IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received August 18, 2018, accepted September 13, 2018, date of publication September 18, 2018, date of current version October 19, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2870887

Sound Field Reproduction via the Alternating Direction Method of Multipliers Based Lasso Plus Regularized Least-Square

MAOSHEN JIA[1], (Senior Member, IEEE), JIAMING ZHANG[1], (Student Member, IEEE), YUXUAN WU[1], (Student Member, IEEE), AND JING WANG[2], (Member, IEEE)

[1]Beijing Key Laboratory of Computational Intelligence and Intelligent System, Faculty of Information Technology, Beijing University of Technology, Beijing 100124, China

[2]School of Information and Electronic, Beijing Institute of Technology, Beijing 100081, China

Corresponding authors: Maoshen Jia (jiamaoshen@bjut.edu.cn) and Jing Wang (wangjing@bit.edu.cn)

This work was supported by the China Postdoctoral Science Foundation under Project 2017M610731.

ABSTRACT This paper proposes a 3-D sound field reproduction (SFR) approach through the combination of alternating direction method of multipliers (ADMM)-based least-absolute shrinkage and selection operator (Lasso) and regularized least square (LS). The proposed SFR method is split into two parts through the pressure matching optimization of loudspeaker positions and the computation of driving signals. At the first part, a plurality of candidate positions of loudspeakers in planar array is given and, then, the active speaker selection method is proposed based on ADMM complex Lasso algorithm for selecting the optimal loudspeaker positions. Afterwards, regularized LS is adopted to calculate the selected loudspeaker weights and control the total power. The numerical simulation experiments demonstrate that the proposed SFR scheme outperforms the existing sparse loudspeakers' placement and weight optimization algorithms especially in under-sampled sound fields. Meanwhile, the evaluations also confirmed that the proposed method could significantly reduce the computational complexity of the active loudspeaker selection compared to the state-of-the-art Lasso-based SFR. Effectively, the proposed method uses a relatively small number of loudspeakers for a satisfying reproduction quality.

INDEX TERMS Sound field reproduction, alternating direction method of multipliers, Lasso, least squares.

I. INTRODUCTION

Sound Field Reproduction (SFR) aims to generate a desired or target sound field using an array of loudspeakers within a given region or matching points of space. It has been frequently used in virtual auditory environments, such as auditory display, immersive gaming and communication systems, etc. The ultimate goal of SFR is to seek the driving signals (complex amplitudes) of the loudspeakers in order to minimize the reproduction error in the listening zone. For this purpose, SFR has become an ongoing research topic in recent decades.

There are three main approaches for SFR: Wave Field Synthesis (WFS) [1], [2], Higher Order Ambisonics (HOA) [3], [4] and Pressure Matching (PM) [5]–[7]. The first approach, the WFS, is based on the Huygens–Fresnel principle, which means that any wavefront can be synthesized from a series of elementary spherical waves. By using the Kirchhoff-Helmholtz integral, the WFS system is designed under the condition that the driving signals of loudspeakers need to be equivalent to the distribution of the sound pressure gradient of the desired sound field [8]. Afterwards, the desired sound field can be accurately reproduced in a large listening area. However, the aperture between loudspeakers is limited by the spatial sampling theorem. The second approach, the HOA, is based on the Cylindrical Harmonic (CH) or Spherical Harmonic (SH) representations of a sound field [9]. The CH or SH expression of the desired field facilitate finding the magnitude and phase for the loudspeakers. Although HOA can be applied in many sound playback configurations, the number of required loudspeakers increases quadratically as the order increases. In addition, the reproduction setup is relatively complicated [10].

The third approach that will be used in this paper is the PM-based SFR. For arbitrarily loudspeaker array geometries, PM approach is designed to derive the loudspeaker weights based on the Least-Square (LS) criterion and the goal of PM approach is to minimize the reproduction error at a set of spatially distributed matching points in a predefined sound control zone. In practice, the matrix inversion problems in the loudspeaker weight calculation is often ill-posed and sensitive to measurement noise, so that the regularization is necessary in PM-based SFR. LS constraint regularized with the l_2-norm of the loudspeaker weights (a.k.a, Tikhonov regularization or ridge regression) is usually employed in SFR systems [5], [11], [12], which is appropriate for power constraint usage scenarios and improves the system robustness. However, the selection of the optimum loudspeaker positions is not considered in such l_2-penalization method thus all the loudspeakers in the array will be activated, this phenomenon may lead to a blurry spatial sound image [13]. Besides, SFR error function is a non-convex problem in terms of loudspeaker locations, which is NP-hard in general [14]. As an alternative, Lilis *et al.* [15] make the problem convex by exploiting l_1-penalization instead of l_2-penalization. It means that the globally optimum speaker locations and the corresponding weight signals can be solved using a compressive sensing idea where the loudspeaker weights are penalized with the l_1-norm, namely, the so-called Least-absolute shrinkage and selection operator (Lasso). Due to the selectivity of Lasso, it is clear that the solution is sparse, i.e., only a few loudspeakers in the candidate positions remain simultaneously active. Typical Lasso-based SFR system include [13], [15], and [16] that exhibit a substantial SFR error reduction compared with l_2-regularized LS, but all these systems perform SFR in the absence of power limitations.

Recent works have been focusing on the combination of the Tikhonov regularization and Lasso to make a trade-off between the number of active loudspeakers and the reproduction error such that the SFR can be implemented with power constraint. For instance, Radmanesh and Burnett [17] proposed a two-stage Lasso-LS optimization framework for 2-dimensional (2D) multi-zone SFR using circular speaker array. Both the loudspeaker locations and weights are optimized to achieve the minimum reproduction error at all matching points. Nevertheless, the major difficulty of this SFR scheme is the computational complexity of Lasso when the number of candidate loudspeaker increases, which is also a common trouble for Lasso-based SFR. To tackle this issue, an Efficient Harmonic Nested (EHN) dictionary algorithm is presented in [18] where the Harmonic Nested Arrays (HNA) is designed for each frequency called candidate subarray and then an EHN dictionary is formed by sequentially removing the previously selected speaker locations of HNA. In other words, EHN dictionary is a subset of the original dictionary (i.e., all candidate loudspeakers) that cannot incorporate every situation. Moreover, this algorithm works well only for linear loudspeaker array. As far as the 3D SFR by using the planar array, Khalilian *et al.* [19], [20] proposed a SFR method based on Singular Value Decomposition (SVD). First, in this method, an "ideal" Acoustic Transform Function (ATF) matrix is created by using SVD matrices of the uniform distribution of omni-directional loudspeaker array. Afterwards, the appropriate locations over all candidate loudspeakers are selected to match the modified ATF matrix and the driven signal is obtained by LS constraint. In contrast to the SVD-based SFR, a method for deriving the positions and the driven signal of loudspeakers was proposed by utilizing Constrained Matching Pursuit (CMP) [21]. This algorithm uses the iterative approach to select the location of each loudspeaker whose ATF matrix is the most correlated with the reproduction error. Both of them maintain an ideal performance and acceptable SFR error.

In this paper, we propose a 3-dimensional (3D) SFR method for the application scenario of immersive spatial sound by using a planar loudspeaker array. In this SFR system, several candidate loudspeaker positions are given for selection and the final loudspeaker array is obtained by selecting the loudspeakers that have a significant contribution to SFR. For achieving an effective loudspeaker placement and calculating the corresponding driven signals of selected loudspeakers under power limitation, the proposed SFR system is split into two parts by exploiting the structure of two stage Lasso-LS. During the first step, an active speaker selection method is proposed based on the framework of Alternating Direction Method of Multipliers (ADMM). ADMM is particularly useful for solving optimization problems that are too large to be handled by generic optimization solvers. As far as for solving l_1-norm regularization problems, ADMM technique keeps smoothing term in algorithmic design but takes advantage of the structure of the l_1 norm instead of using it to approximate the non-smooth l_1 terms [22]. Once the active loudspeaker is determined, ridge regression method is utilized to find the driven signal of all active loudspeakers in the second step. In existing Lasso-based SFR, the generic solver is the least-angle regression (LARS) algorithm, which yields the full piecewise linear solution path of the regression coefficients, or with a low-complexity procedure such as the Coordinate Descent (CD) method. They are the most popular algorithms but may not be the most efficient or up-to-date implementation [13]. Motivated by the abovementioned issues, this work proposes an efficient and fast method to accelerate the procedure of optimal loudspeaker selection in the context of holding the integrity of the original dictionary. Moreover, we extend ADMM algorithm to solve the Lasso problem in the complex case for selecting the optimal loudspeaker positions. Simulation results demonstrate that the proposed ADMM-based Lasso plus LS (AL-LS) model achieves a better SFR performance and the time delay of SFR can be reduced to the great extent.

Summing up, the main contribution of this paper is twofold: 1) Utilizing ADMM to design an active loudspeaker selection strategy and 2) Joint optimization of l_1 and l_2 regularization to obtain a loudspeaker weight estimator in 3D environments via a planar speaker array.

The remainder of this paper is organized as follows: Section II explains the SFR based on pressure matching

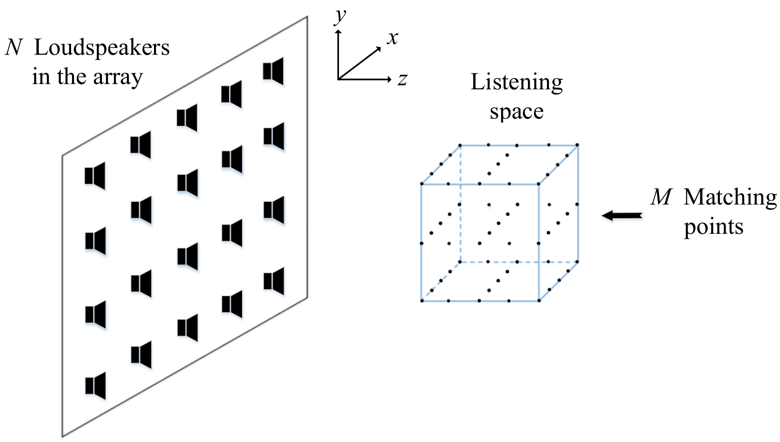

FIGURE 1. Illustration of 3D SFR structure using a planar array of loudspeaker.

and LS approach. Section III introduces the structure of the proposed SFR method in detail. Experimental results are presented and discussed in Section IV, while the conclusion is given in Section V.

