it becomes necessary to assess the performance constraints associated with the microphone array configuration available in the considered scenario. Factors that could potentially influence these limitations include the spacing between microphones and the overall array geometry [36, 46].

In this paper, the method proposed in [43] is extended in two aspects, resulting in two alternative approaches. First, in order to add robustness in scenarios with a mismatch between the signal model and the generated data, we propose to use an additional nonnegativity constraint on the residual noise term when solving the group-sparse optimization problem. We refer to this as the *Group-Lasso Least Squares* (GL-LS) approach. Note that GL-LS is still based on the broadband signal model presented in [43]. Second, in addition to the new non-negativity constraint, we also propose an optimization problem based on a narrowband signal model, where the DOAs and PSDs can be jointly estimated by iterative, group-wise reweighting. This is referred to as the *Group-Lasso with* l_1*-reweighting* (GL-L1) approach. An efficient implementation of both approaches using the alternating direction method of multipliers (ADMM) is presented, as opposed

to the use of CVX [47] as it was done in our previous work [43].

In order to evaluate the performance of the proposed approaches, a series of simulations subject to different types of model mismatch are performed, by introducing off-grid DOAs, non ideal microphone responses, and reverberation for the case of a single point source. We also analyze scenarios involving two point sources of different broadband power and varying angular separation between them. Finally, we consider the case of three point sources when using higher-order directivity patterns. These simulations greatly extend the aspects considered for evaluating the proposed approaches in comparison with those presented in our previous work [43]. We show that most of the times, having the PSD re-estimation step with least squares as performed with GL-LS is beneficial, but also that redesigning the system of equations in a frequency-dependent manner as performed in GL-L1 may help in distinguishing closely located sources.

This paper is structured as follows. In Section 2, we present the signal model. In Section 3, we explain the proposed approaches. In Section 4, we explain the ADMM-based implementation. In Section 5, we present the simulations setup, the results obtained, and discussion. Finally, in Section 6, we conclude with a summary of the work presented and future work.

2 Signal model

2.1 Rotational microphone signal model

In the short-time Fourier transform (STFT) domain, the signal recorded by a single, directional microphone rotating in the horizontal plane is modeled as

$$Y(k,n) = \sum_{p=1}^{P} S_p(k,n) \int_0^{2\pi} a^*(k,\theta-\gamma_n) H_p(k,\theta) d\theta + D(k,n), \tag{1}$$

where $(\cdot)^*$ denotes the complex conjugate, k the discrete frequency index, n the observation frame index, θ the look direction, and γ_n the microphone orientation at frame n. We assume that there are a total of P point sources in the far field and that P is known. The expression in (1) describes that the resulting microphone signal $Y(k, n)$ is composed of the sum of the P point source signals $S_p(k,n)$ arriving from P distinct directions, ϑ_1 to ϑ_P, weighted by the direction-dependent microphone response $a(k,\theta-\gamma_n)$, relative to the microphone orientation γ_n, and by the room transfer function (RTF) $H_p(k,\theta)$ from the p-th far-field source to the microphone, added to diffuse or sensor noise $D(k, n)$. If we consider that the recording is performed in anechoic conditions, then $H_p(k,\theta) = \delta(\theta-\vartheta_p)$, $\forall k$ and for $p = 1,\ldots,P$, such that the expression in (1) is reduced to

$$Y(k,n) = \sum_{p=1}^{P} a^*(k,\vartheta_p-\gamma_n) S_p(k,n) + D(k,n). \tag{2}$$

Assuming that the source signals are uncorrelated and stationary during the entire observation, such that their PSDs remain constant across different time frames, and that the microphone response is real-valued, the microphone signal PSD $\phi_Y(k,n)$ can be described as follows:

$$\phi_Y(k,n) = \sum_{p=1}^{P} |a(k,\vartheta_p-\gamma_n)|^2 \phi_{S_p}(k) + \phi_D(k,n), \tag{3}$$

where $\phi_D(k,n)$ is the noise PSD for frequency k and time frame n, and $\phi_{S_p}(k)$ is the PSD for frequency k corresponding to the p-th source at position ϑ_p. As the directional microphone is oriented towards different directions γ_n for different observation frames n, the relative positions of the sound sources with respect to the microphone do not remain the same, and consequently their PSD values $\phi_{S_p}(k)$ are multiplied with different squared microphone response coefficients $|a(k,\vartheta_p-\gamma_n)|^2$ over different time frames.

2.2 Grid-based system of equations

If we assume that measurements of $\phi_Y(k,n)$ are available for multiple time frames, with $n = 1, \ldots, N$, a linear system of equations can be built for estimating the point source DOAs ϑ_p and PSDs $\phi_{S_p}(k)$, for $p = 1, \ldots, P$ and $k = 1, \ldots, K$, where K denotes the number of frequency bins. As the directional weighting factor $|a(k,\vartheta_p-\gamma_n)|^2$ in (3) depends on the unknown source DOA, we use a grid of candidate positions defined between 0 and 2π and an overcomplete dictionary of corresponding microphone response coefficients in order to build the linear system of equations, as previously proposed in [43].

For a single candidate angle, denoted as θ_l, a vector $\boldsymbol{\varphi}_{S,\theta_l}$ with PSD values for all frequencies is defined as

$$\boldsymbol{\varphi}_{S,\theta_l} = [\phi_S(1,\theta_l) \quad \ldots \quad \phi_S(K,\theta_l)]^\top, \tag{4}$$

where $(\cdot)^\top$ denotes the transpose and $\phi_S(k,\theta_l)$ is the PSD at candidate position θ_l for frequency k. We stack different $\boldsymbol{\varphi}_{S,\theta_l}$ vectors for all candidate DOAs from a given L-element angular grid, with $P \ll L$ and $N < L$, in order to obtain the vector we ultimately aim to estimate as

$$\boldsymbol{\varphi}_S = \left[\boldsymbol{\varphi}_{S,\theta_1}^\top \ \ldots \ \boldsymbol{\varphi}_{S,\theta_L}^\top\right]^\top, \tag{5}$$

which will be used to obtain $\hat{\vartheta}_p$ and $\hat{\phi}_{S_p}(k)$ such that $\hat{\vartheta}_p \in \{\theta_1, \ldots, \theta_L\}$, for $p = 1, \ldots, P$.

One possible construction of the linear system of equations is defined as follows. Firstly, a vector $\boldsymbol{\varphi}_{Y,n}$ containing the PSD values for the microphone observation frame n is defined as

$$\boldsymbol{\varphi}_{Y,n} = [\phi_Y(1,n) \ \ldots \ \phi_Y(K,n)]^\top. \tag{6}$$

By stacking different $\boldsymbol{\varphi}_{Y,n}$ vectors for all observations $n = 1$ to $n = N$, we obtain a vector $\boldsymbol{\varphi}_Y$ as

$$\boldsymbol{\varphi}_Y = \left[\boldsymbol{\varphi}_{Y,1}^\top \ \ldots \ \boldsymbol{\varphi}_{Y,N}^\top\right]^\top. \tag{7}$$

Similarly, we also define the vectors for the diffuse component as

$$\boldsymbol{\varphi}_{D,n} = [\phi_D(1,n) \ \ldots \ \phi_D(K,n)]^\top, \tag{8}$$

$$\boldsymbol{\varphi}_D = \left[\boldsymbol{\varphi}_{D,1}^\top \ \ldots \ \boldsymbol{\varphi}_{D,N}^\top\right]^\top. \tag{9}$$

In order to build an overcomplete microphone response matrix involving all possible candidate DOAs, we firstly define a vector containing the squared, microphone response for a candidate angle θ_l relative to the microphone orientation γ_n as

$$\mathbf{a}(\theta_l - \gamma_n) = \left[|a(1,\theta_l - \gamma_n)|^2 \ \ldots \ |a(K,\theta_l - \gamma_n)|^2\right]. \tag{10}$$

Hence, the microphone response matrix is defined as

$$\bar{\mathbf{A}} = \begin{bmatrix} \mathbf{A}_{1,1} & \ldots & \mathbf{A}_{1,L} \\ \vdots & \ddots & \vdots \\ \mathbf{A}_{N,1} & \ldots & \mathbf{A}_{N,L} \end{bmatrix}, \tag{11}$$

where $\mathbf{A}_{n,l} = \mathrm{diag}(\mathbf{a}(\theta_l - \gamma_n))$ and $\mathrm{diag}(\cdot)$ denotes the creation of a diagonal matrix from the elements of a given vector. The linear system of equations can then be written as

$$\boldsymbol{\varphi}_Y = \bar{\mathbf{A}}\boldsymbol{\varphi}_S + \boldsymbol{\varphi}_D, \tag{12}$$

which corresponds to a matrix representation of (3) for different observation frames, and which we refer to as the narrowband system of equations in the remainder.

By ensuring that $\gamma_1 \neq \gamma_2 \neq \cdots \neq \gamma_N$, with $0 \leq \gamma_n \leq 2\pi$, $\forall\, n$, and assuming that $\boldsymbol{\varphi}_Y$ and $\bar{\mathbf{A}}$ are known, source localization in terms of DOA estimation can be achieved by solving the proposed linear system of equations. From the estimated vector $\hat{\boldsymbol{\varphi}}_S$, we can identify in which direction θ_l within the angular grid there are peaks in power, indicating the point source DOAs $\hat{\vartheta}_p$ and their PSDs $\hat{\phi}_{S_p}(k)$, assuming that $\hat{\vartheta}_p \in \{\theta_1, \ldots, \theta_L\}$.

While this narrowband system of equations has all frequency bins decoupled from each other, it is also possible to build a broadband version of it. By considering an $N \times N$ identity matrix denoted as $\mathbf{I}_N$ and a K-element column vector of ones denoted as $\mathbf{1}_K$, we define the broadband vectors and matrix:

$$\widetilde{\boldsymbol{\varphi}}_Y = \left(\mathbf{I}_N \otimes \mathbf{1}_K^\top\right)\boldsymbol{\varphi}_Y, \tag{13}$$

$$\widetilde{\boldsymbol{\varphi}}_D = \left(\mathbf{I}_N \otimes \mathbf{1}_K^\top\right)\boldsymbol{\varphi}_D, \tag{14}$$

$$\widetilde{\mathbf{A}} = \left(\mathbf{I}_N \otimes \mathbf{1}_K^\top\right)\bar{\mathbf{A}}, \tag{15}$$

where $\otimes$ denotes the Kronecker product. This operation over the previously defined vectors $\boldsymbol{\varphi}_Y$ and $\boldsymbol{\varphi}_D$ results in the sum over all frequency bins of the PSD values from the microphone signal and the diffuse component, respectively. It also results in a "dimensionality reduction" of the original matrix $\bar{\mathbf{A}}$ such that $\mathbf{A}_{n,l}$ is replaced by $\mathbf{a}(\theta_l - \gamma_n)$ stacked together accordingly. Using (13)–(15), the linear system of equations can be written as

$$\widetilde{\boldsymbol{\varphi}}_Y = \widetilde{\mathbf{A}}\boldsymbol{\varphi}_S + \widetilde{\boldsymbol{\varphi}}_D, \tag{16}$$

which we refer to as the broadband system of equations in the remainder, and which is equivalent to the system of equations used in [43]. Like the narrowband system presented in (12), the broadband system in (16) can be solved for estimating the point source DOAs, with the caveat of having lost the frequency-dependent information in the microphone observation vector by summing the target input PSD vector over all frequencies. Consequently, an additional PSD re-estimation step is required when using the broadband system of equations as will be explained further in the following section.

3 Proposed approaches

In this section, we explain the broadband and narrowband approach proposed for estimating the DOAs and PSDs of point sources modeled as previously described in Section 2. The first approach involves the solution to a group-sparse constrained optimization problem for estimating the DOAs, constructed based on the

broadband system of equations in (16), followed by a least-squares step for re-estimating the point sources' PSDs. This approach is here named GL-LS. The second approach, involving the solution to a group-sparse constrained optimization problem based on the narrowband system of equations in (12), considers an iterative process of reweighting the group-sparsity penalty present in the formulation for jointly estimating the DOAs and PSDs of the point sources. This approach is here named GL-L1.

