Sound field reproduction using the lasso and OMP method with fixed directivity loudspeakers

Bokai Du[1], Michael Kohnen[2], Michael Vorländer[2]

[1] Northwestern Polytechnical university,710072, Xi'an, E-Mail:dubokai123@mail.nwpu.edu.cn

[2] Institute of Technical Acoustics,52072, Aachen, E-Mail: mko@akustik.rwth-aachen.de

Abstract

A spherical array of loudspeakers is usually installed in sound field reproduction system. For these systems, the loudspeaker signal design is traditionally relied on least-squares(LS) and high-order Ambisonics (HOA) methods. In some cases of sound incidence to the sweet spot, the desired sound field comes only from one or a small number of sources. Taking this sparsity into consideration, as a result, the desired sound field can be reproduced when only a small number of loudspeakers in the spherical array are activated. In this paper, the performance of the least-absolute shrinkage and selection operator (Lasso) and orthogonal matching pursuit (OMP) are compared with the LS method. Simulation results were investigated to evaluate the performance differences between the Lasso and OMP method with the LS method. Furthermore the effect of using fixed directivity loudspeakers were evaluated in terms of simulation.

Introduction

Sound field reproduction(SFR) is an important aspect for acoustic virtual reality, where loudspeaker array is used to reproduce not only the content of the sound, but also the spatial property of the sound field.

For modern reproduction system, there are three most popular approaches to SFR: Wave field synthesis(WFS) [1], High-order ambisonics(HOA) [2]and direct approximation. WFS is based on Hyugen's principle and the Kirchoff-Helmholtz(K-H) boundary surface integral equation and a planar loudspeaker array is often used. The second method decomposes the primary sound field into spherical harmonics domain and mode-matching is applied to design the loudspeaker signals. In direct approximation, the desired sound field is sampled using an array of microphones and the design of proper loudspeaker signals is achieved by minimize the difference between target sound pressures and reproduced sound pressures stirred by loudspeaker array at sample points. Usually, least-square(LS) criterion is used to design the loudspeaker signals and limit the loudspeaker energy.

However, the accuracy of LS method depends on the number of microphones. Generally speaking, in order to achieve better performance, especially for the high frequency part of sound field, a large number of microphones is required to make sample points dense enough. Futhermore, the reproduction problem is often ill-posed and regularization is necessary. However, in some cases the desired sound field comes only from one or a small number of sources, which means the sound field can be reproduced using just a limited number of loudspeakers. As a result, Lasso method was proposed to make full use of this sparsity to achieve better performance with less microphones [3].

In this work, another algorithm called orthogonal matching pursuit (OMP) [4] is applied to solve reproduction problem. The matching pursuit algorithm is an iterative approach to construct a linear expansion to approximate the original input vector using a given set of vectors called a dictionary. For every step, a vector from the dictionary which is most correlated to the current approximation error (for the first step, the error is the original input vector) is selected. Fixed directivity loudspeaker is modeled as a weighted combination of monopole and dipole [5]. The performance of fixed-directivity loudspeaker array is also investigated. Its performance is compared with the omnidirectional one.

The paper is organized as follows. In section 2, the sound filed reproduction problem is stated. Least-square and matching pursuit method will also be presented in this section. For section 3, the OMP method is introduced and applied to the reproduction problem. The simulation results are presented in section 4. Concluding remarks are given in section 5.

Problem statement and theory

In this paper, a spherical loudspeaker array is used. Assuming that the number of matching points in the listening area and the number of loudspeakers are M and L, respectively. In the frequency domain, the sound pressure can be expressed as:

$$p_{m,l}(\omega) = w_l(\omega) G_{m,l}(\omega) \tag{1}$$

where $w_l(\omega)$ is l-th loudspeaker's driving signal and $G_{m,l}(\omega)$ is the acoustic transfer function(ATF) between the l-th loudspeaker and m-th matching point. In this work, the reproduction system is assumed to be placed in free field. Therefore, for a monopole placed at position $\vec{r}_l$ is expressed as free space Green function:

$$G_{m,l}(\omega) = \frac{1}{4\pi} \cdot \frac{e^{jkr_{m,l}}}{r_{m,l}} \tag{2}$$

Where $r_{m,l} = \|\vec{r}_l - \vec{r}_m\|_2$ and $\vec{r}_m$ is the position of m-th matching point. $k = \omega / c$ is the wave number, and c is the sound speed in the air.

The sound pressure at each matching point is the summation of the sound pressure stirred by each loudspeaker of the reproduction array:

$$p_m = \sum_{l=1}^{L} w_l G_{m,l} \tag{3}$$

For the whole reproduction system, the reproduced sound pressure at matching points sound pressures is expressed in matrix form:

$$\mathbf{p}_{rep} = \mathbf{G}\mathbf{w} \tag{4}$$

Where $\mathbf{p}_{rep}$ is a M by 1 vector of reproduce sound pressure at matching points p_m, **w** is an L by 1 vector of loudspeaker weights w_l, **G** is a M by L matrix whose (m,l)-th entry is equal to $G_{m,l}(\omega)$.