II. PRELIMINARIES

A. SOUND FIELD GENERATION

The SFR structure using planar array is shown in Figure 1. Assume that the number of matching points in the listening area and the number of loudspeakers in the array are M and N, respectively.

The sound pressure at the m-th matching point generated by the n-th speaker is defined as:

$$p^{\mathrm{t}}_{m,n}(t) = w^{\mathrm{t}}_n(t) * g_{m,n}(t) \tag{1}$$

where $m(1 \leq m \leq M)$ and $n(1 \leq n \leq N)$ are the indexes of the matching point and the candidate loudspeaker, respectively. $w^{\mathrm{t}}_n(t)$ is the n-th loudspeaker excitation in time domain, $g_{m,n}(t)$ denotes the impulse response between the n-th loudspeaker and the m-th matching point, $*$ is convolution operation, and t is the time index. After converting the time domain equation to the frequency domain, the sound pressure can be rewritten in the following form:

$$p_{m,n}(f) = w_n(f) \cdot G_{m,n}(f) \tag{2}$$

where $w_n(f)$ and $G_{m,n}(f)$ are the representation of $w^{\mathrm{t}}_n(t)$ and $g_{m,n}(t)$ in the frequency domain, respectively. In this work, both the loudspeakers and sources are modeled to be monopoles radiating in free-field. Therefore, $G_{m,n}(f)$, also called Acoustic Transform Function (ATF), can be expressed as free space Green's function:

$$G_{m,n}(f) = \frac{1}{4\pi} \cdot \frac{e^{-jk\|\mathbf{x}_n - \mathbf{y}_m\|_2}}{\|\mathbf{x}_n - \mathbf{y}_m\|_2} \tag{3}$$

where $\mathbf{x}_n$ and $\mathbf{y}_m$ are the positions of the n-th loudspeaker and the m-th matching point, $k = 2\pi f/c$ is the wave number, and $c = 343$m/s is the sound propagation speed.

Let $\mathbf{w}(f) \in \mathbb{C}^N$ denote the vector formed by aggregating each loudspeaker weight at frequency f, that is, $\mathbf{w}(f) = [w_1(f), w_2(f), \cdots, w_N(f)]^{\mathrm{T}}$, where T stands for transpose operation. Meanwhile, let $\mathbf{p}^{\mathrm{d}}(f) \in \mathbb{C}^M$ and $\mathbf{p}^{\mathrm{r}}(f) \in \mathbb{C}^M$ be the sound pressure of the desired and the reproduced sound field sampled at matching points $\{\mathbf{y}_m\}_{m=1}^{M}$. For simplicity in exposition, f will be omitted hereafter, the preceding defined vectors are written as $\mathbf{w}$, $\mathbf{p}^{\mathrm{d}}$ and $\mathbf{p}^{\mathrm{r}}$. Suppose there exists an omnidirectional point source located at $\mathbf{s} = (s_x, s_y, s_z)$ with complex amplitude of A, the desired sound field is computed as:

$$\mathbf{p}^{\mathrm{d}} = A \cdot \frac{e^{-jk\|\mathbf{s} - \mathbf{y}_m\|_2}}{4\pi\ \|\mathbf{s} - \mathbf{y}_m\|_2} \tag{4}$$

For each frequency f, the reproduced sound field $\mathbf{p}^{\mathrm{r}}$ can be obtained by:

$$\mathbf{p}^{\mathrm{r}} = \mathbf{G}\mathbf{w} \tag{5}$$

where $\mathbf{G} \in \mathbb{C}^{M \times N}$ whose (m, n)-th entry is equal to $G_{m,n}(f)$.

B. SOUND FIELD REPRODUCTION USING LS CRITERION

PM is achieved by minimizing the approximation error between the synthesized field and the desired field. The proximity of the reproduced field with the desired one is quantified using LS criterion thus the loudspeaker weight can be estimated by:

$$\begin{aligned}\tilde{\mathbf{w}} &= \arg\min_{\mathrm{w}} \left\|\mathbf{p}^{r} - \mathbf{p}^{\mathrm{d}}\right\|_2^2 \\ &= \arg\min_{\mathrm{w}} \left\|\mathbf{G}\mathbf{w} - \mathbf{p}^{\mathrm{d}}\right\|_2^2\end{aligned} \tag{6}$$

The unique closed-form solution of (6) is given by (7) when the Green's function matrix $\mathbf{G}$ is tall, i.e., for $M > N$:

$$\tilde{\mathbf{w}} = \mathbf{G}^{+}\mathbf{p}^{\mathrm{d}} = \left(\mathbf{G}^{\mathrm{H}}\mathbf{G}\right)^{1} \mathbf{G}^{\mathrm{H}}\mathbf{p}^{\mathrm{d}} \tag{7}$$

where $\tilde{\mathbf{w}}$ is the estimated loudspeaker weight vector, the superscript $^{+}$ and $^{\mathrm{H}}$ represent Moore–Penrose pseudo-inverse and conjugate transpose. However, inverse problems

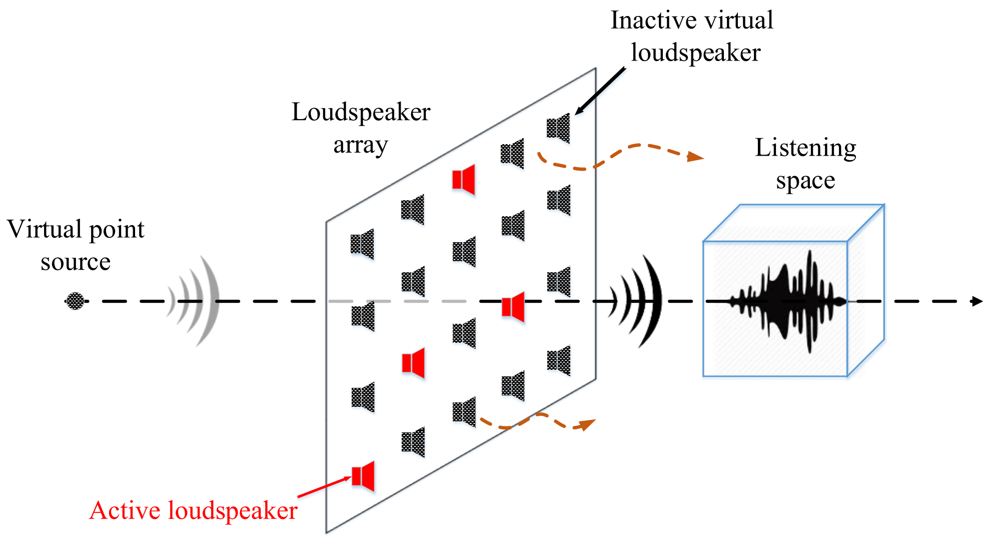

FIGURE 2. **The proposed 3D SFR system configuration.**

are not always well-posed which may lead to inaccuracy of the reproduced sound field [13]. Therefore, regularization is necessary in SFR design. To solve (6), the ensuing section elaborates the proposed SFR system by combining Lasso and regularized LS.

III. PROPOSED SOUND FIELD REPRODUCTION SYSTEM

A. SFR SYSTEM MODEL

As illustrated in Figure 2, our proposed 3D SFR model uses a planar loudspeaker array to reproduce the desired sound field generated by a virtual point source. The main goal of this work is to design a loudspeaker placement strategy and find the corresponding driving signals, in order to achieve a satisfying reproduction quality with a relatively small number of loudspeakers. Suppose there are N_a loudspeakers are available, intuitively speaking, the simplest way for loudspeaker placement is the benchmark configuration where N_a loudspeakers distributed uniformly. However, it is not an effective method, because the loudspeaker which are located at boundary or corner contribute less than those located near the center of the plane of the array. On the contrary, if these N_a loudspeakers are placed densely near the center, LS-based SFR may allocate excessive power to them, which produces many artifacts and degrades the overall SFR quality [15]. Therefore, we design a self-adaptive loudspeaker selection strategy, where the N candidate virtual loudspeaker positions are provided and then the most efficient N_a loudspeakers out of N loudspeakers ($N \gg N_a$) are selected for SFR. In this case, the procedure of loudspeaker selection is to find a sparse solution of loudspeaker weight vector $\mathbf{w}$, which can be converted into a convex optimization problem. Afterwards, LS is adopted to optimize the previously selected loudspeaker weight. While this approach has been applied in 2D multi-zone SFR [17], [18], [22], it remains challenge in 3D environment. The conversion from 2D SFR to 3D SFR means a dramatic increase in data volume, which results in the expansion of Green's function matrix $\mathbf{G}$. In addition, the conventional algorithm for loudspeaker selection and weight estimation is accompanied by the high computational complexity, accordingly, it cannot guarantee the low latency of the SFR system. In this section, we design a 3D SFR method using the two stage Lasso-LS for immersive applications, an ADMM-based loudspeaker selection strategy is proposed in the first stage to select N_a loudspeakers out of N loudspeakers. Then, the complex excitations of N_a loudspeakers will be calculated by utilizing LS.