3.1 Group Lasso followed by least squares (GL-LS)

Assuming that $N < KL$, the linear system in (16) is underdetermined. Therefore, the following Group Lasso [48] optimization problem is proposed to be solved:

$$\underset{\boldsymbol{\varphi}_S}{\text{minimize}} \quad \frac{1}{2}\left\|\widetilde{\boldsymbol{\varphi}}_Y - \widetilde{\mathbf{A}}\boldsymbol{\varphi}_S\right\|_2^2 + \lambda \sum_{l=1}^{L} \left\|\boldsymbol{\varphi}_{S,\theta_l}\right\|_2 \tag{17a}$$

$$\text{subject to} \quad \boldsymbol{\varphi}_S \geq 0 \tag{17b}$$

$$\widetilde{\boldsymbol{\varphi}}_Y - \widetilde{\mathbf{A}}\boldsymbol{\varphi}_S \geq 0 \tag{17c}$$

The nonnegativity constraint in (17b) is necessary for complying with the intrinsic nonnegativity property of PSD values [49], whereas the nonnegativity constraint over the error term, as formulated in (17c), is included as a means to model the noise signal $\phi_D(k,n)$ present in the microphone observations, such that more robustness can be achieved in case of a possible model mismatch, for instance between the assumed matrix $\widetilde{\mathbf{A}}$ and the actual microphone response. The Group Lasso formulation includes the regularization term $\lambda \sum_{l=1}^{L} \left\|\boldsymbol{\varphi}_{S,\theta_l}\right\|_2$, which enforces sparsity between so-called different groups [48]. When assuming that only a limited number of point sources are present in space ($P \ll L$), using this group-sparsity penalty is a way of ensuring that only a few of the subvectors $\boldsymbol{\varphi}_{S,\theta_l}$ composing $\boldsymbol{\varphi}_S$ will be activated, that is, have a magnitude significantly different from zero. After solving the optimization problem (17a)-(17c) and obtaining $\hat{\boldsymbol{\varphi}}_S$, and therefore, $\hat{\boldsymbol{\varphi}}_{S,\theta_1}, \ldots, \hat{\boldsymbol{\varphi}}_{S,\theta_L}$, the PSD values can be averaged over the K frequency bins for each of the L candidate directions, allowing the point source DOAs to be estimated by finding the indices of θ_l for which there are peaks in the average PSD. For a total of P sources assumed to be present, P peaks should then be identified.

One may note that the resulting PSD estimates for all directions in the angular grid obtained from solving (17a)-(17c) will be inherently biased, due to the group-sparsity penalty included in the optimization problem [50]. Moreover, the summation over frequency of the PSD values present in $\widetilde{\boldsymbol{\varphi}}_Y$, see (13), results in the loss of frequency-dependent information on the detected point sources' PSDs. These limiting factors motivate the use of a re-estimation step for the PSD values using the estimated DOAs, as previously proposed in [43].

The new PSD vectors are defined as follows:

$$\boldsymbol{\phi}_Y(k) = [\phi_Y(k,1) \quad \ldots \phi_Y(k,N)]^\top, \tag{18}$$

$$\boldsymbol{\phi}_S(k) = \left[\phi_{S_1}(k) \ldots \phi_{S_P}(k)\right]^\top, \tag{19}$$

$$\boldsymbol{\phi}_D(k) = [\phi_D(k,1) \ldots \phi_D(k,N)]^\top. \tag{20}$$

We define a new matrix $\mathbf{A}_S(k) \in \mathbb{R}^{N\times P}$ which now contains squared microphone response coefficients for only the directions $\hat{\vartheta}_1, \ldots, \hat{\vartheta}_P \in \{\theta_1, \ldots, \theta_L\}$ where the P sources are assumed to be located, based on the preceding DOA estimation:

$$\mathbf{A}_S(k) = \begin{bmatrix} |a(k,\hat{\vartheta}_1 - \gamma_1)|^2 & \ldots & |a(k,\hat{\vartheta}_P - \gamma_1)|^2 \\ \vdots & & \vdots \\ |a(k,\hat{\vartheta}_1 - \gamma_N)|^2 & \ldots & |a(k,\hat{\vartheta}_P - \gamma_N)|^2 \end{bmatrix}. \tag{21}$$

Using the PSD signal model in (3), a new linear system of equations for the microphone signal PSD is then formulated, for each frequency bin, as

$$\boldsymbol{\phi}_Y(k) = \mathbf{A}_S(k)\boldsymbol{\phi}_S(k) + \boldsymbol{\phi}_D(k). \tag{22}$$

If $P \leq N$, a constrained least-squares approach can be used for solving the overdetermined linear system and estimating the PSD values of the point sources:

$$\underset{\boldsymbol{\phi}_S(k)}{\text{minimize}} \quad \frac{1}{2}\left\|\boldsymbol{\phi}_Y(k) - \mathbf{A}_S(k)\boldsymbol{\phi}_S(k)\right\|_2^2 \tag{23a}$$

$$\text{subject to} \quad \boldsymbol{\phi}_S(k) \geq 0 \tag{23b}$$

Hence, in this re-estimation step, we avoid the bias induced by the Group Lasso formulation presented in the DOA estimation step and allow for a more accurate PSD estimation for the stationary point sources.

When compared to our previous work [43], the GL-LS approach described in this subsection presents an extension of the previous idea of solving a group-sparse constrained optimization problem for estimating DOAs and PSDs of point sources by including the additional nonnegativity constraint over the error term, in order to provide robustness against model mismatches that may be present.

3.2 Group Lasso with l_1-reweighting (GL-L1)

As an alternative to the method proposed in Section 3.1, we consider employing the narrowband input vector $\boldsymbol{\varphi}_Y$

and response matrix $\bar{\mathbf{A}}$ (see the model in (12)) and construct the following optimization problem:

$$\underset{\boldsymbol{\varphi}_S}{\text{minimize}} \quad \frac{1}{2}\left\|\boldsymbol{\varphi}_Y - \bar{\mathbf{A}}\boldsymbol{\varphi}_S\right\|_2^2 + \lambda \sum_{l=1}^{L} w_l \left\|\boldsymbol{\varphi}_{S,\theta_l}\right\|_2 \tag{24a}$$

$$\text{subject to} \quad \boldsymbol{\varphi}_S \geq 0 \tag{24b}$$

$$\boldsymbol{\varphi}_Y - \bar{\mathbf{A}}\boldsymbol{\varphi}_S \geq 0 \tag{24c}$$

In this approach, we include the use of group weights, denoted w_l for $l = 1, \ldots, L$, in order to allow for a group-wise, iterative reweighting process aimed at enhancing sparsity within the final solution obtained [51, 52]. This is motivated by the goal of jointly estimating the point source DOAs and PSDs from the solution to the constrained optimization problem, without performing a least-squares re-estimation step as in the GL-LS approach, in which the optimization problem in (17) is solved only for estimating the source's DOAs.

The group sparsity is enhanced by successively solving the optimization problem in (24), while updating, between iterations, the sparsity penalization of each estimated group separately as a function of their corresponding norms [51, 53, 54]. The weight updates can be expressed as

$$w_l^{(i+1)} = 1/(\|\hat{\boldsymbol{\varphi}}_{S,\theta_l}^{(i)}\|_2 + \epsilon) \quad \text{for} \quad l = 1, \ldots, L, \tag{25}$$

where i denotes the reweighting iteration index, and $\epsilon > 0$ is a fixed parameter introduced for avoiding a division by zero.

After reaching convergence from the group-wise iterative reweighting procedure, i.e., repeatedly re-estimating $\hat{\boldsymbol{\varphi}}_{S,\theta_l}$ while updating the sparsity penalty weights for different groups individually, the point source DOA estimation is performed in a similar fashion as proposed in Section 3.1. Assuming a total number of P point sources, the P largest peaks in frequency-averaged power obtained from $\hat{\boldsymbol{\varphi}}_S$ are picked, and the estimated DOAs $\hat{\vartheta}_p$ are obtained from the indices of θ_l for the corresponding groups selected.

Simultaneously, the point source PSDs are obtained as the estimated values of the corresponding subvectors contained in $\hat{\boldsymbol{\varphi}}_S$ with angles $\hat{\vartheta}_p$, for $p = 1, \ldots, P$. Hence, in this approach, the PSD estimates are obtained jointly with the DOA estimates, without performing an additional least-squares re-estimation step as in the proposed GL-LS approach.

The GL-L1 approach described in this subsection presents an extension of our previous work [43] in more aspects than GL-LS, namely the employment of the additional nonnegativity constraint over the error term, the use of the narrowband linear system of equations, and the iterative group reweighting process for jointly estimating DOAs and PSDs without a least-squares re-estimation step.

4 Implementation

In order to solve the optimization problems defined in (17) and (24) in an efficient way, we employ the alternating direction method of multipliers (ADMM) algorithm [55], as opposed to the use of CVX [47] as it was done in our previous work [43]. Since in both proposed approaches the vector $\boldsymbol{\varphi}_S$ to be estimated is the same, and their distinction only consists in the construction of the input vectors and matrices employed, as well as in the use of group weights, we firstly describe the used ADMM implementation in general terms and then further clarify each method's full algorithm afterwards.

Let a general optimization problem involving the same group sparsity and nonnegativity constraints present in (17) and (24) be described as

$$\underset{\mathbf{x}}{\text{minimize}} \quad \frac{1}{2}\|\mathbf{b} - \mathbf{A}\mathbf{x}\|_2^2 + \lambda \sum_{l=1}^{L} w_l \|\mathbf{x}_l\|_2 \tag{26a}$$

$$\text{subject to} \quad \mathbf{x} \geq 0 \tag{26b}$$

$$\mathbf{b} - \mathbf{A}\mathbf{x} \geq 0 \tag{26c}$$

where $\mathbf{x}_l$ corresponds to the l-th group composing the vector $\mathbf{x}$. We can use auxiliary variables, denoted $\mathbf{u}_1$ and $\mathbf{u}_2$, to rewrite the problem as

$$\underset{\mathbf{u}_1 \geq 0,\ \mathbf{u}_2 \geq 0}{\text{minimize}} \quad \frac{1}{2}\|\mathbf{u}_1\|_2^2 + \lambda \sum_{l=1}^{L} w_l \left\|\mathbf{u}_{2,l}\right\|_2 \tag{27a}$$

$$\text{subject to} \quad \mathbf{u}_1 = \mathbf{b} - \mathbf{A}\mathbf{x} \tag{27b}$$

$$\mathbf{u}_2 = \mathbf{x} \tag{27c}$$

where $\mathbf{u}_{2,l}$ corresponds to the l-th group composing the vector $\mathbf{u}_2$, such that it can be represented as

$$\mathbf{u}_2 = \begin{bmatrix} \mathbf{u}_{2,1}^\top & \ldots & \mathbf{u}_{2,L}^\top \end{bmatrix}^\top. \tag{28}$$

The augmented Lagrangian [55] of the optimization problem in (27) can be written as

$$\mathcal{L}_\rho(\mathbf{x},\mathbf{u}_1,\mathbf{u}_2,\mathbf{d}_1,\mathbf{d}_2) = \frac{1}{2}\|\mathbf{u}_1\|_2^2 + \lambda\sum_{l=1}^{L} w_l\|\mathbf{u}_{2,l}\|_2 + \frac{\rho}{2}\|\mathbf{b}-\mathbf{A}\mathbf{x}-\mathbf{u}_1-\mathbf{d}_1\|_2^2 + \frac{\rho}{2}\|\mathbf{x}-\mathbf{u}_2-\mathbf{d}_2\|_2^2 \tag{29}$$

where $\mathbf{d}_1$ and $\mathbf{d}_2$ are the dual variables, and ρ can be interpreted as a dual update step size [55].