B. LASSO MODEL STATEMENT FOR SFR

LS-based SFR aims to search a loudspeaker weight vector $\mathbf{w}$ to minimize the reproduction error $\left\|\mathbf{Gw} - \mathbf{p}^{\mathrm{d}}\right\|_2^2$, which falls short in limiting the number of simultaneously active loudspeakers and allocates power to all loudspeakers of uniformly spaced array. In practical scenarios, we hope to use as few loudspeakers as possible. Hence, it is necessary to select the most efficient loudspeaker locations out of N candidate positions for reducing the number of active loudspeaker. Inspired by variable selection and compressive sampling ideas, the sparse solution of (6) can be derived by solving the following constrained minimization problem:

$$\hat{\mathbf{w}} = \arg\min_{\mathbf{w}} \frac{1}{2}\left\|\mathbf{Gw} - \mathbf{p}^{\mathrm{d}}\right\|_2^2 + \mu \left\|\mathbf{w}\right\|_0 \tag{8}$$

where the l_0-norm $||\mathbf{w}||_0$ indicates the number of nonzero entries in $\mathbf{w}$ (i.e. the number of active loudspeakers), and μ is the regularization parameter. Solving (8) is in general NP-hard which requires exhaustive search over all subsets of columns of the Green's function matrix $\mathbf{G}$ [24]. Alternatively, l_1-norm is usually chosen as a proxy for the l_0 counting norm, because it is closest convex approximation to l_0 quasi-norm [25]. By replacing the l_0-norm of the regularization term $\mu||\mathbf{w}||_0$ with the l_1-norm, the procedure of obtaining the loudspeaker weight can be seen as a Lasso problem:

$$\hat{\mathbf{w}}_{\text{Lasso}} = \arg\min_{\mathbf{w}} \frac{1}{2}\left\|\mathbf{G}\mathbf{w} - \mathbf{p}^{\text{d}}\right\|_2^2 + \lambda \|\mathbf{w}\|_1 \tag{9}$$

where the vector one-norm is given by $\|\mathbf{w}\|_1 = \sum_{n=1}^{N} |w_n|$. The positive Lasso penalty parameter λ regulates the sparsity level of $\mathbf{w}$ (i.e., controls the number of active speaker). The formula (9) can be solved by generic methods, such as LARS, sub-gradient or CD algorithm. From numerical point of view, Lasso model has non-smooth terms in objective functions (e.g., In Equation (9), the right-side term is non-smooth)—this non-smoothness excludes prohibitively straightforward applications of some mature algorithms in optimization that rely on the acquisition of gradient information [22]. Hence, in many Lasso-based SFR systems, CD is employed as Lasso solver that is a Non-gradient optimization algorithm. Although the computational complexity of CD is lower than LARS and sub-gradient [26], it is also a time-consuming task and cannot be perform in real-time. These drawbacks limit the direct application of generic optimization techniques and raise the necessity of particular consideration in algorithmic design of low-latency SFR. To overcome this restriction, the present paper draws from convex optimization advances to introduce a novel approach for selecting the optimal loudspeaker positions, namely, ADMM technique.

C. ADMM COMPLEX LASSO ALGORITHM FOR ACTIVE LOUDSPEAKER SELECTING

Recently, ADMM has been applied to solve structured or large scale convex optimization problems arising from different applications in diverse areas, such as compressed sensing [27] and machine learning [28]. It enjoys the superior convergence properties of the method of multipliers and the decomposability property of dual ascent. Specifically, it uses variable splitting to decompose the problem into easily solvable sub-problems [29]. With respect to l_1-norm regularization problems for selecting the active loudspeakers, the formula (9) can be seen as the sum of two separable convex functions, which meets the form of ADMM. Therefore, we propose the ADMM-based algorithm to derive the solution of the loudspeaker weight vector.

In this proposed SFR scheme, there are N candidate loudspeaker positions for selection, the driving signals of all candidate loudspeakers are denoted as $\mathbf{w} \in \mathbb{C}^N$. The n-th column of the Green's function matrix $\mathbf{G} \in \mathbb{C}^{M\times N}$ is the ATF of the n-th candidate loudspeaker at all matching points. The sound pressure set of the desired sound field is $\mathbf{p}^{\text{d}} \in \mathbb{C}^M$ that is calculated at M matching points in listening space. Each element of $\mathbf{w}$, $\mathbf{p}^{\text{d}}$ and $\mathbf{G}$ is complex value. For dealing with problems involving complex variables, we map $\mathbf{w}$, $\mathbf{p}^{\text{d}}$ and $\mathbf{G}$ into real-valued matrices:

$$\overline{\mathbf{w}} \triangleq \begin{bmatrix} \Re(\mathbf{w}) \\ \Im(\mathbf{w}) \end{bmatrix} \tag{10}$$

$$\overline{\mathbf{p}}^{\text{d}} \triangleq \begin{bmatrix} \Re(\mathbf{p}^{\text{d}}) \\ \Im(\mathbf{p}^{\text{d}}) \end{bmatrix} \tag{11}$$

$$\overline{\mathbf{G}} \triangleq \begin{bmatrix} \Re(\mathbf{G}) & -\Im(\mathbf{G}) \\ \Im(\mathbf{G}) & \Re(\mathbf{G}) \end{bmatrix} \tag{12}$$

where $\Re(\cdot)$ and $\Im(\cdot)$ denote the real and imaginary parts of a complex number, respectively. This operation doubles the original dimension, that is, $\overline{\mathbf{w}} \in \mathbb{R}^{2N}$, $\overline{\mathbf{p}}^{\text{d}} \in \mathbb{R}^{2M}$ and $\overline{\mathbf{G}} \in \mathbb{R}^{2M\times 2N}$. Now, the equivalent real model of Lasso problem is:

$$\hat{\overline{\mathbf{w}}}_{\text{Lasso}} = \arg\min_{\overline{\mathbf{w}}} \frac{1}{2}\left\|\overline{\mathbf{G}}\overline{\mathbf{w}} - \overline{\mathbf{p}}^{\text{d}}\right\|_2^2 + \lambda \|\overline{\mathbf{w}}\|_1 \tag{13}$$

where $\hat{\overline{\mathbf{w}}}_{\text{Lasso}}$ is the sparse solution, which means that loudspeakers in many candidate locations are not be active. In this subsection, we want to select the proper loudspeaker locations on a planar grid automatically and control the number of active loudspeaker equal to N_a, this procedure is introduced in detail as follow.

The convergence of ADMM is proven for the problem with an objective function that is sum of two separable convex functions with linear constraints. Here, we re-formulate (13) as the following optimization problem:

$$\begin{aligned} &\text{minimize}\, f(\overline{\mathbf{w}}) + g(\overline{\mathbf{z}}) \\ &\text{subject to}\ \overline{\mathbf{w}} - \overline{\mathbf{z}} = 0 \end{aligned} \tag{14}$$

where $f(\overline{\mathbf{w}}) = (1/2)\left\|\overline{\mathbf{G}}\overline{\mathbf{w}} - \overline{\mathbf{p}}^{\text{d}}\right\|_2^2$, $\overline{\mathbf{z}}$ is the intermediate variable when splitting the original problem (13), and $g(\overline{\mathbf{z}}) = \lambda\|\overline{\mathbf{z}}\|_1$. Note that both $f(\cdot)$ and $g(\cdot)$ are convex functions. The augmented Lagrangian function of this problem is given by:

$$L_\rho(\overline{\mathbf{w}}, \overline{\mathbf{z}}, \overline{\boldsymbol{\eta}}) = f(\overline{\mathbf{w}}) + g(\overline{\mathbf{z}}) + \overline{\boldsymbol{\eta}}^{\text{T}}(\overline{\mathbf{w}} - \overline{\mathbf{z}}) + \frac{\rho}{2}\|\overline{\mathbf{w}} - \overline{\mathbf{z}}\|_2^2 \tag{15}$$

where $\overline{\boldsymbol{\eta}}$ is dual variable or Lagrange multiplier and $\rho > 0$ is the augmented Lagrangian parameter.

The iteration operation is needed for solving (14), which consists of the three steps. At k-th iteration, we first fix $\overline{\mathbf{z}}$ and $\overline{\boldsymbol{\eta}}$ to minimize the augmented Lagrangian over $\overline{\mathbf{w}}$ (i.e., $\overline{\mathbf{w}}$-minimization):

$$\overline{\mathbf{w}}^{k+1} = \arg\min_{\overline{\mathbf{w}}} L_\rho(\overline{\mathbf{w}}, \overline{\mathbf{z}}^k, \overline{\boldsymbol{\eta}}^k) \tag{16}$$

Afterwards, we fix $\overline{\mathbf{w}}$ and $\overline{\boldsymbol{\eta}}$ to minimize the augmented Lagrangian over $\overline{\mathbf{z}}$ (i.e., $\overline{\mathbf{z}}$-minimization):

$$\overline{\mathbf{z}}^{k+1} = \arg\min_{\overline{\mathbf{z}}} L_\rho(\overline{\mathbf{w}}^{k+1}, \overline{\mathbf{z}}, \overline{\boldsymbol{\eta}}^k) \tag{17}$$

Finally, the dual variable $\overline{\boldsymbol{\eta}}$ is updated (i.e., $\overline{\boldsymbol{\eta}}$-minimization):

$$\overline{\boldsymbol{\eta}}^{k+1} = \overline{\boldsymbol{\eta}}^k + \overline{\mathbf{w}}^{k+1} - \overline{\mathbf{z}}^{k+1} \tag{18}$$

For convenience, let $\overline{\mathbf{u}} = (1/\rho)\overline{\boldsymbol{\eta}}$ be the scaled version of the dual variable, the scaled form of (16)-(18) can be expressed as:

$$\overline{\mathbf{w}}^{k+1} = \arg\min_{\overline{\mathbf{w}}} \left\{ f(\overline{\mathbf{w}}) + \rho/2\left\|\overline{\mathbf{w}} - \overline{\mathbf{z}}^k + \overline{\mathbf{u}}^k\right\|_2^2 \right\} \tag{19}$$

Algorithm 1 ADMM Complex Lasso Algorithm for Obtaining the Positions and Weights of the Active Loudspeakers

Input: $\overline{\mathbf{G}}$ ▶Green's function matrix mapped into real-valued model
Input: $\overline{\mathbf{p}}^{\text{d}}$ ▶desired sound field mapped into real-valued model
Input: λ ▶Lasso regularization parameter
Input: ρ ▶augmented Lagrangian parameter
Input: ε^{abs} ▶upper limit of absolute tolerances
Input: ε^{rel} ▶upper limit of relative tolerance
Output: $\hat{\mathbf{w}}_{\text{Lasso}}$ ▶sparse complex loudspeaker weight
1. Initialize $\overline{\mathbf{w}}^0 = \overline{\mathbf{z}}^0 = \overline{\boldsymbol{\eta}}^0 = \mathbf{0}$, $k = 0$ and $i = \sqrt{-1}$.
2. **while**(1) **do**
3. $\quad \overline{\mathbf{w}}^{k+1} \leftarrow \left(\overline{\mathbf{G}}^{\text{T}}\overline{\mathbf{G}} + \rho\mathbf{I}\right)^{-1}\left[\overline{\mathbf{G}}^{\text{T}}\overline{\mathbf{p}}^{\text{d}} + \rho\left(\overline{\mathbf{z}}^k - \overline{\mathbf{u}}^k\right)\right]$ ▶ where $\mathbf{I}$ is the identity matrix
4. $\quad \overline{\mathbf{z}}^{k+1} \leftarrow S_{\lambda/\rho}\left(\overline{\mathbf{w}}^{k+1} + \overline{\mathbf{u}}^k\right)$
5. $\quad \overline{\mathbf{u}}^{k+1} \leftarrow \overline{\mathbf{u}}^k + \overline{\mathbf{w}}^{k+1} - \overline{\mathbf{z}}^{k+1}$
6. **if** stopping criterion is satisfied **then** ▶stopping criterion is defined in (23)-(26)
7. $\quad$ break
8. **end if**
9. $k = k + 1$
10. **end while**
11. **for** $j = 1 : N$ **do** ▶convert real value to complex value to obtain the estimated loudspeaker weight
12. $\quad \hat{\mathbf{w}}_{\text{Lasso}}(j) = \overline{\mathbf{w}}(j) + i \cdot \overline{\mathbf{w}}(j + N)$
13. **end for**
14. **return** $\hat{\mathbf{w}}_{\text{Lasso}}$

$$\overline{\mathbf{z}}^{k+1} = \arg\min_{\overline{\mathbf{z}}}\left\{g(\overline{\mathbf{z}}) + \rho/2\left\|\overline{\mathbf{w}}^{k+1} - \overline{\mathbf{z}} + \overline{\mathbf{u}}^k\right\|_2^2\right\} \tag{20}$$

$$\overline{\mathbf{u}}^{k+1} = \overline{\mathbf{u}}^k + \overline{\mathbf{w}}^{k+1} - \overline{\mathbf{z}}^{k+1} \tag{21}$$

More specifically, by substituting $f(\overline{\mathbf{w}})$ and $g(\overline{\mathbf{z}})$ to (19)-(21), the required steps of the update rules are presented in **Algorithm 1**.

A couple of remarks in **Algorithm 1** are now emphasized.

Remark 1 (Use Algorithm 1 to Choose N_a Active Loudspeakers in the Planar Array): In the above ADMM-based algorithm, the Lasso regularization parameter λ tunes the number of nonzero elements of $\hat{\mathbf{w}}_{\text{Lasso}}$. As λ decreases, the candidate loudspeakers of array will be successively activated, but not vice-versa. This work aims to select N_a active loudspeakers, i.e., $\left\|\hat{\mathbf{w}}_{\text{Lasso}}\right\|_0 = N_a$. The penalty parameter λ can be searched over the set $[0, \left\|\mathbf{G}^{\text{H}}\mathbf{p}^{\text{d}}\right\|_\infty)$ with a step size of $10^{-4} \cdot \left\|\mathbf{G}^{\text{H}}\mathbf{p}^{\text{d}}\right\|_\infty$ until $\hat{\mathbf{w}}_{\text{Lasso}}$ satisfy the condition $\left\|\hat{\mathbf{w}}_{\text{Lasso}}\right\|_0 = N_a$, then the current value of λ is chosen as the Lasso penalty factor. On the contrary, if $\left\|\hat{\mathbf{w}}_{\text{Lasso}}\right\|_0 \neq N_a$ for any optional λ, we select the λ which results in $\left\|\hat{\mathbf{w}}_{\text{Lasso}}\right\|_0$ greater than N_a slightly. Then the first N_a largest elements in $\hat{\mathbf{w}}_{\text{Lasso}}$ are preserved and other elements are set to be zero. It should be note that a unique value of λ corresponds to a certain N_a. For different SFR applications, the number of active loudspeaker may be different, the corresponding values of λ can be pre-calculated and stored in a lookup table.

Remark 2 (Soft Thresholding Function in Step 4 of the Algorithm 1): The soft thresholding function $S_{\lambda/\rho}(\cdot)$ during $\overline{\mathbf{z}}$-updating is obtained as follow [30]:

$$S_\kappa(\theta) = (1 - \frac{\kappa}{|\theta|})_+\theta = \begin{cases} \theta - \kappa & \text{if } \theta > \kappa \\ 0 & \text{if } |\theta| \leq \kappa \\ \theta + \kappa & \text{if } \theta < -\kappa \end{cases} \tag{22}$$

Remark 3 (Matrix Factorization in Step 3 of the Algorithm 1): It is worth to notice that since $\rho > 0$, $\overline{\mathbf{G}}^{\text{T}}\overline{\mathbf{G}}$ is positive semi-definite matrix which is always invertible over $\overline{\mathbf{w}}$-minimization. To further alleviate the cost of calculations, the Cholesky decomposition of $(\overline{\mathbf{G}}^{\text{T}}\overline{\mathbf{G}} + \rho\mathbf{I}) = \mathbf{H}\mathbf{H}^{\text{T}}$ is calculated once at the outset in $O(N^3)$ flops, and then re-use these cached computations across all the solves in $O(N^2)$ flops. Moreover, if $\overline{\mathbf{G}}$ is fat (i.e., $M < N$), the Sherman-Morrison-Woodbury inversion formula [31] is adopted to substitute a factorization of the much smaller matrix $\left(\mathbf{I} + \overline{\mathbf{G}}\overline{\mathbf{G}}^{\text{T}}/\rho\right)$ for the factorization of $\left(\overline{\mathbf{G}}^{\text{T}}\overline{\mathbf{G}} + \rho\mathbf{I}\right)$. Then, subsequent iterations can be carried out with a faster and lower-dimensional operation at cost of $O(MN)$.

Remark 4 (Definition of Stopping Criterion in Step 6-8 of the Algorithm 1): For declaring termination, the setup of stopping criterion is depended on the primal residual $\mathbf{e}_{\text{pri}}^{k1}$ and dual residual $\mathbf{e}_{\text{dual}}^{k1}$. The iteration is terminated when the primal and dual residuals satisfy a stopping criterion (which can vary depending on the requirements of the application). A typical criterion is to stop when:

$$\left\|\mathbf{e}_{\text{pri}}^{k1}\right\|_2 = \left\|\overline{\mathbf{w}}^{k1} - \overline{\mathbf{z}}^{k+1}\right\|_2 \leq \varepsilon^{\text{pri}} \tag{23}$$

$$\left\|\mathbf{e}_{\text{dual}}^{k1}\right\|_2 = \left\|\rho\left(\overline{\mathbf{u}}^{k1} - \overline{\mathbf{u}}^k\right)\right\|_2 \leq \varepsilon^{\text{dual}} \tag{24}$$

where the tolerances ε^{pri} and $\varepsilon^{\text{dual}}$ can be chosen using an absolute plus relative criterion:

$$\varepsilon^{\text{pri}} = \sqrt{N}\varepsilon^{\text{abs}} + \varepsilon^{\text{rel}} \max\left\{\left\|\overline{\mathbf{w}}^{k1}\right\|_2^2, \left\|\overline{\mathbf{z}}^{k+1}\right\|_2^2\right\} \quad (25)$$

$$\varepsilon^{\text{dual}} = \sqrt{N}\varepsilon^{\text{abs}} + \varepsilon^{\text{rel}} \left\|\rho\overline{\mathbf{u}}^{k1}\right\|_2^2 \quad (26)$$

A reasonable value for the absolute and relative tolerances in this paper are $\varepsilon^{\text{abs}} = 10^{-4}$ and $\varepsilon^{\text{rel}} = 10^{-2}$ chosen based on a series of informal experimental results. It should be noted that those two tolerances may not be unique, which need to be set according to actual situation (see [32] for more details).

Remark 5 (Update of the Augmented Lagrangian Parameter of the Algorithm 1): For improving the convergence and making performance less dependent on the initial value of the parameter ρ, we update this parameter in each iteration as follow:

$$\rho^{k+1} = \begin{cases} \tau^{\text{incr}}\rho^k & \text{if } \left\|\mathbf{e}_{\text{pri}}^{k1}\right\|_2 > \mu \left\|\mathbf{e}_{\text{dual}}^{k1}\right\|_2 \\ \rho^k/\tau^{\text{decr}} & \text{if } \left\|\mathbf{e}_{\text{dual}}^{k1}\right\|_2 > \mu \left\|\mathbf{e}_{\text{pri}}^{k1}\right\|_2 \\ \rho^k & \text{otherwise} \end{cases} \quad (27)$$

where typical choices of $\mu = 10$ and $\tau^{\text{incr}} = \tau^{\text{decr}} = 2$ [33], the initial value $\rho^0 = 0.1$ is set to be slightly low to start with. The idea behind this parameter iteration is to try to keep the primal and dual residual norms within a factor of μ of one another as they both converge to zero.

After we find the $\hat{\mathbf{w}}_{\text{Lasso}}$ through Algorithm 1, the number of nonzero entries in $\hat{\mathbf{w}}_{\text{Lasso}}$ is counted as N_a which represents the number of active loudspeakers that are employed in the second stage.