Considering an ADMM iteration indexed as j, we define the following short hands:

$$\boldsymbol{\zeta}_1(j) = \mathbf{b}-\mathbf{A}\mathbf{x}(j+1)-\mathbf{d}_1(j), \tag{30}$$

$$\boldsymbol{\zeta}_2(j) = \mathbf{x}(j+1)-\mathbf{d}_2(j). \tag{31}$$

The updates for each variable in (29) are

$$\mathbf{x}(j+1) = \arg\min_{\mathbf{x}} \ \mathcal{L}_\rho(\mathbf{x},\mathbf{u}_1,\mathbf{u}_2,\mathbf{d}_1,\mathbf{d}_2), \tag{32}$$

$$\mathbf{u}_1(j+1) = \arg\min_{\mathbf{u}_1\geq 0} \ \frac{1}{2}\|\mathbf{u}_1\|_2^2 + \frac{\rho}{2}\|\boldsymbol{\zeta}_1(j)-\mathbf{u}_1\|_2^2, \tag{33}$$

$$\mathbf{u}_2(j+1) = \arg\min_{\mathbf{u}_2\geq 0} \ \lambda\sum_{l=1}^{L} w_l\|\mathbf{u}_{2,l}\|_2 + \frac{\rho}{2}\|\boldsymbol{\zeta}_2(j)-\mathbf{u}_2\|_2^2, \tag{34}$$

$$\mathbf{d}_1(j+1) = \mathbf{u}_1(j+1)-\boldsymbol{\zeta}_1(j), \tag{35}$$

$$\mathbf{d}_2(j+1) = \mathbf{u}_2(j+1)-\boldsymbol{\zeta}_2(j). \tag{36}$$

The first update in (32) is computed as

$$\mathbf{x}(j+1) = \left(\mathbf{A}^\top\mathbf{A}+\mathbf{I}\right)^{-1}[\mathbf{A}^\top(\mathbf{b}-\mathbf{u}_1(j)-\mathbf{d}_1(j)) + \mathbf{u}_2(j)+\mathbf{d}_2(j)]. \tag{37}$$

The $\mathbf{u}_1$ update in (33) takes into account its nonnegativity constraint and is computed as

$$\mathbf{u}_1(j+1) = \max\left(0, \frac{\rho\boldsymbol{\zeta}_1(j)}{1+\rho}\right), \tag{38}$$

where $\max(\cdot)$ denotes the element-wise max operator.

The update of $\mathbf{u}_2$ in (34) is computed as

$$\mathbf{u}_{2,l}(j+1) = \mathrm{T}\left(\max(\boldsymbol{\zeta}_{2,l}(j),0), w_l\frac{\lambda}{\rho}\right), \tag{39}$$

for $l = 1,\ldots,L$, where $\mathrm{T}(\cdot)$ represents a group-wise shrinkage function, defined as

$$\mathrm{T}(\mathbf{z},\kappa) = \left[\max\left(1-\frac{\kappa}{\|\mathbf{z}\|_2},0\right)\right]\mathbf{z}. \tag{40}$$

After computing the updates of $\mathbf{d}_1$ and $\mathbf{d}_2$ as in (35) and (36), respectively, the whole iterative process is repeated until convergence [55]. For more detailed derivations of each ADMM update equation, we refer to [55].

As previously mentioned, the difference in implementation between the proposed methods GL-LS and GL-L1 relies on how the input data structure for the ADMM algorithm here explained is chosen. In the case of GL-LS, we use $\widetilde{\boldsymbol{\varphi}}_Y$ and $\widetilde{\mathbf{A}}$ as input vector and response matrix, respectively, and we set all group weights w_l equal to one. After running the ADMM scheme once, the result is used for estimating the source DOAs and re-estimating the source PSDs as explained in Section 3.1. In the case of GL-L1, we use $\boldsymbol{\varphi}_Y$ and $\bar{\mathbf{A}}$ as input, the weights are also initially set equal to one, and the ADMM scheme is repeated until convergence while re-updating the group weights as described in Section 3.2, so that a final joint DOA and PSD estimation is obtained. A summary of the GL-LS implementation is presented in Algorithm 1, while a summary of the GL-L1 implementation is presented in Algorithm 2.

Regarding the computational complexity of each proposed approach, numerous factors will affect the overall cost of both GL-LS and GL-L1. Firstly, when analyzing a single iteration within the ADMM scheme, we can observe that the most computationally demanding update occurs in (37), and that its cost will depend on the dimensions of the response matrix being employed [55]. The use of the wideband signal model by GL-LS and the narrowband signal model by GL-L1 results in an assymptotic complexity of $\mathcal{O}(NKL)$ and $\mathcal{O}(NK^2L)$, respectively, for each of the proposed approaches. In the case of GL-LS, the cost of the additional PSD re-estimation step through the solution of a least squares problem will assymptotically be $\mathcal{O}(N^3)$. Finally, the total runtime experienced by each proposed approach will depend on the convergence criterion selected for the ADMM scheme and the number of iterations required for satisfying it. Therefore, even though the use of the proposed approaches may significantly differ in overall computational cost depending on the number of microphone orientations N being considered and setup conditions that can affect the convergence rate, the GL-LS approach is currently more computationally efficient than the GL-L1 approach. One may note, however, that the sparse structure of the matrix employed in the narrowband signal model, see (11), allows for using sparsity-aware methods for matrix computations [56, 57] that can potentially reduce the complexity of GL-L1.

initiate $\mathbf{A} = \widetilde{\mathbf{A}}$, $\mathbf{b} = \widetilde{\boldsymbol{\varphi}}_Y$
  $\hat{\boldsymbol{\varphi}}_S = \mathbf{0}^{KL\times 1}$
  $w_l = 1$ for $l = 1, \ldots, L$
initiate $j := 0$, $\mathbf{u}_2(0) = \hat{\boldsymbol{\varphi}}_S$,
  $\mathbf{x}(0) = \mathbf{d}_1(0) = \mathbf{d}_2(0) = \mathbf{0}^{KL\times 1}$
repeat {ADMM scheme}
  update $\mathbf{x}(j+1)$ as in (37)
  update $\mathbf{u}_1\big(j+1\big)$ as in (38)
  for $l \leftarrow 1$ to L **do**
    update $\mathbf{u}_{2,l}$ as in (39)
  end for
  $\mathbf{d}_1(j+1) = \mathbf{u}_1(j+1) - \boldsymbol{\zeta}_1(j)$
  $\mathbf{d}_2(j+1) = \mathbf{u}_2(j+1) - \boldsymbol{\zeta}_2(j)$
  $j \leftarrow j+1$
until convergence
store $\hat{\boldsymbol{\varphi}}_S = \mathbf{u}_2(\text{end})$
estimate $\hat{\vartheta}_p$ for $p = 1, \ldots, P$ as the P largest peaks averaged over frequency in $\hat{\boldsymbol{\varphi}}_S$
estimate $\hat{\boldsymbol{\phi}}_S(k)$ in (23) for $k = 1, \ldots, K$

Algorithm 1 The proposed GL-LS algorithm

initiate $i := 0$, $\mathbf{A} = \bar{\mathbf{A}}$, $\mathbf{b} = \boldsymbol{\varphi}_Y$,
  $\hat{\boldsymbol{\varphi}}_S^{(0)} = \mathbf{x}_{\text{save}} = \mathbf{d}_{1,\text{save}} = \mathbf{d}_{2,\text{save}} = \mathbf{0}^{KL\times 1}$,
  $w_l^{(0)} = 1$ for $l = 1, \ldots, L$
repeat {reweighting scheme}
  initiate $j := 0$, $\mathbf{u}_2(0) = \hat{\boldsymbol{\varphi}}_S^{(i)}$,
    $\mathbf{x}(0) = \mathbf{x}_{\text{save}}$, $\mathbf{d}_1(0) = \mathbf{d}_{1,\text{save}}$, $\mathbf{d}_2(0) = \mathbf{d}_{2,\text{save}}$
  repeat {ADMM scheme}
    update $\mathbf{x}(j+1)$ as in (37)
    update $\mathbf{u}_1\big(j+1\big)$ as in (38)
    for $l \leftarrow 1$ to L **do**
      update $\mathbf{u}_{2,l}$ as in (39)
    end for
    $\mathbf{d}_1(j+1) = \mathbf{u}_1(j+1) - \boldsymbol{\zeta}_1(j)$
    $\mathbf{d}_2(j+1) = \mathbf{u}_2(j+1) - \boldsymbol{\zeta}_2(j)$
    $j \leftarrow j+1$
  until convergence
  store $\hat{\boldsymbol{\varphi}}_S^{(i)} = \mathbf{u}_2(\text{end})$, $\mathbf{x}_{\text{save}} = \mathbf{x}(\text{end})$, $\mathbf{d}_{1,\text{save}} = \mathbf{d}_1(\text{end})$, $\mathbf{d}_{2,\text{save}} = \mathbf{d}_2(\text{end})$
  update $w_l^{(i+1)} = 1/(\|\hat{\boldsymbol{\varphi}}_{S,\theta_l}^{(i)}\|_2 + \epsilon)$ for $l = 1, \ldots, L$
  $i \leftarrow i+1$
until convergence
estimate $\hat{\vartheta}_p$ for $p = 1, \ldots, P$ as the P largest peaks averaged over frequency in $\hat{\boldsymbol{\varphi}}_S$
store $\hat{\boldsymbol{\phi}}_S(k)$ based on $\hat{\boldsymbol{\varphi}}_S$ and $\hat{\vartheta}_p$

Algorithm 2 The proposed GL-L1 algorithm

5 Simulations

In order to evaluate the performance of both proposed approaches in terms of DOA and PSD estimation, simulations are done in MATLAB considering different setups. We aim to observe the advantages or disadvantages of using GL-LS or GL-L1 under different conditions, ranging from different types of model mismatch in Sections 5.2 to 5.3 to variations of the acoustic scene in Sections 5.4 to 5.5 and of the microphone directivity pattern in Section 5.6.

In Table 1, a summary of the main parameters used in each subsection is presented. For all simulations, the sampling frequency is 16 kHz, the source signals are stationary, and the microphone signal PSD $\phi_Y(k,n)$ is estimated with Welch's method, using a 512-point Hann window corresponding to a length of 32 ms and 50% overlap over a duration of 500 ms for each microphone orientation γ_n, and therefore, each observation frame n. For a total of N observations, N different microphone orientations uniformly distributed over 360° are simulated. The L candidate directions used for building the microphone response matrices are defined according to a uniformly spaced angular grid given a certain resolution in degrees. The signal-to-noise ratio (SNR) is defined as $\text{SNR} = \sigma_{S_1}^2/\sigma_D^2$, where $\sigma_{S_1}^2$ denotes the broadband power of the first source ($p = 1$), and σ_D^2 denotes the broadband power of the diffuse component. The regularization parameter λ is heuristically set as a function of $\|\widetilde{\mathbf{A}}^\top \widetilde{\boldsymbol{\varphi}}_Y\|_\infty$ and $\|\bar{\mathbf{A}}^\top \boldsymbol{\varphi}_Y\|_\infty$ for the GL-LS and GL-L1 approaches, respectively, with $\|\cdot\|_\infty$ denoting the l_∞-norm. The choice of λ may be suboptimal; however, it is motivated by the objective of obtaining solutions with both approaches that are generalizable for all different scenarios tested. In the implementation of the GL-LS and GL-L1 approaches, the weights are initialized as $w_l = 1$, for $l = 1, \ldots, L$, and in the case of GL-L1, the reweighting process is performed twice, resulting in three repetitions of the ADMM scheme. For each scenario considered, a total number of 100 Monte Carlo realizations are simulated. With the exception of Section 5.4, all scenarios are simulated under anechoic conditions, and with the exception of Section 5.5, the simulated source signals correspond to speech-shaped noise, obtained by filtering white Gaussian noise with a 16th order linear prediction filter based on a male speech signal from [58].