D. LS-BASED WEIGHT ESTIMATION FOR ACTIVE LOUDSPEAKER

The previously attained sparse vector $\hat{\mathbf{w}}_{\text{Lasso}} = \left\{\hat{w}_{\text{Lasso}}^i\right\}_{i=1}^N$ contains N_a nonzero components. The index set representing those nonzero elements (the indexes of the active loudspeakers selected in the preceding stage) can be extracted as $\mathbf{I}^{\text{nz}} = \left\{i_1^{\text{nz}}, i_2^{\text{nz}}, \cdots, i_{N_a}^{\text{nz}}\right\}$, $\mathbf{I}^{\text{nz}} \subset \{1, 2, \cdots, N\}$, i.e.,

$$\hat{w}_{\text{Lasso}}^i \begin{cases} = 0 & \text{if } i \notin \mathbf{I}^{\text{nz}} \\ \neq 0 & \text{if } i \in \mathbf{I}^{\text{nz}} \end{cases} \quad (28)$$

Meanwhile, the new Green's function matrix $\mathbf{G}_{\text{a}} = \left\{\mathbf{g}_1^{\text{a}}, \mathbf{g}_2^{\text{a}}, \cdots, \mathbf{g}_{N_a}^{\text{a}}\right\}$ whose column vector $\mathbf{g}_b^{\text{a}}$ ($b = 1, 2, \ldots, N_a$) is formed as follow:

$$\mathbf{g}_b^{\text{a}} = \mathbf{g}_{i_b^{\text{nz}}} \quad (29)$$

where $\mathbf{g}_{i_b^{\text{nz}}}$ represents the i_b^{nz}-th column vector of the original Green's function matrix $\mathbf{G}$. After this operation, $\mathbf{G}_{\text{a}}$ with size of $M \times N_a$ is the ATF of all active loudspeakers at all matching points. The second stage aims to minimize the squared error between the desired and reproduced sound field with power constraint:

$$\hat{\mathbf{w}}_{\text{LS}} = \arg\min_{\text{w}} \left\|\mathbf{G}_{\text{a}}\mathbf{w} - \mathbf{p}^{\text{d}}\right\|_2^2 \quad \text{s.t. } \|\mathbf{w}\|_2^2 < p_{\max} \quad (30)$$

where $p_{\max}$ denotes the maximum allowable power. The minimization in (29) can be expressed as the form of the l_2-constransint on the loudspeaker weight:

$$\hat{\mathbf{w}}_{\text{LS}} = \arg\min_{\text{w}} \left\|\mathbf{G}_{\text{a}}\mathbf{w} - \mathbf{p}^{\text{d}}\right\|_2^2 + \gamma \|\mathbf{w}\|_2^2 \quad (31)$$

where $\gamma > 0$ is the l_2 regularization parameter which is used for controling the total loudspeaker power. The closed-form solution of (30) is given by:

$$\hat{\mathbf{w}}_{\text{LS}} = \left(\mathbf{G}_{\text{a}}^{\text{H}}\mathbf{G}_{\text{a}} + \gamma\mathbf{I}\right)^1 \mathbf{G}_{\text{a}}^{\text{H}}\mathbf{p}^{\text{d}} \quad (32)$$

where $\hat{\mathbf{w}}_{\text{LS}}$ is the final active loudspeaker weight for SFR. In the meantime, the reproduced sound field can be obtained by:

$$\mathbf{p}^{\text{r}} = \mathbf{G}_{\text{a}}\hat{\mathbf{w}}_{\text{LS}} \quad (33)$$

Using the above steps, the optimal active loudspeaker location and their corresponding complex amplitudes are found in the frequency domain.

IV. NUMERICAL EXPERIMENTS AND SIMULATION RESULTS

In this section, a series of simulations under various conditions are presented to evaluate the performance of the proposed SFR method.

A. TEST CONDITIONS

The tests consider the immersive audio applications, which is based on the 3D SFR structure of Figure 2. For the sake of facilitative observing, we first set up spatial Cartesian coordinates. Suppose the loudspeaker array occupies 3m×3m square centered at the origin in the xy plane of $z = 0$, while the desired field is generated by a virtual monochromatic source with complex amplitude equal to $8e^{i\theta}$ ($\theta = 0°$ in this paper) located 8m away from the array, whose coordinate is $(0, 0, -8)$. Besides, the listening space is a cube with side lengths of 1m and located 1m away from the speaker array in the direction of positive z-axis, i.e., the central coordinate is $(0, 0, 1.5)$. All matching points are uniformly distributed over the listening area, the distance between the adjacent points is 0.2m, thus the total number of matching points is $M = 125$. The number of candidate loudspeakers and the active loudspeakers in the array are set to be $N = 625$ and $N_a = 25$, respectively. The LS regularization parameter γ is calculated by Newton method from [11], which limits the maximum available loudspeaker power $p_{\max}$. The SFR technique based on Singular Value Decomposition (SVD) [19], [20] and Constrained Matching Pursuit (CMP) [20], [21] are served as reference approach.

B. SOUND FIELD REPRODUCTION ERROR ANALYSIS FOR MONOCHROMATIC PRIMARY SOURCE

The first experiment is to give the intuitionistic SFR results within and without the reproduction zone. Figure 3 shows a snapshot of real parts of the desired field radiated by the point

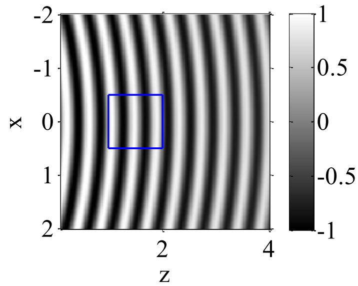

FIGURE 3. Cross-section of the desired sound field radiated by a point source located at (0, 0, −8) with complex amplitude $8e^{i\theta}$ ($\theta = 0°$), frequency $f = 800$Hz.

source at $f = 800$Hz in the $y = 0$ plane. Notice that the reproduction zone (listening region) is marked by the blue square.

Herein, we perform the test under the following two conditions: (a) unconstrained loudspeaker power and (b) limiting the total loudspeaker power with $p_{\max} = 2$.

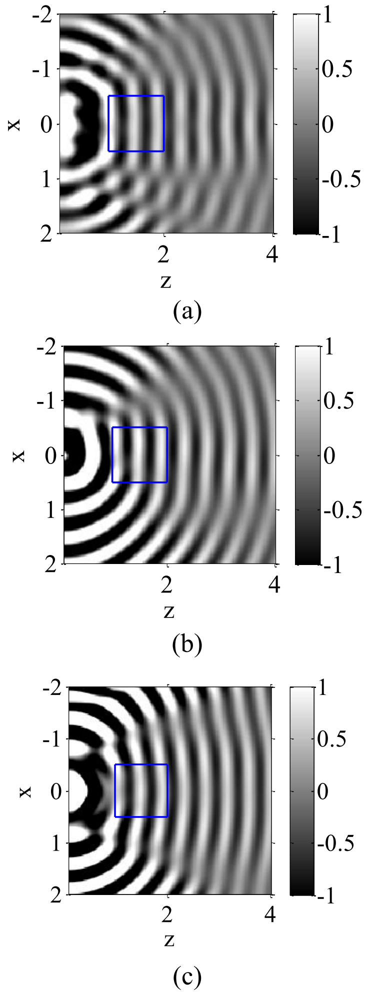

FIGURE 4. Reproduced sound field visualization without power limitation using (a) CMP-based algorithm (b) SVD-based algorithm (c) AL-LS based algorithm.

The Figure 4 shows the cross-sections of the reproduced sound field for the three methods with unlimited power, while Figure 5 shows the results with $p_{\max} = 2$. Figure 4(a) and Figure 5(a) represent the sound fields reproduced by CMP-based algorithm. Figure 4(b) and Figure 5(b) represent the sound fields reproduced by SVD-based algorithm. Moreover, the sound fields reproduced by the proposed ADMM-based Lasso plus LS (named 'AL-LS' in the following experiments) SFR approach are depicted in Figure 4(c) and Figure 5(c).

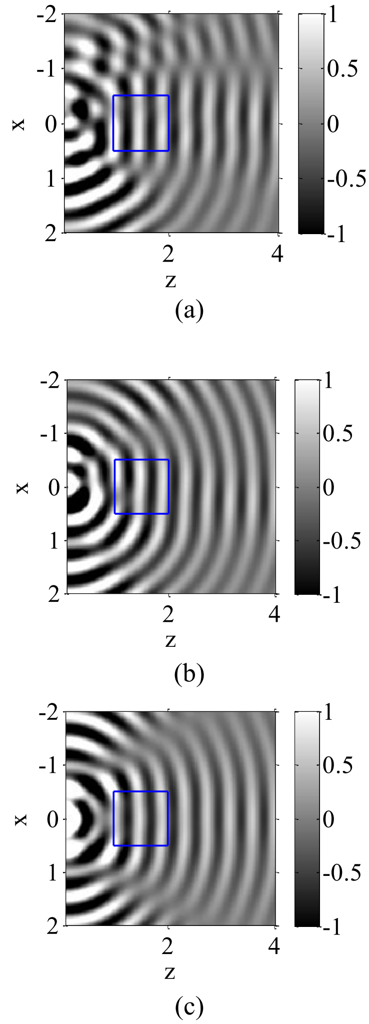

FIGURE 5. Reproduced sound field visualization under power constraint $p_{\max} = 2$ using (a) CMP-based algorithm (b) SVD-based algorithm (c) AL-LS based algorithm.

As shown in Figure 4 and Figure 5, all the three methods restore the desired field inside the reproduction zone reasonably well under the two test conditions. However, the reproduced sound field is highly intensified in the region $\boldsymbol{D} = \{(x, z)| -2 \leq x \leq 2, 0 \leq z \leq 2\}$ (outside the listening region) for unconstrained power scenario (see Figure 4), which result in the enlargement of average reproduction error. In the power-limited scenario (see Figure 5), the overall energy distribution is more evenly than that in the unconstrained power scenario for the three methods. This phenomenon reveal that the power limitation is indispensable in practical applications and the total loudspeaker power should be kept as small as possible whilst ensuring the reproduction quality inside the listening zone.

Additionally, as depicted in Figure 5, the proposed SFR method has the minimum reproduction distortion compared to the reference approaches, which confirms that a more balanced quality of the reproduced sound field can be maintained both inside and outside the reproduction zone.

To examine the SFR performance inside the listening region quantitatively, the following experiments are performed under various power limitations and multiple frequencies. The reproduction error is calculated at densely distributed matching points $M_1 = 125000$ (i.e., the distance of the adjacent points is 2cm) to ensure that the reproduced sound field is over-sampled. Because 2cm is smaller than half of the wavelength of the highest frequency (2kHz) in this simulation ($\lambda/2 = c/2f \approx 8.58$cm). The Normalized Mean Square Error (NMSE) of the reproduced sound field relative to the desired sound field is adopted for performance evaluation, which is defined as:

$$\mathrm{NMSE(dB)} = 10 \cdot \lg\left(\frac{\left\|(\mathbf{p}^{\mathrm{d}})_{M_1} - (\mathbf{p}^{\mathrm{r}})_{M_1}\right\|_2^2}{\left\|(\mathbf{p}^{\mathrm{d}})_{M_1}\right\|_2^2}\right) \tag{34}$$

where $(\mathbf{p}^{\mathrm{d}})_{M_1}$ and $(\mathbf{p}^{\mathrm{r}})_{M_1}$ denote the desired sound field and the reproduced sound field at M_1 matching points, respectively. In the following numerical simulations, the benchmark SFR system, SVD-based algorithm and CMP-based algorithm are chosen as the reference methods. It should be noted that a planar array consists of uniformly placed $N_a = 25$ loudspeakers is used for reproducing in benchmark SFR system and the driven signals of those loudspeakers are estimated by LS. The loudspeaker placement of the proposed AL-LS method, SVD-based and CMP-based method vary when the source frequency and $p_{\max}$ change. This phenomenon is shown in Figure 9. Besides, the number of active loudspeakers both in SVD-based and CMP-based SFR methods is also set to be $N_a = 25$.

First, we make a comparison of NMSE in the power constrained case between the proposed method and the reference algorithms. Herein, $p_{\max}$ is set in the range [0.1, 2] with the step size 0.1 and the source frequency is 800Hz.

As illustrated in Figure 6, the NMSE is gradually decreasing as $p_{\max}$ uniformly increases from 0.1 to 2 for all SFR methods, because higher available power allows for better approximation of the desired field. It can be observed that our proposed AL-LS SFR method has the lowest NMSE in the power range [0.4, 2] except at the very low power. The reason is that the active loudspeaker placement of our proposed AL-LS approach is globally optimal solution. Therefore, the NMSE of the proposed SFR method is the lowest in most cases. In addition, when $p_{\max} \geq 1.5$, the reduction in NMSE is not significant for all systems.

The next experiment compares the NMSE in the frequency range [200, 2000] Hz at $p_{\max} = 0.3$ and $p_{\max} = 2$. In this test, we take the practical application scenarios into account,

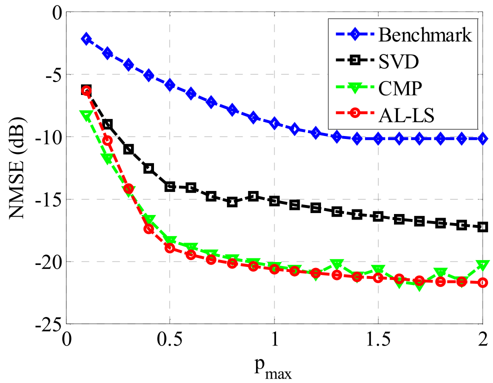

FIGURE 6. The NMSE versus maximum available power $p_{\max}$ inside the listening space for benchmark SFR, SVD-based SFR, CMP-based SFR and the proposed AL-LS SFR. The selected frequency is $f = 800$Hz.

that is, the active loudspeaker locations are chosen at a fixed frequency only once instead of selecting them in every frequency, and all selected active speakers at that frequency will be used in all frequency band. To maintain consistency with the previous simulations, the designed frequency is also set to be $f = 800$Hz, which close to the mid-point of the frequency range. The test results of NMSE at multiple frequencies with different power constraints are shown in Figure 7 and Figure 8.

According to Figure 7, it is observed that the NMSE increases as the frequency increases in terms of all SFR methods, because the loudspeaker weight is calculated under the condition that the desired sound field $(\mathbf{p}^{\mathrm{d}})_M$ is under-sampled at higher frequencies. As seen in Figure 7(a), although the performance of SVD-based SFR is slightly better than that of the proposed method at lower frequencies, the AL-LS approach outperforms the other methods over the frequency range [400, 2000] and it has the comparable error levels with CMP-based SFR. Additionally, the observations in Figure 7(b) maintain consistency with Figure 7(a), i.e., all SFR methods have the similar reproduction quality at lower frequencies, but the superiority of our proposed AL-LS is evident at middle or higher frequencies.

Furthermore, we make a comparison between those four SFR methods with unlimited power. The test conditions are the same as the previous experiment and the results are shown in Figure 8.

Figure 8 indicates that the NMSE of benchmark SFR is the highest. This is because the benchmark system has more dispersed loudspeaker distribution, the total energy of these loudspeakers can not be concentrate on the reproduction zone efficiently. Furthermore, the SVD-based SFR, CMP-based SFR and the proposed method have the subequal reproduction error across the frequency range. The results in this figure are also confirmed with Figure 4.

The last experiment in this subsection compares the effect of the number of matching points M toward the NMSE. We fix the volume of listening cube as $1\mathrm{m}^3$ and increase the total number of matching points M when calculate the placement and weights of the loudspeakers.

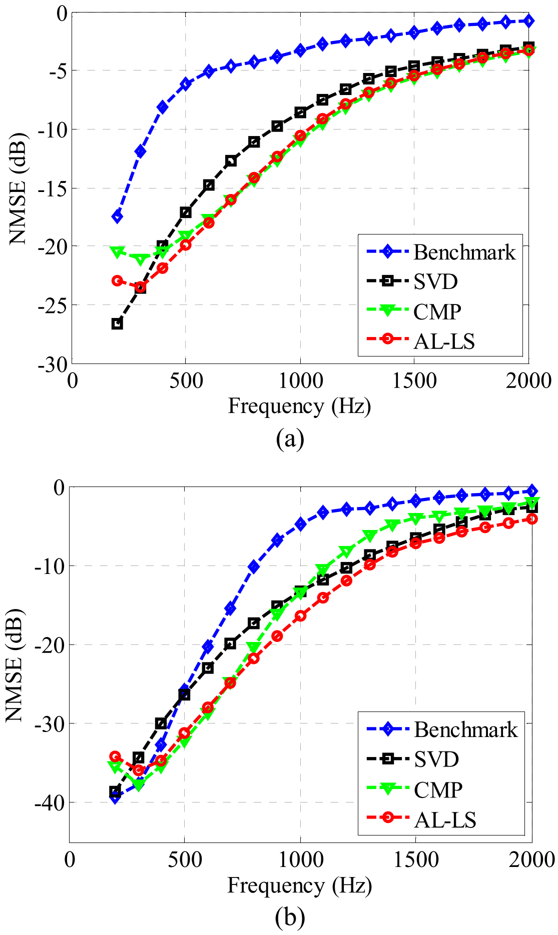

FIGURE 7. The NMSE of the proposed AL-LS based SFR and the reference SFR methods for single-tone primary source over the frequency range [200, 2000] Hz with the interval of 100Hz, when (a) $p_{\max} = 0.3$ and (b) $p_{\max} = 2$.

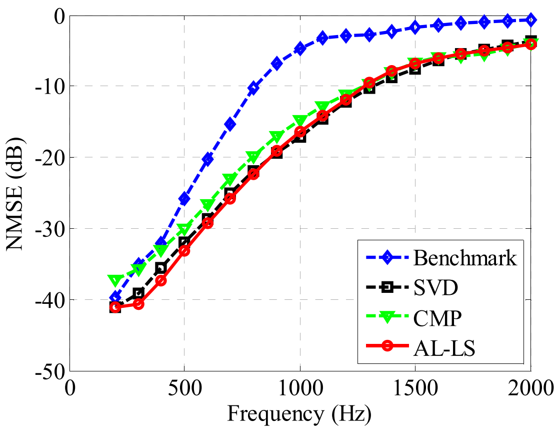

FIGURE 8. Reproduction error of benchmark SFR, SVD-based SFR, CMP-based SFR and AL-LS based SFR without power limitation.

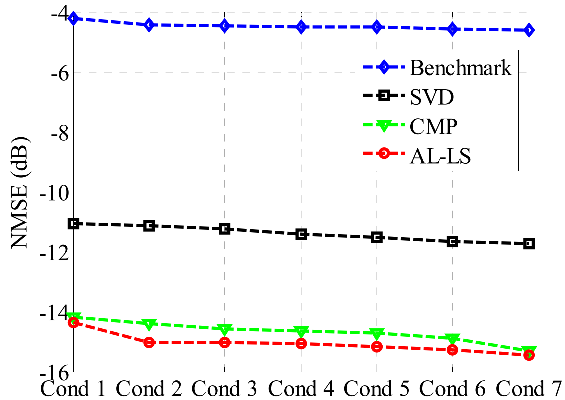

FIGURE 9. Reproduction error of benchmark SFR, SVD-based SFR, CMP-based SFR and AL-LS based SFR with different number of matching points M.

As illustrated in Figure 9, the x-axis label 'Cond 1' represents $M = 125$ $(5 \times 5 \times 5)$, 'Cond 2' represents $M = 512$ $(8 \times 8 \times 8)$, 'Cond 3' represents $M = 1000$ $(10 \times 10 \times 10)$, 'Cond 4' represents $M = 1728$ $(12 \times 12 \times 12)$, 'Cond 5' represents $M = 3375$ $(15 \times 15 \times 15)$, 'Cond 6' represents $M = 8000$ $(20 \times 20 \times 20)$, and 'Cond 7' represents $M = 15625$ $(25 \times 25 \times 25)$. The results shows that when M increases rapidly, the NMSE decrease slightly, and the performance of the proposed method is always better than reference methods.

C. EXHIBITION OF THE PLACEMENT OF THE LOUDSPEAKER ARRAY FOR MONOCHROMATIC PRIMARY SOURCE

From the detailed SFR error analysis, it is demonstrated that the AL-LS approach possess the satisfactory reproduction quality compared to other tested items. The reason is that the procedure of obtaining active loudspeaker locations is a convex optimization problem, the solution of which owns global optimality. In this sub-section, we exhibit the loudspeaker array for different SFR methods. The active loudspeaker locations of our proposed SFR approach and the reference methods in the context of power constraints are presented in Figure 10, where the designed frequency is also $f = 800$Hz.

The left and right column of Figure 10 show the active speaker placement with $p_{\max} = 0.3$ and $p_{\max} = 2$, respectively. Figure 10(a) and 10(b) show the benchmark SFR placement with N_a active loudspeakers distributed uniformly in the array. This structure does not change with frequency f or maximum allowable power $p_{\max}$.

Figure 10(c) and 10(d) show the SVD-based placement, and Figure 10(e) and 10(f) show the CMP-based placement. It should be noted that the loudspeaker positions is recalculated when $p_{\max}$ changes, which may lead to higher computational complexity.