In Section 5.1, the performance measures used to evaluate the proposed approaches are defined. In Sections 5.2 and 5.3, we present simulations in which a single point source is present, whereas in Sections 5.5 and 5.6, the performance is evaluated in scenarios with 2 and 3 sources, respectively. The parameters varied for each simulated setup are further explained in each corresponding subsection.

5.1 Performance measures

The DOA estimation is evaluated by computing the mean absolute error (MAE) between the estimated DOA and the true source DOA, for each point source p separately, and can be expressed as

$$\text{MAE}_p = \frac{1}{N_r}\sum_{r=1}^{N_r} |\vartheta_p^r - \hat{\vartheta}_p^r|, \tag{41}$$

where N_r denotes the total number of Monte Carlo realizations and r is the realization index.

The point source PSD estimation is evaluated by computing the normalized mean squared error (NMSE), for each point source separately, as

$$\text{NMSE}_p = \frac{1}{N_r}\sum_{r=1}^{N_r} \frac{\sum_{k=1}^{K}(\phi_{S_p}^r(k) - \hat{\phi}_{S_p}^r(k))^2}{\sum_{k=1}^{K}(\phi_{S_p}^r(k))^2}. \tag{42}$$

5.2 Different grid resolutions

Firstly, we consider a simulation setup for evaluating the performance of each method proposed as a function of the grid resolution selected for building the microphone response matrices, containing all candidate positions for a single point source. In addition, different SNR values are used to evaluate the methods' robustness to additive diffuse noise.

Table 1 Main simulation parameters for each subsection

Subsection	P	N	Grid resolution	SNR	λ (GL − LS, GL − L1)
5.2	1	6	[1°, 5°, 10°, 15°, 20°, 30°, 40°]	[0 dB, 3 dB, 5 dB, 10 dB, 15 dB, 20 dB]	$0.1\|\widetilde{\mathbf{A}}^\top\widetilde{\boldsymbol{\varphi}}_Y\|_\infty$, $0.1\|\bar{\mathbf{A}}^\top\boldsymbol{\varphi}_Y\|_\infty$
5.3	1	6	10°	10 dB	$0.1\|\widetilde{\mathbf{A}}^\top\widetilde{\boldsymbol{\varphi}}_Y\|_\infty, 0.1\|\bar{\mathbf{A}}^\top\boldsymbol{\varphi}_Y\|_\infty$
5.4	1	6	10°	10 dB	$0.1\|\widetilde{\mathbf{A}}^\top\widetilde{\boldsymbol{\varphi}}_Y\|_\infty, 0.1\|\bar{\mathbf{A}}^\top\boldsymbol{\varphi}_Y\|_\infty$
5.5	2	6	10°	10 dB	$0.005\|\widetilde{\mathbf{A}}^\top\widetilde{\boldsymbol{\varphi}}_Y\|_\infty, 0.005\|\bar{\mathbf{A}}^\top\boldsymbol{\varphi}_Y\|_\infty$
5.6	3	9	10°	10 dB	$0.005\|\widetilde{\mathbf{A}}^\top\widetilde{\boldsymbol{\varphi}}_Y\|_\infty$, $0.005\|\bar{\mathbf{A}}^\top\boldsymbol{\varphi}_Y\|_\infty$

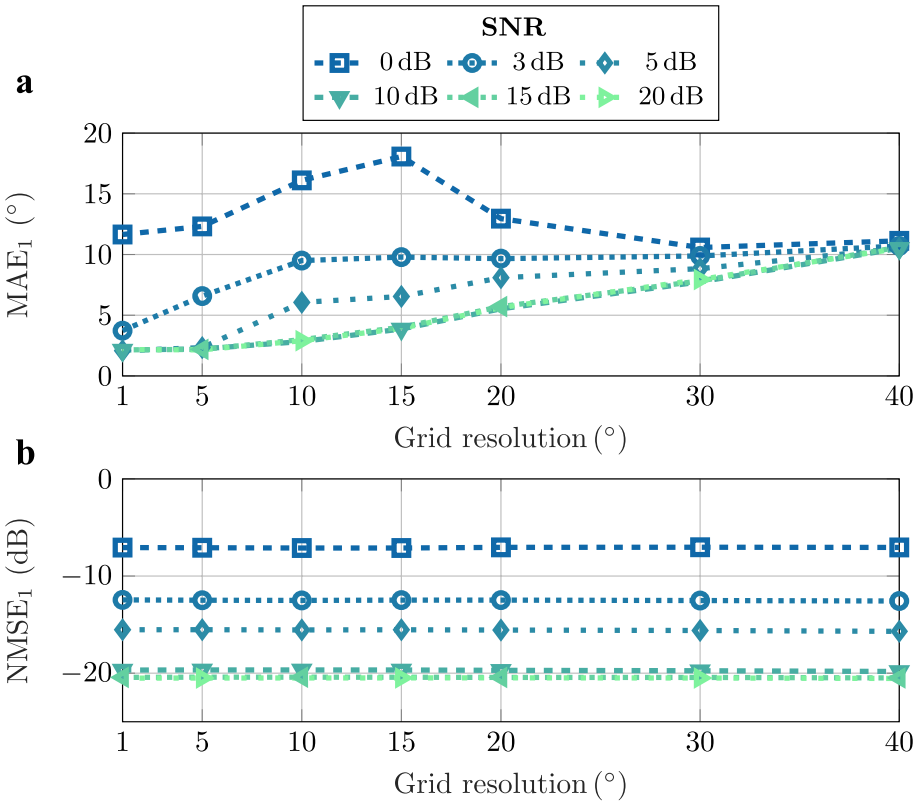

Fig. 1 Estimation errors obtained with GL-LS for different grid resolutions and SNRs

A single point source emitting stationary speech-shaped noise of variance $\sigma_{S_1}^2$ is simulated in anechoic conditions, with its DOA being randomly generated between 0° and 360° for each Monte Carlo realization, yielding the possibility of the source position being on or off-grid. A total of $N = 6$ observation frames are used, with the microphone orientations uniformly distributed over 360° (i.e., 0°, 60°, 120°, 180°, 240°, and 300°) for building the linear systems of equations, i.e., (16) and (12) for GL-LS and GL-L1, respectively. We assume the observations are made with an ideal cardioid microphone with flat frequency response, defined as $a_{\text{cardioid}}(k, \theta) = 0.5 + 0.5\cos(\theta)$, $\forall k$, and that the microphone is static during each observation frame. The diffuse component is white Gaussian noise, with variance σ_D^2. The grid resolution is varied from 1° to 40°, and the SNR is varied from 0 to 0 dB. The regularization parameter λ is heuristically set to $0.1\|\widetilde{\mathbf{A}}^\top \widetilde{\boldsymbol{\varphi}}_Y\|_\infty$ and $0.1\|\bar{\mathbf{A}}^\top \boldsymbol{\varphi}_Y\|_\infty$ for GL-LS and GL-L1, respectively.

The estimation errors obtained when using GL-LS and GL-L1 for all combinations of grid resolution and SNR considered are presented in Figs. 1 and 2, respectively. When analyzing the DOA estimation in terms of MAE, we can observe that for both methods, the performance seems to converge to a certain minimum achievable error corresponding to a quarter of the grid resolution with an increase in SNR, which is a result of both methods picking the closest grid point to the point source's actual position. In that case, the MAE linearly increases as the grid resolution varies from 5° to 40°, whereas for the case of a one-degree resolution, the error indicates a possible limitation as a function of the angular variation of the cardioid directivity pattern in estimating the true source DOAs. For SNRs lower than 10 dB, the MAE obtained with the GL-LS approach varies more for different grid resolutions than the MAE obtained with GL-L1. This could indicate that while having a finer grid resolution can be beneficial in the DOA estimation, employing the narrowband system of equations where frequency-dependent information is preserved, as in the GL-L1 approach, can positively affect the robustness towards diffuse noise when trying to localize a point source that does not present a flat spectrum, such as the one simulated in this scenario.

When analyzing the PSD estimation in terms of NMSE, we observe that, for GL-LS, the NMSE seems to only depend on the SNR and not on the grid resolution, indicating that although the PSD re-estimation step via least squares depends on the previously estimated point source DOA, a mismatch between the chosen angular

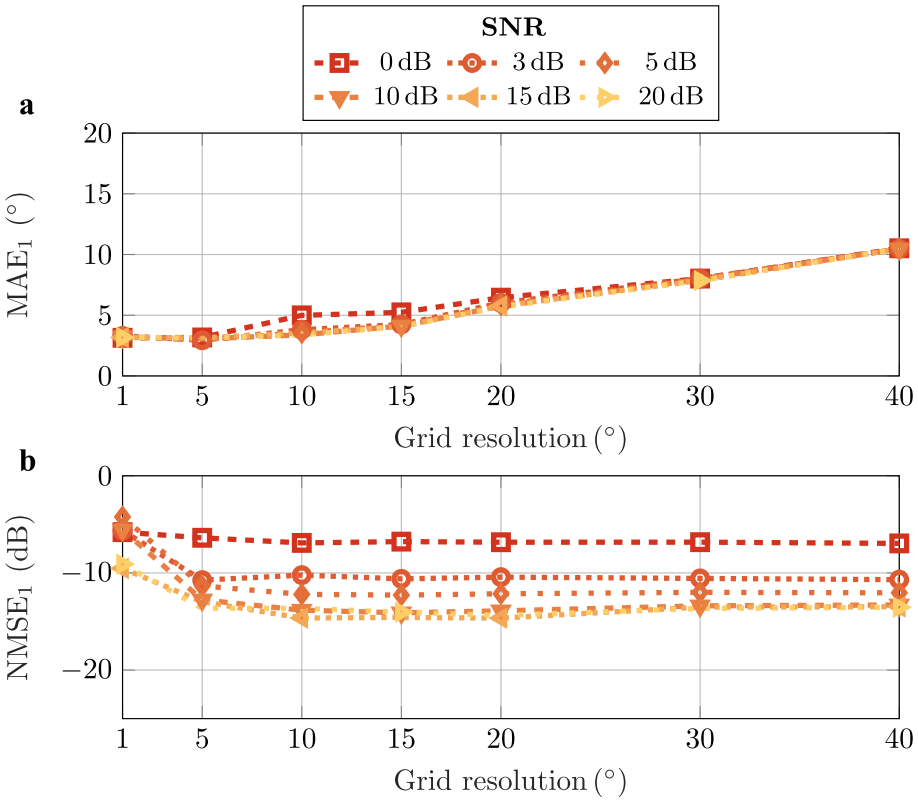

Fig. 2 Estimation errors obtained with GL-L1 for different grid resolutions and SNRs

grid for building the microphone response matrix and the source's true DOA does not impact the PSD estimation accuracy. This is also observed when GL-L1 is used with a grid resolution above 10°. When comparing both methods, we observe that GL-LS overall achieves a lower PSD estimation error than GL-L1 for different SNR values.

5.3 Different levels of microphone response mismatch

In this set of simulations, we aim to investigate the effect of a mismatch between the microphone response used for generating the input vectors $\widetilde{\boldsymbol{\varphi}}_Y$ and $\bar{\boldsymbol{\varphi}}_Y$ used by GL-LS and GL-L1, respectively, and the assumed microphone response when building the linear system of equations to be solved in each proposed approach. In practical scenarios, such a mismatch can occur when it is assumed that the microphone being used presents an ideally flat frequency response, whereas, in reality, it becomes more directional for higher frequencies instead. By fixing the use of an ideal cardioid microphone response when building the linear system of equations, a performance comparison between both approaches presented can be done for different cases of model mismatch caused by the use of the frequency-dependent microphone responses when generating the input vectors $\widetilde{\boldsymbol{\varphi}}_Y$ and $\bar{\boldsymbol{\varphi}}_Y$ used by GL-LS and GL-L1, respectively.