Figure 10(g) and 10(h) indicate the placement of the proposed method with $p_{\max} = 0.3$ and $p_{\max} = 2$. It is clear that the loudspeaker placements both in Figure 10(g) and 10(h) are same. Because the first stage ADMM-based Lasso does not consider the power limitation in active loudspeakers selection and the selected positions are only related to frequency, i.e.,

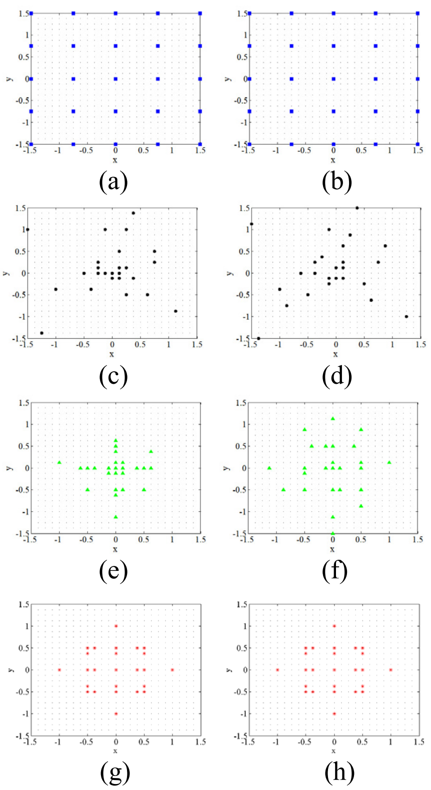

FIGURE 10. Active loudspeaker placement positions under $p_{max} = 0.3$ (left column) and $p_{max} = 2$ (right column) with designed frequency $f = 800$Hz. (a) (b) are benchmark placement, (c) (d) are SVD-based placement, (e) (f) are CMP-based placement and (g) (h) are AL-LS based placement.

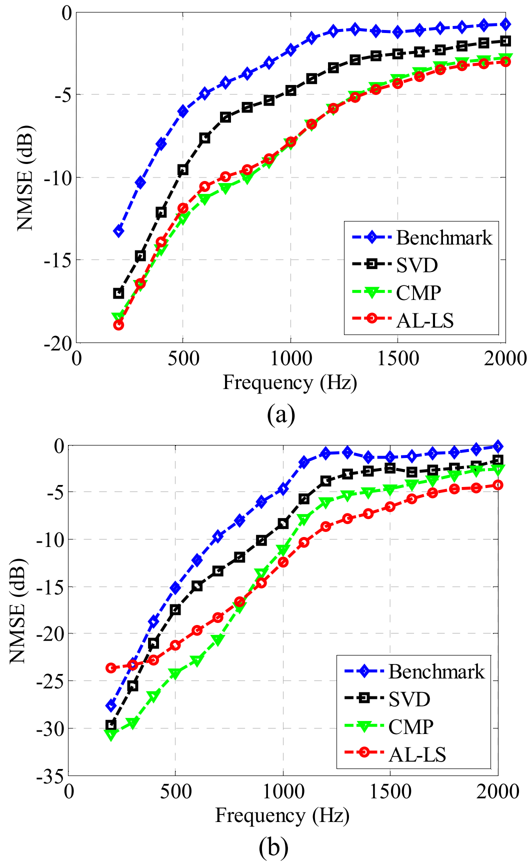

FIGURE 11. The NMSE of the proposed AL-LS based SFR and the reference SFR methods for two primary sources over the frequency range [200, 2000] Hz with the interval of 100Hz, when (a) $p_{max} = 0.3$ and (b) $p_{max} = 2$.

the placement of active loudspeakers is calculated once with a given frequency.

D. SOUND FIELD REPRODUCTION ERROR ANALYSIS FOR MULTIPLE PRIMARY SOURCES

The preceding sub-section IV-B compares our proposed SFR approach with other methods for single-tone primary source sound field reproduction, while this part consider the reproduction of two primary sources.

In this experiment, the two sources are located at $(\sqrt{28}, 0, -6)$ and $(0, \sqrt{28}, -6)$ with complex excitations of $\sqrt{32}$. The reproduced source frequency range is from 200Hz to 2000 Hz. Meanwhile, the designed frequency is 800Hz and the maximum available powers are $p_{max} = 0.3$ and $p_{max} = 2$.

The test results of NMSE at multiple frequencies under different power constraints for two primary sources are shown in Figure 11.

By checking Figure 11(a), our proposed AL-LS SFR model has an analogous NMSE compared with CMP-based method and both of them outperform other methods when $p_{max} = 0.3$.

Figure 11(b) depicts that the proposed SFR approach works better than the reference methods when $f > 800$Hz with $p_{max} = 2$. However, the reproduction quality of AL-LS method degrades when $f < 800$Hz. The reason is due to the face that the process of active loudspeaker selection does not address the power constraint, which may leads to increase the SFR error for a specific p_{max} in certain parts of the listening space.

In summary, the extensive simulations presented above demonstrate that the proposed approach can effectively reduce the reproduction error compared to the reference methods, especially for higher frequencies and an under-sampled sound field.

E. COMPUTATIONAL EVALUATIONS OF THE ACTIVE LOUDSPEAKER SELECTION

The active loudspeaker selection is the most complex part in many SFR system using the Lasso, the complexity of this searching process can be reduced to the great extent by applying our proposed ADMM-based algorithm.

This subsection is devoted to the evaluation of the execution efficiency of the proposed ADMM-based algorithm from a computational perspective. The augmented Lagrangian parameter ρ and the Lasso penalty parameter λ are pre-calculated as 1.0 and 0.021, respectively.

Firstly, for designed frequency $f = 800$Hz, we give the diagram of the number of iterations and the convergence curve of the proposed algorithm in Figure 12, which corresponds to the inputs $\overline{\mathbf{G}} \in \mathbb{R}^{2M\times 2N}$ and $\overline{\mathbf{p}}^{\mathrm{d}} \in \mathbb{R}^{2M}$.

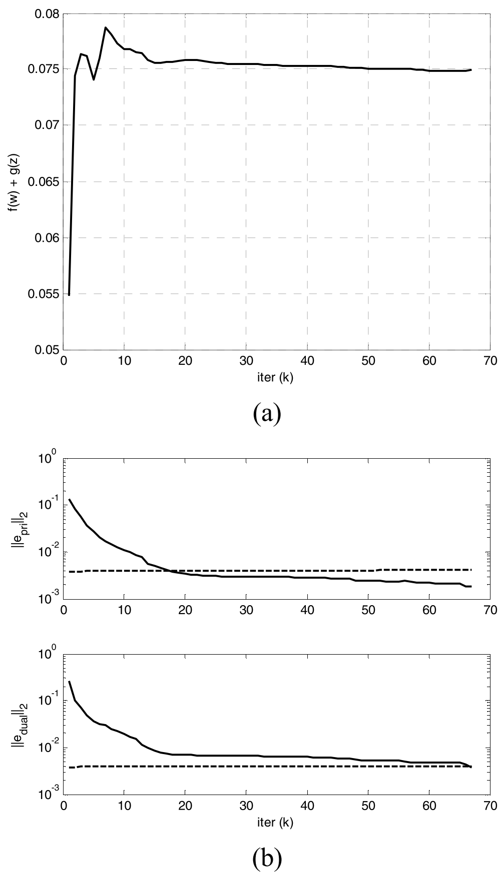

FIGURE 12. **Convergence curves of (a) the objective function (b) the primal and dual residuals of the ADMM. The dashed line in (b) show the upper limit of the two residuals.**

Combined with Figure 12(a) and 12(b), we find that the algorithm converges and terminates after 67 iterations. The objective function $f(\overline{\mathbf{w}}) + g(\overline{\mathbf{z}})$ is nearly convergent and the primal residual $\mathbf{e}_{\mathrm{pri}}^{k1}$ has converged after iterating 20 times. Hence, the ultimate iteration numbers depends on the dual residual $\mathbf{e}_{\mathrm{dual}}^{k1}$. It can be observed that the more iteration numbers leads to the more accurate solutions but the computational costs increase as well. Therefore, we cannot pursuit a single factor such as less iteration numbers or extremely precise results. In consequence, we need to make a trade-off between the accuracy of solutions and the implementation efficiency by selecting the appropriate parameters for termination checks (more details are introduced in [32]).

Finally, the running time of ADMM-based Lasso and CD-based Lasso used for l_1-minimization are compared. The operation environment is set as follows. Hardware: both of the two SFR methods are implemented on a desktop machine with a 3.2 GHz Intel (R) Core (TM) i5-6500 CPU, 8 GB of RAM and an Intel (R) HD Graphics 530 GPU. Software: Equipped with Windows 10 Spring Creators Update 1803 Version and Matlab R2016a. Additionally, due to the SVD-based and CMP-based methods are not Lasso-based SFR systems, this experiment would not consider those two algorithms.

TABLE 1. **Runtime of ADMM-based Lasso and CD-based Lasso when *M* changes from 125 to 15625.**

Number of matching points	Time (s)	
	ADMM-based Lasso	CD-based Lasso
M=125 (5×5×5)	0.128s	3.875s
M=512 (8×8×8)	0.672s	7.602s
M=1000 (10×10×10)	0.759s	10.433s
M=8000 (20×20×20)	1.296s	42.489s
M=15625 (25×25×25)	2.487s	77.031s
Speed-up	11.31~32.79 times	

To examine the execution speed of ADMM-based Lasso and CD-based Lasso, we fix the dimensionality of N as 625 (the number of all candidate loudspeakers) and N_a as 25 (the number of active loudspeakers). The matching points are set inside the listening space of 1m × 1m × 1m cube, and the number of these points changes from 125 to 15625. The results are shown in Table 1.

Table 1 lists the computing time of two different algorithms when M changes from 125 to 15625. Obviously, the runtime rises with the incensement of data volume. However, the speed of ADMM-based Lasso can reach up to 32.79 times faster than CD-based Lasso when $M = 8000$. Even in the worst condition, the ADMM algorithm is still 11.31 times faster than CD-based Lasso. This implies that ADMM-based Lasso is more advantageous in practical large-scale or low-latency SFR applications.

In general, the reason guarantees the efficiency of the proposed SFR technique is two-fold. The first is the superiority of ADMM itself, and the second is the trick for matrix inversion computation. Recall that the real-valued Green's function matrix $\overline{\mathbf{G}}$ has $2M \times 2N$ entries, when $\overline{\mathbf{G}}$ is fat (i.e., $M = 125$ or 512, $N = 625$), the Sherman-Morrison-Woodbury inversion formula (described in subsection III-B) can be further utilized to reduce the computational costs.

V. CONCLUSION

This paper proposes an efficiently 3D SFR technique through the optimization of omni-directional loudspeaker locations and weights in a planar array based on two-stage Lasso-LS, which has the potential of being applied to immersive audio applications. At the first stage, the ADMM algorithm is adopted to overcome the persistent problem of the high computational complexity of Lasso, in order to promote the efficiency during the optimal loudspeaker selection. Regularized LS is then employed to control the total active loudspeaker power and estimate the final loudspeaker weight. Simulation results indicate that our proposed SFR approach has the comparable reproduction quality with CMP-based SFR, and outperforms the SVD-based SFR and Benchmark systems under various power limitation across the frequency range. The results also confirmed that the execution speed of the proposed ADMM-based Lasso solver is much faster than the state-of-the-art Lasso solver. Summing up, the advantages of the proposed SFR method is the relatively low reproduction error and the high efficiency when we calculate the loudspeaker weight. Future research could include the investigation of AL-LS in reverberation environment.

REFERENCES

[1] J. Ahrens, *Analytic Methods of Sound Field Synthesis*. Berlin, Germany: Springer, 2012.

[2] J. Ahrens and S. Spors, "Sound field reproduction using planar and linear arrays of loudspeakers," *IEEE Trans. Audio, Speech, Language Process.*, vol. 18, no. 8, pp. 2038–2050, Nov. 2010.

[3] J. Donley, C. Ritz, and W. B. Kleijn, "Reproducing personal sound zones using a hybrid synthesis of dynamic and parametric loudspeakers," in *Proc. Asia–Pacific Signal Inf. Process. Assoc. Annu. Summit Conf.*, Jeju, South Korea, Dec. 2016, pp. 1–5.

[4] A. Gupta and T. D. Abhayapala, "Three-dimensional sound field reproduction using multiple circular loudspeaker arrays," *IEEE Trans. Audio, Speech, Language Process.*, vol. 19, no. 5, pp. 1149–1159, Jul. 2011.

[5] P.-A. Gauthier and A. Berry, "Sound-field reproduction in-room using optimal control techniques: Simulations in the frequency domain," *J. Acoust. Soc. Amer.*, vol. 117, pp. 662–678, Feb. 2005.

[6] M. Kolundzija, C. Faller, and M. Vetterli, "Reproducing sound fields using MIMO acoustic channel inversion," *J. Audio Eng. Soc.*, vol. 59, no. 10, pp. 721–734, 2011.

[7] H. Khalilian, I. V. Bajic, and R. G. Vaughan, "A simulation study of a three-dimensional sound field reproduction system for immersive communication," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 25, no. 5, pp. 980–995, May 2017.

[8] S. Koyama, S. Shimauchi, and H. Ohmuro, "Sparse sound field representation in recording and reproduction for reducing spatial aliasing artifacts," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Florence, Italy, May 2014, pp. 4443–4447.

[9] W. Zhang, P. N. Samarasinghe, H. Chen, and T. D. Abhayapala, "Surround by sound: A review of spatial audio recording and reproduction," *Appl. Sci.*, vol. 7, no. 5, p. 532, May 2017.

[10] J. Y. Hong, J. He, B. Lam, R. Gupta, and W.-S. Gan, "Spatial audio for soundscape design: Recording and reproduction," *Appl. Sci.*, vol. 7, no. 6, p. 627, Jun. 2017.

[11] T. Betlehem and C. Withers, "Sound field reproduction with energy constraint on loudspeaker weights," *IEEE Trans. Audio, Speech, Language Process.*, vol. 20, no. 8, pp. 2388–2392, Oct. 2012.

[12] H. Khalilian, I. V. Bajić, and R. G. Vaughan, "3D sound field reproduction using diverse loudspeaker patterns," in *Proc. IEEE Int. Conf. Multimedia Expo Workshops*, San Jose, CA, USA, Jul. 2013, pp. 1–4.

[13] P.-A. Gauthier, P. Lecomte, and A. Berry, "Source sparsity control of sound field reproduction using the elastic-net and the lasso minimizers," *J. Acoust. Soc. Amer.*, vol. 141, no. 4, pp. 2315–2326, Apr. 2017.

[14] B. K. Natarajan, "Sparse approximate solutions to linear systems," *SIAM J. Comput.*, vol. 24, no. 2, pp. 227–234, 1995.

[15] G. N. Lilis, D. Angelosante, and G. B. Giannakis, "Sound field reproduction using the Lasso," *IEEE Trans. Audio, Speech, Language Process.*, vol. 18, no. 8, pp. 1902–1912, Nov. 2010.

[16] Q. Feng, F. Yang, and J. Yang, "Time-domain sound field reproduction using the group Lasso," *J. Acoust. Soc. Amer.*, vol. 143, no. 2, pp. EL55–EL60, Jan. 2018.

[17] N. Radmanesh and I. S. Burnett, "Generation of isolated wideband sound fields using a combined two-stage Lasso-LS algorithm," *IEEE Trans. Audio, Speech, Language Process.*, vol. 21, no. 2, pp. 378–387, Feb. 2013.

[18] N. Radmanesh, I. S. Burnett, and B. D. Rao, "A Lasso-LS optimization with a frequency variable dictionary in a multizone sound system," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 24, no. 3, pp. 583–593, Mar. 2016.

[19] H. Khalilian, I. V. Bajić, and R. G. Vaughan, "Towards optimal loudspeaker placement for sound field reproduction," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Vancouver, BC, Canada, May 2013, pp. 321–325.

[20] H. Khalilian, I. V. Bajić, and R. G. Vaughan, "Comparison of loudspeaker placement methods for sound field reproduction," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 24, no. 8, pp. 1364–1379, Aug. 2016.

[21] H. Khalilian, I. V. Bajić, and R. G. Vaughan, "Loudspeaker placement for sound field reproduction by constrained matching pursuit," in *Proc. IEEE Workshop Appl. Signal Process. Audio Acoust.*, New Paltz, NY, USA, Oct. 2013, pp. 1–4.

[22] X. Li, L. Mo, X. Yuan, and J. Zhang, "Linearized alternating direction method of multipliers for sparse group and fused LASSO models," *Comput. Statist. Data Anal.*, vol. 79, pp. 203–211, Nov. 2014.

[23] N. Radmanesh and B. D. Rao, "Frequency-based customization of multizone sound system design," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Shanghai, China, Mar. 2016, pp. 375–379.

[24] E. J. Candès and Y. Plan, "Near-ideal model selection by ℓ_1 minimization," *Ann. Statist.*, vol. 37, no. 5A, pp. 2145–2177, 2009.

[25] E. J. Candès and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 21–30, Mar. 2008.

[26] J. Friedman, T. Hastie, H. Höfling, and R. Tibshirani, "Pathwise coordinate optimization," *Ann. Appl. Statist.*, vol. 1, no. 2, pp. 302–332, 2007.

[27] P. Yin, Y. Lou, Q. He, and J. Xin, "Minimization of ℓ_{1-2} for compressed sensing," *SIAM J. Sci. Comput.*, vol. 37, no. 1, pp. A536–A563, 2015.

[28] R. Zhang and J. Kwok, "Asynchronous distributed ADMM for consensus optimization," in *Proc. Int. Conf. Mach. Learn.*, Beijing, China, Jun. 2014, pp. 1701–1709.

[29] N. Parikh and S. Boyd, "Block splitting for distributed optimization," *Math. Program. Comput.*, vol. 6, no. 1, pp. 77–102, Mar. 2014.

[30] D. L. Donoho and J. M. Johnstone, "Ideal spatial adaptation by wavelet shrinkage," *Biometrika*, vol. 81, no. 3, pp. 425–455, 1994.

[31] G. H. Golub and C. F. Van Loan, *Matrix Computations*, vol. 3. Baltimore, MD, USA: The Johns Hopkins Univ. Press, 2012.

[32] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, Jan. 2011.

[33] S. L. Wang and L. Z. Liao, "Decomposition method with a variable parameter for a class of monotone variational inequality problems," *J. Optim. Theory Appl.*, vol. 109, no. 2, pp. 415–429, 2001.

MAOSHEN JIA (M'13–SM'15) was born in China in 1982. He received the B.E. degree in electronic information engineering from Hebei University, Baoding, China, in 2005, and the Ph.D. degree in electronic science and technology from the Beijing University of Technology, Beijing, China, in 2010.

From 2010 to 2015, he was a Lecturer with the School of Electronic Information and Control Engineering, Beijing University of Technology, where he has been an Associate Professor with the Faculty of Information Technology since 2016. He has authored over 30 articles. His current research interests include multichannel audio signal processing and sound field reproduction.

JIAMING ZHANG (S'16) was born in China in 1994. He received the B.E. degree in electronic information engineering from the Beijing University of Technology, Beijing, China, in 2016, where he is currently pursuing the M.S. degree in information and communication engineering. His current research interests are sound field reproduction.

YUXUAN WU (S'18) was born in Beijing, China, in 1995. He received the B.E. degree in electronic information engineering from the Beijing University of Technology, Beijing, in 2017, where he is currently pursuing the M.S. degree in information and communication engineering. His current research interests are sound field reproduction.

JING WANG (M'10) was born in China in 1980. She received the B.S. and Ph.D. degrees in communication and information system from the Beijing Institute of Technology, Beijing, China, in 2002 and 2007, respectively.

She has ever been a Visiting Scholar with The Chinese University of Hong Kong. She is currently an Associate Professor with the School of Information and Electronics, Research Institute of Communication Technology, Beijing Institute of Technology. Her current research is speech and audio signal processing, multimedia quality assessment, and multimedia communication.