As an additional comparison, we also execute the proposed methods *without* including the additional non-negativity constraint on the error term expressed in (17c) and (24c) for GL-LS and GL-L1, respectively, so that the possible improvement in robustness due to the constraint can be analyzed. In the case of GL-LS, this would correspond to solving the optimization problem presented in [43], and these versions of the proposed approaches are here referred to as GL-LS$_0$ and GL-L1$_0$.

A single point source of speech-shaped noise is again simulated in anechoic conditions, with its DOA being randomly generated for each Monte Carlo realization. The SNR is here fixed at 10 dB and the grid resolution is fixed at 10°. For a normalized frequency value $f \in [0, 1]$, the frequency-dependent directivity patterns, denoted as *Sub-to-cardioid* and *Omni-to-cardioid*, are defined as a linear combination of two directivity functions:

$$a(f, \theta) = (1 - f)a_L(\theta) + fa_H(\theta) \tag{43}$$

where:

$$\begin{aligned} a_H(\theta) &= 0.5 + 0.5\cos(\theta) \\ a_L(\theta) &= 0.75 + 0.25\cos(\theta) \end{aligned} \quad \text{Sub-to-cardioid} \tag{44}$$

or:

$$\begin{aligned} a_H(\theta) &= 0.5 + 0.5\cos(\theta) \\ a_L(\theta) &= 1 \end{aligned} \quad \text{Omni-to-cardioid} \tag{45}$$

The estimation errors obtained when using both versions of each proposed approach with different microphone responses are presented in Fig. 3. We can observe that for the case of generating data with an ideal cardioid, the additional constraint over the error term does not significantly affect the DOA estimation performance in terms of MAE for neither of the methods. However, it does result in slight improvement of the PSD estimation in terms of NMSE for the GL-L1 method. We can also observe that, when an actual mismatch between the microphone response used for building the response matrices and the microphone response used for generating the data is present, the DOA estimation is indeed improved for both approaches when considering the *Sub-to-cardioid* response, but not when considering the use of the *Omni-to-cardioid* response. This is suspected to be due low level of directivity in lower frequencies presented by the microphone response, such that the observed microphone PSD for different orientations does not provide sufficient directional information to appropriately localize the target source.

In terms of PSD estimation, we observe that the NMSE for GL-LS$_0$ and GL-LS do not significantly differ even with a more apparent difference in DOA estimation errors. This may be due to the fact that the GL-LS method is only affected by the additional nonnegativity constraint during the DOA estimation step, and the PSD is then re-estimated via least squares. In the latter step, the microphone response mismatch is still present and therefore not compensated for, possibly yielding similar error levels. A similar effect was observed in Section 5.2 for GL-LS, in which the mean mismatch between the true DOA and the chosen grid candidate, which depends on the grid resolution, did not impact the PSD estimation performance.

When comparing methods GL-L1 and GL-L1$_0$, the NMSE decreases with the inclusion of the additional nonnegativity constraint, even with an increase in MAE for the case of the *Omni-to-cardioid* microphone. This indicates that even if the additional constraint does not improve the DOA estimation, it can still positively affect the PSD estimation. We also observe that the GL-L1 approach seems to be more robust towards model mismatch than GL-LS in terms of PSD estimation, with or without the nonnegativity constraint over the error term, suggesting an advantage of employing the proposed narrowband system of equations instead of its wideband counterpart in this scenario.

5.4 Different levels of reverberation

While the simulations done in anechoic conditions in Sections 5.2 to 5.3 were included to show important performance characteristics of the two proposed approaches,

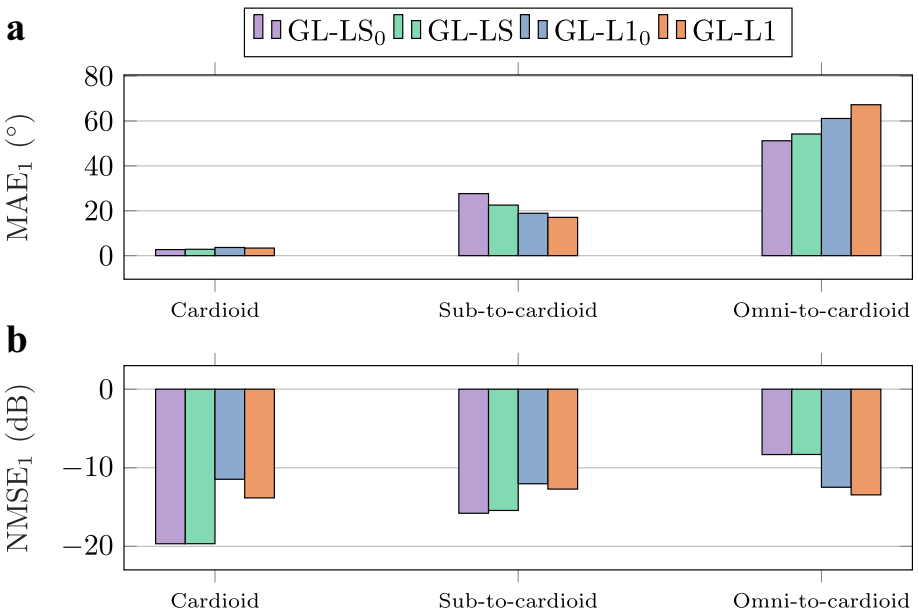

Fig. 3 Estimation errors obtained with GL-LS and GL-L1, as well as with their modified versions which exclude the nonnegativity constraint over the error term (GL-LS$_0$ and GL-L1$_0$, respectively), while using different microphone directivity patterns to generate the observed data

we here consider a more realistic scenario where a sound source is placed in reverberant environments.

A room of dimensions $6.3 \times 5.1 \times 2.5$m is simulated, with a single, ideal cardioid microphone placed at coordinates [3.7, 2.1, 1.5] m. The point source is placed at the same height as the microphone and its DOA is set by randomly generating its coordinates within the room for each Monte Carlo realization, with the constraints of being at least 0.1 m away from the room boundaries and exceeding the setup's critical distance from the microphone, denoted r_c, which varies with the reverberation time T_{60} considered [59]. An illustration of the simulated room with the constraints on the source position is shown in Fig. 4. For each of the $N = 6$ microphone orientations, the room impulse response is generated using the image source method implemented in [60] and convolved with the original point source signal, composed of speech-shaped noise. The grid resolution for building the microphone response matrices is set to 10°, and the SNR is in this case defined as the ratio between the broadband power of the reverberant source signal and the diffuse noise component, with its value fixed at 10 dB. The reverberation time is varied from $T_{60} = 0$ s to $T_{60} = 0.6$ s, where $T_{60} = 0$ s corresponds to the anechoic case, and with reflections being simulated only on the two-dimensional plane in order to concord with the signal model in (1). Finally, the NMSE for evaluating the PSD estimation is computed with respect to a newly defined reference, corresponding to the PSD of the point source signal recorded with the cardioid microphone oriented towards the source's true DOA in the reverberant environment considered. This reference would correspond to the one expressed in (42) and used for evaluating the results obtained when considering anechoic conditions, as the reverberation time would correspond to zero.

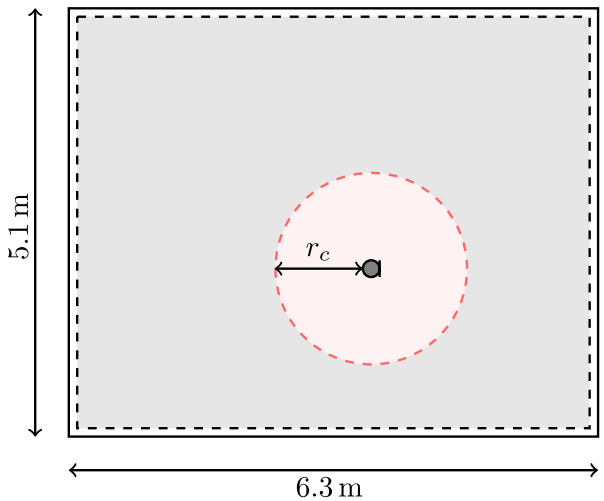

Fig. 4 Illustration of simulated room where the source position is randomly generated within the area represented in gray (at least 10 cm from the wall and farther than the room's critical distance r_c)

The estimation errors obtained when using GL-LS and GL-L1 for all reverberation times considered are presented in Fig. 5. We can observe that the method GL-LS presents overall lower MAE and NMSE than the method GL-L1, indicating the benefit, in this scenario, of summing the signal PSD over frequency. Moreover, both approaches show the tendency of a performance degradation in DOA and PSD estimation with an increase in reverberation time, which is expected as reverberation is not explicitly accounted for in the utilized models (12) and (16).

5.5 Influence of angular separation and power ratio between two sources

So far, we assumed that only a single point source was used when performing the simulations previously presented in Sections 5.2, 5.3, and 5.4. Now, in order to allow a performance comparison between the proposed approaches regarding their capacity in separating different sources, we consider the case where two point sources with different spectral content are recorded by an ideal cardioid microphone. One of the point sources' DOA is randomly selected, whereas the DOA of the other source is then set according to a certain angular separation from the first source, varied from 30° to 180° with a 30°-step. The two sources, indexed as $p = 1$ and $p = 2$, emit colored noise based on a third-octave band filter centered on 1 kHz and 2 kHz, respectively. A power ratio between sources is defined as

$$\mathrm{PR} = \frac{\sigma_{S_2}^2}{\sigma_{S_1}^2}, \tag{46}$$

where $\sigma_{S_1}^2$ and $\sigma_{S_2}^2$ denote the broadband variance of sources $p = 1$ and $p = 2$, respectively. The power ratio is set to 0 dB, 3 dB, and 6 dB according to (46). The grid resolution is fixed at 10° and the SNR, which is still defined with respect to the broadband variance of source $p = 1$, is fixed at 10 dB. The regularization parameter λ is now set to $0.005\|\widetilde{\mathbf{A}}^\top \widetilde{\boldsymbol{\varphi}}_Y\|_\infty$ and $0.005\|\bar{\mathbf{A}}^\top \boldsymbol{\varphi}_Y\|_\infty$ for the GL-LS and GL-L1 methods, respectively, as an attempt to decrease the influence of the group sparsity penalty, since in this scenario more than a single group is expected to be activated. The DOA and PSD estimation errors, computed for each source separately, are presented for the GL-LS and GL-L1 methods in Figs. 6 and 7, respectively. We can observe that, when using GL-LS, the MAE for both sources decreases as the angular separation between them increases, with the error being consistently lower for $p = 2$ when compared to $p = 1$ for PR = 3 dB and

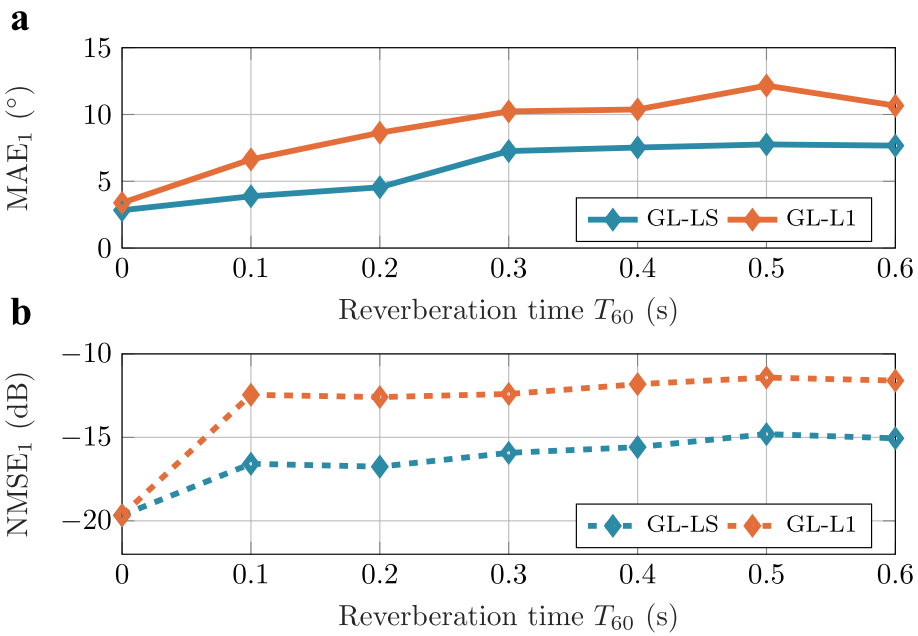

Fig. 5 Estimation errors obtained with GL-LS and GL-L1 for different reverberation times

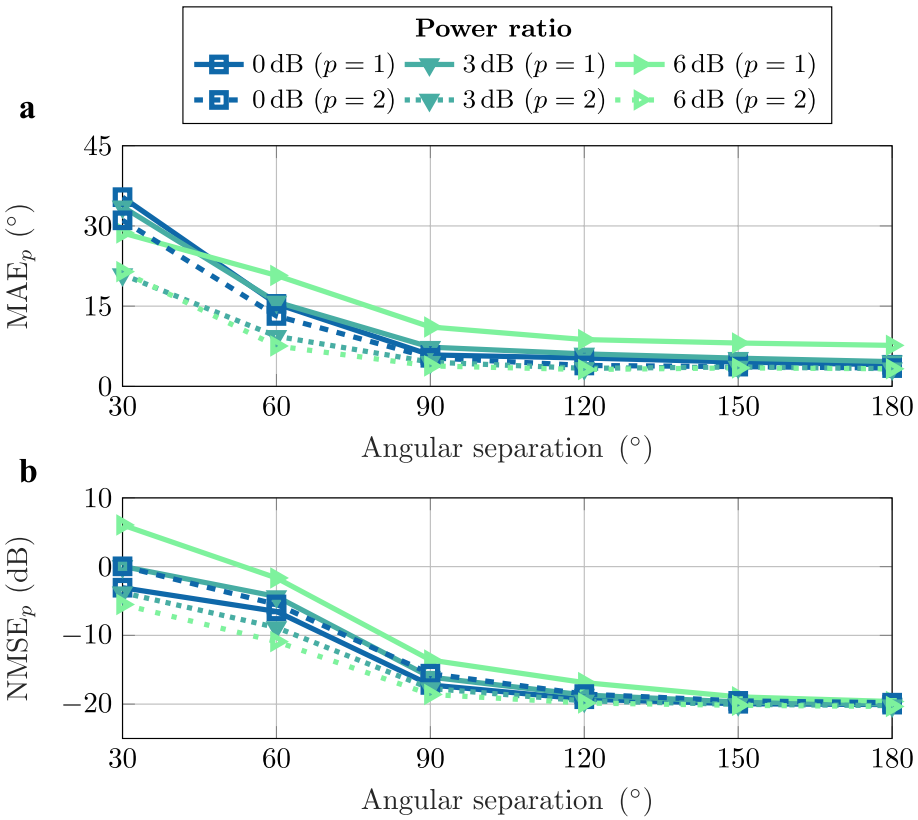

Fig. 6 Estimation errors obtained with GL-LS for different values of angular separation and power ratio between two sources

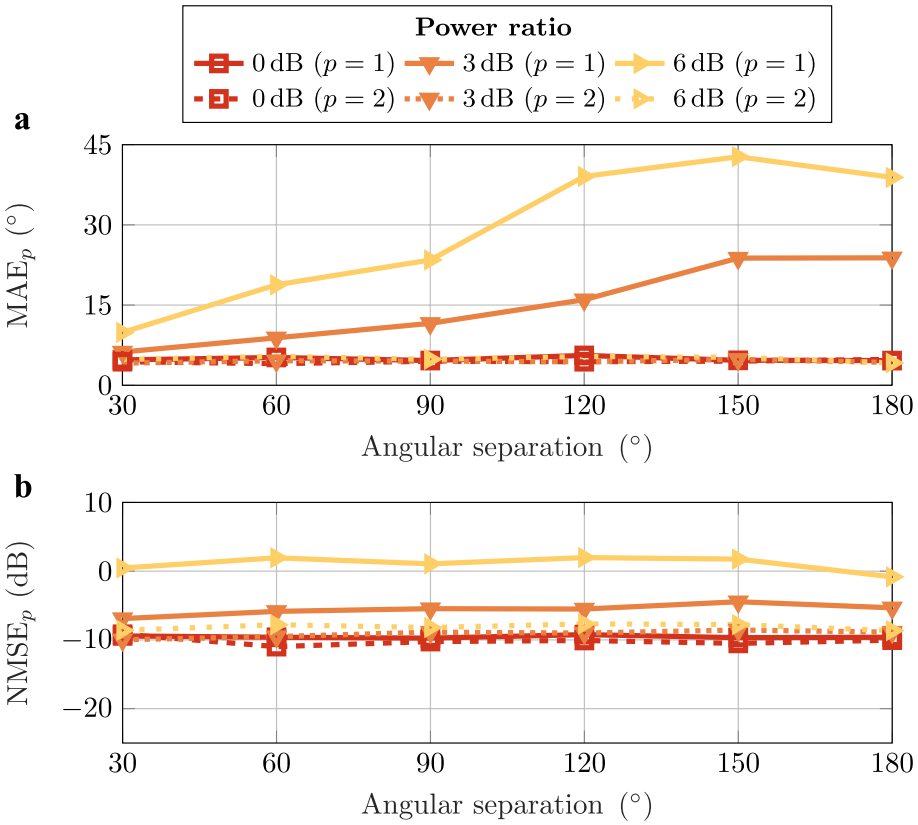

Fig. 7 Estimation errors obtained with GL-L1 for different values of angular separation and power ratio between two sources

PR = 6 dB. We also observe that the difference in terms of MAE between the sources increases as the power ratio increases. These results are due to the second source having higher broadband power than the first, and therefore, being less corrupted by the diffuse noise in comparison. We also observe a similar behavior in the PSD estimation in terms of NMSE, with all errors converging to around −20 dB as the angular separation reaches 180°, and the difference in PSD estimation error between sources increasing with the power ratio.

When considering GL-L1 method, we can observe that, for angular separation values below 60°, its MAE is lower than the one obtained with GL-LS, indicating a benefit of using a frequency-dependent structure when building the proposed linear system of equations to identify closely located sources. However, as opposed to the behavior of GL-LS, the MAE for $p = 1$ increases with angular separation while the MAE for $p = 2$ remains reasonably constant for PR = 3 dB and PR = 6 dB. By further analyzing the multiple realizations of each scenario considered, it was possible to observe that the estimated vector $\hat{\boldsymbol{\varphi}}_S$ often presented, especially for the case when PR > 0 dB, spurious peaks of frequency-averaged PSD neighboring the second source's estimated position ($\hat{\vartheta}_2$), indicating a spreading of the target source's power over multiple, neighboring candidate DOAs. If the frequency-averaged power of a candidate location neighboring source $p = 2$ is greater than the one related to source $p = 1$, then the algorithm will select the incorrect candidate and yield higher DOA estimation errors, which increase with the angular separation between sources and present opposing trends to those of the GL-LS approach. Since the regularization parameter has been heuristically chosen in this work's simulation and may be suboptimal, further tuning could potentially be carried out in order to improve this approach's DOA estimation performance in multi-source scenarios.

Despite an increase in MAE for $p = 1$ as a function of the angular separation, we can observe that, in terms of PSD estimation, the use of the GL-L1 method presents fairly constant NMSE values for both sources and all PR values considered. It is also observed that similarly to the use of GL-LS, the difference in terms of NMSE between two sources increases with the power ratio. Overall, the use of GL-LS showed to achieve lower PSD errors in most cases.

5.6 Influence of the microphone directivity pattern in the case of three sources

As a further investigation on the capacities of the proposed approaches to discriminate between different point sources in space, a new set of simulations is built for a case of three point sources. Upon testing the planned scenario, it was observed that using a cardioid microphone when simulating the recorded signals did not provide enough directional diversity to allow for three distinct peaks to be identified within the estimated vector $\hat{\boldsymbol{\varphi}}_S$. For this reason, we test both methods using microphones with higher-order directivity patterns based on the higher-order differential microphones studied in [61].

A general, second-order microphone directivity pattern, denoted $\Gamma(\theta)$, can be expressed as

$$\Gamma(\theta) = c_0 + c_1 \cos(\theta) + c_2 \cos^2(\theta) \tag{47}$$

where θ denotes the angle and c_0, c_1, and c_2 are real-valued scaling coefficients. By varying the values of c_0, c_1 and c_2, one can obtain different second-order directivity patterns. In this study, we consider simulating three different patterns, here denoted *Cardioid-A*, *Cardioid-B* and *Hypercardioid-2*, based on the different combinations of coefficient values proposed in [61] and presented in Table 2. An illustration of each microphone directivity pattern in absolute values is presented in Fig. 8.

Three point sources of equal power emitting speech-shaped noise are simulated in anechoic conditions, with the DOA of the first source (ϑ_1) being randomly selected between 0° and 360°, and the DOAs of the two remaining sources (ϑ_2 and ϑ_3) being set according to the angular separation considered as $\vartheta_2 = \vartheta_1 + \Delta\theta$ and $\vartheta_3 = \vartheta_1 + 2\Delta\theta$, with $\Delta\theta$ denoting such separation value. As opposed to the previous simulations presented, a total of $N = 9$ microphone orientations uniformly distributed over 360° (i.e., 0°, 40°, 80°, 120°, 160°, 200°, 240°, 280° and 320°) are used when building the linear system of equations used for each method, due to the need for more observation data in order to successfully differentiate the three different point sources. Both the DOA and PSD estimation errors are averaged over all three sources, since they are simulated with equal power, and are denoted as $\overline{\text{MAE}}$ and $\overline{\text{NMSE}}$, respectively.

Table 2 Coefficient values for microphone directivity patterns used

Directivity pattern	$[c_0, c_1, c_2]$
Cardioid	[0.5, 0.5, 0]
Cardioid-A	[0.25, 0.5, 0.4]
Cardioid-B	[0, 0.5, 0.5]
Hypercardioid-2	[-0.2, 0.4, 0.8]

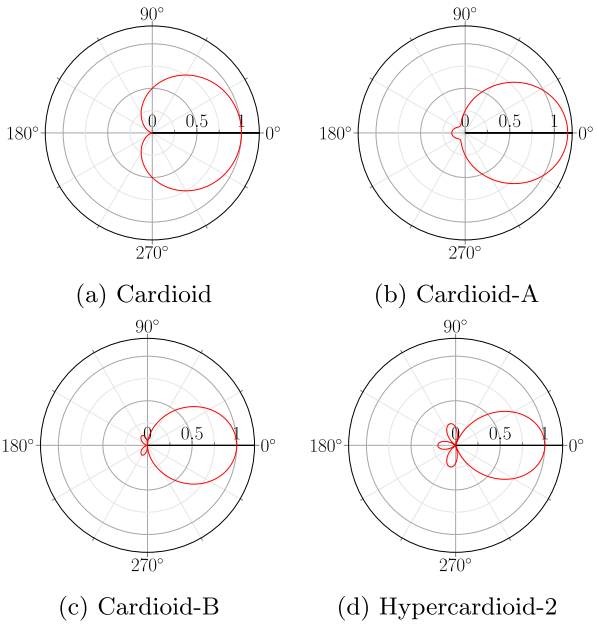

Fig. 8 Microphone directivity patterns considered in this study represented in absolute values

The DOA and PSD estimation errors for the GL-LS and GL-L1 methods while using different second-order directivity patterns and angular separations between sources are presented in Figs. 9 and 10, respectively. When evaluating the DOA estimation, we observe again that the performance of GL-LS strongly depends on the angular separation between sources. However, the choice of microphone directivity pattern has only presented an impact for the case of an angular separation of 60° between sources, in which the *Hypercardioid-2* provided slightly better performance. For the GL-L1 approach, the same pattern yields lower or equal $\overline{\text{MAE}}$ values for all angular separations considered. We again observe that GL-L1 yields lower MAE than GL-LS for the case of sources with angular separation below 60°, indicating the benefit, in this case, of employing the narrowband signal model.

In terms of PSD estimation, it is observed again that, for both proposed approaches, the use of the *Hypercardioid-2* pattern overall yields $\overline{\text{NMSE}}$ values that are lower than or similar to those obtained with the *Cardioid-A* and *Cardioid-B* patterns. The GL-LS approach presents better PSD estimation performance than GL-L1, despite its greater sensitivity to the angular separation between sources when performing the preceding DOA estimation step.

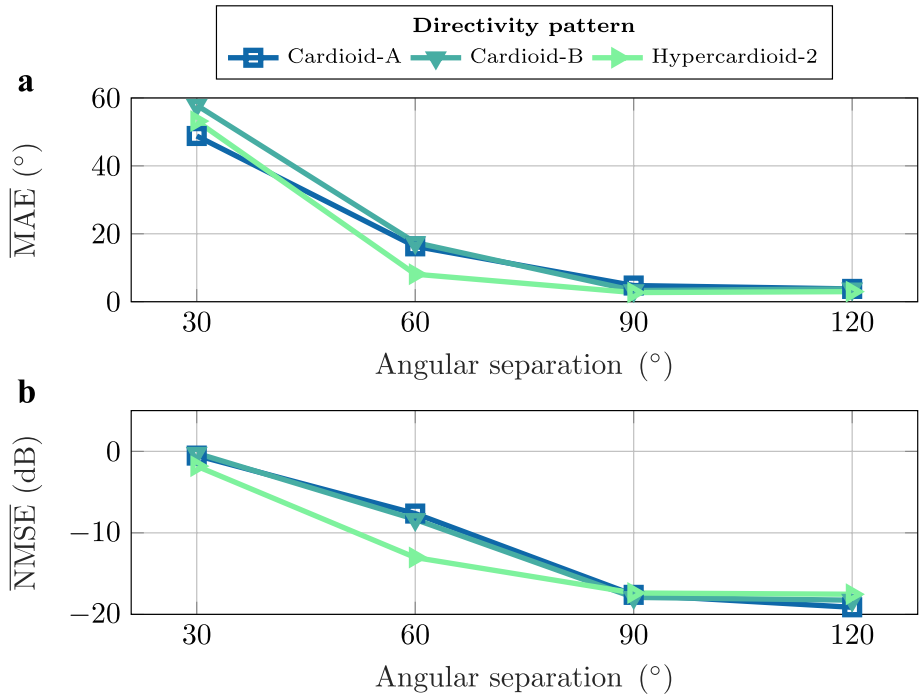

Fig. 9 Estimation errors averaged over three sources and obtained with GL-LS for different values of angular separation, while using different microphone directivity patterns

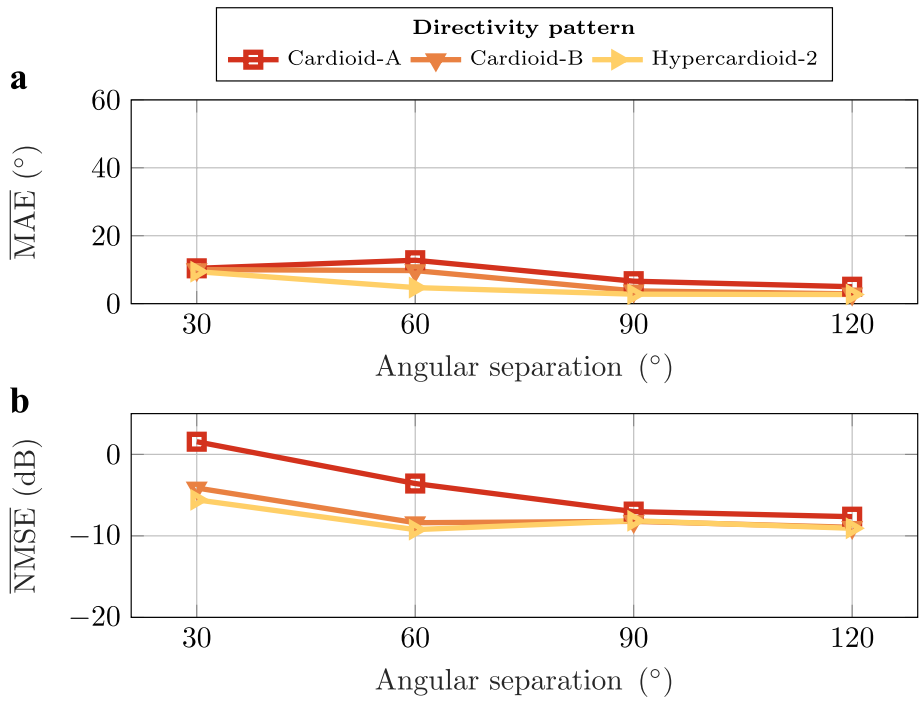

Fig. 10 Estimation errors averaged over three sources and obtained with GL-L1 for different values of angular separation, while using different microphone directivity patterns

5.7 Discussion on the performance limitations of GL-LS and GL-L1

Based on the simulation results presented in this section, it can be observed that for both proposed approaches GL-LS and GL-L1, the DOA and PSD estimation performance can depend on multiple factors, which, when combined, can lead to the necessity of a thorough investigation over the conditions in which their application is to be considered.

In the single-source scenarios simulated in this work, the results indicate that although the intuitive choice of using a finer grid of candidate DOAs can aid at obtaining better DOA estimates, with the MAE being lower bounded by approximately a quarter of the grid resolution, factors such as the noise level and different mismatches between the signal model assumed in the proposed approaches and the practical conditions in which the microphone signals are observed can strongly affect the overall performance of both DOA and PSD estimation. Regarding the model mismatches, it was observed that possible microphone calibration errors in its directivity and room reverberation lead to higher MAE and NMSE values.

In the multi-source scenarios, it was observed that the angular separation and difference in power between sources can significantly affect the DOA estimation performance of both proposed approaches and the PSD estimation performance of GL-LS.

Finally, although the extensive simulations presented in this work can already provide valuable information on the performance trends of the proposed approaches in numerous scenarios, more definitive evaluations can be obtained when considering the case of nonstationary signals and using experimental data. Many of the prevalent applications of DOA and PSD estimation involve speech signals and a combination of model mismatches resulting from multiple factors, which can be simultaneously present in practical setups. Therefore, an investigation of the proposed approaches' behavior under these conditions is required not only for gaining further clarity on the current applicability of the proposed approaches, but also to identify which main aspects should be considered in future work in order to enhance their performance.

6 Conclusion

In this paper, we proposed two approaches for performing DOA and PSD estimation of one or more point sources. The first approach, named GL-LS, is based on a broadband signal model with the PSDs summed over frequency for solving a group-sparse optimization problem with nonnegativity constraints over the desired output vector and the resulting error term, such that the sources' DOAs can be estimated from an overcomplete dictionary of angular candidate positions. Subsequently, a least squares step is performed for re-estimating the point sources' PSDs based on the estimated DOA information. The second approach, named GL-L1, is based on a narrowband signal model structure for solving an analogous optimization problem, which in this case is iteratively, group-wise reweighted for enhancing the solution's sparsity and jointly providing both DOA and PSD estimates.

Both approaches are implemented using ADMM, and simulations were performed for evaluating their performance under different conditions. Compared to the original method which GL-LS and GL-L1 were based on [43], it was observed that, in a scenario involving a microphone response model mismatch, having the additional nonnegativity constraint over the error term can improve the DOA estimation for the case of GL-LS and the PSD estimation for the case of GL-L1. Moreover, in terms of DOA estimation, the GL-L1 approach presented an advantage over GL-LS in scenarios with low SNR or where multiple sources are closely located to each other. Finally, it was shown that having the least squares PSD re-estimation step is beneficial in most scenarios, such that GL-LS outperformed GL-L1 in terms of PSD estimation errors.

Future work includes a further study of the influence of the choice of microphone orientations and directivity pattern when acquiring measurement data over the DOA and PSD estimation performance, expanding the proposed approaches to the multi-channel case and performing experimental tests.

Acknowledgements
Not applicable.

Authors' contributions
All authors jointly developed the methodology and designed the computer simulations presented. ET implemented the algorithms and computer simulations, and all authors jointly interpreted the results obtained. ET drafted the manuscript and all authors read and reviewed the final manuscript.

Authors' information
Not applicable.

Funding
This research work was carried out at the ESAT Laboratory of KU Leuven, in the frame of FWO Mandate SB 1S86520N and FWO Mandate 12ZD622N. The research leading to these results has also received funding from the European Research Council under the European Union's Horizon 2020 research and innovation program/ERC Consolidator Grant: SONORA (no. 773268). This paper reflects only the authors' views and the Union is not liable for any use that may be made of the contained information.

Availability of data and materials
The proposed approaches' implementation is accessible from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Received: 9 June 2023 Accepted: 31 August 2023
Published online: 04 October 2023

References

1. S. Doclo, S. Gannot, M. Moonen, A. Spriet, S. Haykin, K.R. Liu, *in Handbook on array processing and sensor networks, Acoustic beamforming for hearing aid applications*, vol. 9 (Wiley, Hoboken, 2010), pp.269–302
2. P.C. Loizou, *Speech Enhancement: Theory and Practice*, 2nd edn. (CRC Press, Boca Raton, 2013)
3. P.A. Naylor, N.D. Gaubitch, *Speech dereverberation*, vol. 2 (Springer, New York, 2010)
4. M. Brandstein, D. Ward, *Microphone arrays: signal processing techniques and applications* (Springer, New York, 2013)
5. K. Kinoshita, M. Delcroix, S. Gannot, E.A.P. Habets, R. Haeb-Umbach, W. Kellermann, V. Leutnant, R. Maas, T. Nakatani, B. Raj et al., A summary of the reverb challenge: state-of-the-art and remaining challenges in reverberant speech processing research. EURASIP J. Adv. Signal Process. **2016**, 1–19 (2016)
6. S. Gannot, E. Vincent, S. Markovich-Golan, A. Ozerov, A consolidated perspective on multimicrophone speech enhancement and source separation. IEEE/ACM Trans. Audio Speech Lang. Process. **25**(4), 692–730 (2017)
7. E. Vincent, T. Virtanen, S. Gannot, *Audio source separation and speech enhancement* (Wiley, Hoboken, 2018)
8. F. Elvander, R. Ali, A. Jakobsson, T. van Waterschoot, in *Proc. 2019 27th European Signal Process. Conf. (EUSIPCO)*, Offline noise reduction using optimal mass transport induced covariance interpolation (2019), pp. 1–5. https://doi.org/10.23919/EUSIPCO.2019.8903159
9. J. Capon, High-resolution frequency-wavenumber spectrum analysis. Proc. IEEE. **57**(8), 1408–1418 (1969). https://doi.org/10.1109/PROC.1969.7278
10. R. Schmidt, Multiple emitter location and signal parameter estimation. IEEE Trans. Antennas Propagat. **34**(3), 276–280 (1986). https://doi.org/10.1109/TAP.1986.1143830
11. C. Knapp, G. Carter, The generalized correlation method for estimation of time delay. IEEE Trans. Acoust. Speech Signal Process. **24**(4), 320–327 (1976). https://doi.org/10.1109/TASSP.1976.1162830
12. J.H. Dibiase, A high-accuracy, low-latency technique for talker localization in reverberant environments using microphone arrays. Ph.D. thesis (Brown University, Rhode Island, 2000)
13. D. Malioutov, M. Cetin, A. Willsky, A sparse signal reconstruction perspective for source localization with sensor arrays. IEEE Trans. Signal Process. **53**(8), 3010–3022 (2005). https://doi.org/10.1109/TSP.2005.850882
14. S. Fortunati, R. Grasso, F. Gini, M.S. Greco, K. LePage, Single-snapshot DOA estimation by using Compressed Sensing. EURASIP J. Adv. Signal Process. **2014**(1), 120 (2014). https://doi.org/10.1186/1687-6180-2014-120
15. Y. Park, F. Meyer, P. Gerstoft, Sequential sparse Bayesian learning for time-varying direction of arrival. J. Acoust. Soc. Am. **149**(3), 2089–2099 (2021). https://doi.org/10.1121/10.0003802
16. Z. Bai, L. Shi, J.R. Jensen, J. Sun, M.G. Christensen, Acoustic DOA estimation using space alternating sparse Bayesian learning. EURASIP J. Audio, Speech, Music Process. **2021**(1), 14 (2021). https://doi.org/10.1186/s13636-021-00200-z
17. S. Chakrabarty, E.A.P. Habets, in *Proc. 2017 IEEE Workshop Appl. Signal Process. Audio, Acoust. (WASPAA)*, Broadband doa estimation using convolutional neural networks trained with noise signals (IEEE, New Paltz, 2017), pp. 136–140. https://doi.org/10.1109/WASPAA.2017.8170010
18. P.A. Grumiaux, S. Kitić, L. Girin, A. Guérin, A survey of sound source localization with deep learning methods. J. Acoust. Soc. Am. **152**(1), 107–151 (2022). https://doi.org/10.1121/10.0011809
19. R. Martin, Bias compensation methods for minimum statistics noise power spectral density estimation. Signal Process. **86**(6), 1215–1229 (2006). https://doi.org/10.1016/j.sigpro.2005.07.037
20. R. Martin, Noise power spectral density estimation based on optimal smoothing and minimum statistics. IEEE Trans. Speech Audio Process. **9**(5), 504–512 (2001). https://doi.org/10.1109/89.928915
21. Y. Ephraim, D. Malah, Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator. IEEE Trans. Acoust. Speech Signal Process. **32**(6), 1109–1121 (1984). https://doi.org/10.1109/TASSP.1984.1164453
22. T. Gerkmann, R.C. Hendriks, Unbiased MMSE-Based Noise Power Estimation With Low Complexity and Low Tracking Delay. IEEE Trans. Audio Speech Lang. Process. **20**(4), 1383–1393 (2012). https://doi.org/10.1109/TASL.2011.2180896
23. G. Enzner, P. Thune, in *Proc. 2017 IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Robust MMSE filtering for single-microphone speech enhancement (IEEE, New Orleans, 2017), pp. 4009–4013. https://doi.org/10.1109/ICASSP.2017.7952909
24. I. Cohen, B. Berdugo, Noise estimation by minima controlled recursive averaging for robust speech enhancement. IEEE Signal Process. Lett. **9**(1), 12–15 (2002). https://doi.org/10.1109/97.988717
25. I. Cohen, Noise spectrum estimation in adverse environments: Improved minima controlled recursive averaging. IEEE Trans. Speech Audio Process. **11**(5), 466–475 (2003). https://doi.org/10.1109/TSA.2003.811544
26. N. Fan, J. Rosca, R. Balan, in *Proc. 2007 IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Speech Noise Estimation using Enhanced Minima Controlled Recursive Averaging (IEEE, Honolulu, 2007), pp. IV–581–IV–584. https://doi.org/10.1109/ICASSP.2007.366979
27. J.-M. Kum, J.-H. Chang, Speech Enhancement Based on Minima Controlled Recursive Averaging Incorporating Second-Order Conditional MAP Criterion. IEEE Signal Process. Lett. **16**(7), 624–627 (2009). https://doi.org/10.1109/LSP.2009.2019351
28. S. Rangachari, P.C. Loizou, A noise-estimation algorithm for highly non-stationary environments. Speech Commun. **48**(2), 220–231 (2006). https://doi.org/10.1016/j.specom.2005.08.005
29. R. Hendriks, J. Jensen, R. Heusdens, Noise Tracking Using DFT Domain Subspace Decompositions. IEEE Trans. Audio Speech Lang. Process. **16**(3), 541–553 (2008). https://doi.org/10.1109/TASL.2007.914977
30. T. Dietzen, S. Doclo, M. Moonen, T. van Waterschoot, Square root-based multi-source early PSD estimation and recursive RETF update in reverberant environments by means of the orthogonal Procrustes problem. IEEE/ACM Trans. Audio Speech Lang. Process. **28**, 755–769 (2020)
31. N. Pan, J. Benesty, J. Chen, On single-channel noise reduction with rank-deficient noise correlation matrix. Appl. Acoust. **126**, 26–35 (2017). https://doi.org/10.1016/j.apacoust.2017.05.010
32. Q. Zhang, A. Nicolson, M. Wang, K.K. Paliwal, C. Wang, DeepMMSE: A Deep Learning Approach to MMSE-Based Noise Power Spectral Density Estimation. IEEE/ACM Trans. Audio Speech Lang. Process. **28**, 1404–1415 (2020). https://doi.org/10.1109/TASLP.2020.2987441
33. H. Zhao, S. Zarar, I. Tashev, C.H. Lee, in *Proc. 2018 IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Convolutional-Recurrent Neural Networks for Speech Enhancement (IEEE, Calgary, 2018), pp. 2401–2405. https://doi.org/10.1109/ICASSP.2018.8462155
34. X. Li, S. Leglaive, L. Girin, R. Horaud, Audio-Noise Power Spectral Density Estimation Using Long Short-Term Memory. IEEE Signal Process. Lett. **26**(6), 918–922 (2019). https://doi.org/10.1109/LSP.2019.2911879
35. M. Kolbk, Z.H. Tan, J. Jensen, Speech Intelligibility Potential of General and Specialized Deep Neural Network Based Speech Enhancement Systems. IEEE/ACM Trans. Audio Speech Lang. Process. **25**(1), 153–167 (2017). https://doi.org/10.1109/TASLP.2016.2628641
36. J. Benesty, J. Chen, Y. Huang, *Microphone Array Signal Processing , Springer Topics in Signal Processing*, vol. 1 (Springer, Berlin, Heidelberg, 2008). https://doi.org/10.1007/978-3-540-78612-2

37. R.C. Hendriks, T. Gerkmann, J. Jensen, *DFT-Domain Based Single-Microphone Noise Reduction for Speech Enhancement: A Survey of the State of the Art*, vol. 9 (Morgan & Claypool, San Rafael, 2013)
38. A. Plinge, F. Jacob, R. Haeb-Umbach, G.A. Fink, Acoustic Microphone Geometry Calibration: An overview and experimental evaluation of state-of-the-art algorithms. IEEE Signal Process. Mag. **33**(4), 14–29 (2016). https://doi.org/10.1109/MSP.2016.2555198
39. K. Kokkinakis, B. Azimi, Y. Hu, D.R. Friedland, Single and multiple microphone noise reduction strategies in cochlear implants. Trends Amplif. **16**(2), 102–116 (2012). https://doi.org/10.1177/1084713812456906
40. A. Saxena, A.Y. Ng, in, IEEE Int. Conf. Robot. Autom. (ICRA) Proceedings. Learning sound location from a single microphone **2009**, 1737–1742 (2009). https://doi.org/10.1109/ROBOT.2009.5152861
41. Y. Hioka, R. Drage, T. Boag, E. Everall, in *Proc. 2018 Int. Workshop Acoustic Signal Enhancement (IWAENC)*, Direction of arrival estimation using a circularly moving microphone (IEEE, Tokyo, 2018), pp. 91–95. https://doi.org/10.1109/IWAENC.2018.8521297
42. S. Jesus, M. Porter, Y. Stephan, X. Demoulin, O. Rodriguez, E. Coelho, Single hydrophone source localization. IEEE J. Oceanic Eng. **25**(3), 337–346 (2000). https://doi.org/10.1109/48.855379
43. E. Tengan, M. Taseska, T. Dietzen, T. van Waterschoot, in *Proc. 2021 29th European Signal Process. Conf. (EUSIPCO)*, Direction-of-arrival and power spectral density estimation using a single directional microphone (2021), pp. 221–225. https://doi.org/10.23919/EUSIPCO54536.2021.9616239
44. Q. Shen, W. Liu, W. Cui, S. Wu, Y.D. Zhang, M.G. Amin, Low-complexity direction-of-arrival estimation based on wideband co-prime arrays. IEEE/ACM Trans. Audio Speech Lang. Process. **23**(9), 1445–1456 (2015). https://doi.org/10.1109/TASLP.2015.2436214
45. Y. Hioka, K. Niwa, in *Proc. 2014 Int. Workshop Acoustic Signal Enhancement (IWAENC)*, PSD estimation in beamspace for source separation in a diffuse noise field (Juan-les-Pins, France, 2014), pp. 85–88. https://doi.org/10.1109/IWAENC.2014.6953343
46. K. Niwa, T. Kawase, K. Kobayashi, Y. Hioka, in *Proc. 2016 Int. Workshop Acoustic Signal Enhancement (IWAENC)*, PSD estimation in beamspace using property of M-matrix (Xi'an, China, 2016), pp. 1–5. https://doi.org/10.1109/IWAENC.2016.7602965
47. M. Grant, S. Boyd. CVX: Matlab software for disciplined convex programming, version 2.1. https://cvxr.com/cvx (2014). Accessed 12 Sep 2023
48. M. Yuan, Y. Lin, Model selection and estimation in regression with grouped variables. J Royal Statistical Soc B. **68**(1), 49–67 (2006). https://doi.org/10.1111/j.1467-9868.2005.00532.x
49. B. Porat, *A Course in Digital Signal Processing* (John Wiley, New York, 1997)
50. T. Hastie, R. Tibshirani, M. Wainwright, *Statistical Learning with Sparsity: The Lasso and Generalizations* (CRC Press LLC, New York, 2015)
51. E.J. Candès, M.B. Wakin, S.P. Boyd, Enhancing sparsity by reweighted ℓ_1 minimization. J. Fourier Anal. Appl. **14**(5–6), 877–905 (2008). https://doi.org/10.1007/s00041-008-9045-x
52. F. Elvander, T. Kronvall, S. Adalbjörnsson, A. Jakobsson, An adaptive penalty multi-pitch estimator with self-regularization. Signal Process. **127**, 56–70 (2016). https://doi.org/10.1016/j.sigpro.2016.02.015
53. R. Chartrand, Wotao Yin, in *Proc. 2008 IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Iteratively reweighted algorithms for compressive sensing (IEEE, Las Vegas, 2008), pp. 3869–3872. https://doi.org/10.1109/ICASSP.2008.4518498
54. D. Wipf, S. Nagarajan, Iterative Reweighted l_1 and l_2 Methods for Finding Sparse Solutions. IEEE J. Sel. Top. Signal Process. **4**(2), 317–329 (2010). https://doi.org/10.1109/JSTSP.2010.2042413
55. S. Boyd, Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers. Found. Trends Mach. Learn. **3**(1), 1–122 (2010). https://doi.org/10.1561/2200000016
56. J.R. Gilbert, C. Moler, R. Schreiber, Sparse matrices in MATLAB: Design and implementation. SIAM J. Matrix Anal. Appl. **13**(1), 333–356 (1992). https://doi.org/10.1137/0613024
57. I.S. Duff, A.M. Erisman, J.K. Reid, *Direct methods for sparse matrices* (Oxford University Press, Oxford, 2017). https://doi.org/10.1093/acprof:oso/9780198508380.001.0001
58. Bang and Olufsen, Music for Archimedes. CD B&O 101 (1992)
59. H. Kuttruff, *Room acoustics* (CRC Press, Boca Raton, 2016)
60. E. Habets, Room impulse response generator. Tech. rep. (2006)
61. E. De Sena, H. Hacihabiboglu, Z. Cvetkovic, On the Design and Implementation of Higher Order Differential Microphones. IEEE Trans. Audio Speech Lang. Process. **20**(1), 162–174 (2012). https://doi.org/10.1109/TASL.2011.2159204

Publisher's Note