2.1 Sound field reproduction using Least-Square

When we use LS method, sound field reproduction is achieved by minimizing the squared approximation error between the target sound pressure $\mathbf{p}_{des}$ and reproduced sound pressure $\mathbf{p}_{rep}$, namely:

$$w: \arg\min_{w} \|\mathbf{p}_{rep} - \mathbf{G}\mathbf{w}\|_2^2 + \lambda \|\mathbf{w}\|_2^2 \tag{5}$$

Where λ is a preselected regularization parameter,it is used to avoid obtaining some physical unrealisable solution (large magnitude), as a result, the driving signal can be easily obtained as:

$$\mathbf{w} = \left[\mathbf{G}^{\mathbf{H}}\mathbf{G} + \boldsymbol{\lambda}\mathbf{I}\right]^{-1} \mathbf{G}^{\mathbf{H}} \mathbf{p}^{des} \tag{6}$$

2.2 Lasso solution

LS-based SFR falls shorts in limiting the number of active loudspeakers and allocates the power to all loudspeakers of the array. For the case when desired sound field is stirred by one or a small number of sources, sound field can be reproduced with a small number of loudspeakers. Lasso method was proposed to select a limited number of loudspeakers in the array to reproduce the target sound field, it shows better performance than the LS method in this case, especially when the sound field is under-sampled. A sparse solution is obtained by solving the following problem [3]:

$$\mathbf{w} = \arg\min \frac{1}{2} \|\mathbf{G}\mathbf{w} - \mathbf{p}^{des}\|_2^2 + \lambda \|\mathbf{w}\|_1 \tag{7}$$

Where $\|\mathbf{w}\|_1 = \sum_{l=1}^{L} |w_l|$. A spares **w** is obtained, which means only a limited number of loudspeakers are active. The positive λ regulated the sparsity level of the solution, and $\lambda \in [0, \|\mathbf{G}\mathbf{p}^{des}\|_\infty)$. when $\lambda = 0$, it goes back to a least-square problem, when $\lambda > \|\mathbf{G}\mathbf{p}^{des}\|_\infty$, all of elements in solution are zero.

2.3 Orthogonal matching pursuit

Matching pursuit (MP) [6] is an iterative procedure for selecting a linear combination of vectors approximate a given vector. In order to make matching pursuit converge faster, orthogonal matching pursuit is used in this paper. For MP, atoms in dictionary are usually not orthogonal to each other, as a result, residual error is not orthogonal to previously selected atoms, which means one atom may be selected twice. In OMP, for every step, residual error is always orthogonal to all of selected atoms. Compared with MP, the same atom will not be selected twice in OMP and the converge speed is faster. At the same time, the power constrain should be taken into consideration to limit the array effort. The steps of this algorithm are described in detail below.

Step 1. For a fixed spherical loudspeaker array with L loudspeakers, loudspeakers are indexed by $l \in \{1,2,...L\}$. N loudspeaker will be selected to reproduce the target sound field.

Step 2. Compute the ATF at M sample points from the *l-th* loudspeaker and the results are stored in the g_l. The dictionary is $G = [g_1, g_2 \cdots g_L]$.

Step 3. let n=1 and $\mathbf{e}_1 = \mathbf{p}^{des}$.

Step 4. find the vector in the dictionary that has the largest inner product with the current residual error vector $\mathbf{e}_n$, denote it as $\boldsymbol{\varphi}_n$ .Record the index of selected vector to identify the corresponding loudspeaker.

Step 5. combine all of selected vectors into a matrix $\boldsymbol{\varphi}_t = [\boldsymbol{\varphi}_1, \boldsymbol{\varphi}_2, ...\boldsymbol{\varphi}_n]$, and define the orthogonal projection operator of column vector of ϕ_t as:

$$\mathbf{P} = \boldsymbol{\varphi}_t (\boldsymbol{\varphi}_t^H \boldsymbol{\varphi}_t^H)^{-1} \boldsymbol{\varphi}_t^H \tag{8}$$

Step 6. Find the corresponding complex weights of the selected n loudspeakers under a certain power constrain, selected loudspeaker weights vector can be expressed as:

$$\mathbf{w} = (\boldsymbol{\varphi}_t^H \boldsymbol{\varphi}_t + \lambda \mathbf{I})^{-1} \boldsymbol{\varphi}_t^H \mathbf{e}_n \tag{9}$$

Where λ is a parameter to limit the total energy of loudspeakers.

Step 7. Compute the new approximation error:

$$\mathbf{e}_{n+1} = \mathbf{e}_n - \mathbf{P}\mathbf{e}_n \tag{10}$$

Step 8. if n=N,stop; otherwise n=n+1 and go back to step 4.

At last, the given vector $\mathbf{p}^{des}$ can be expressed as:

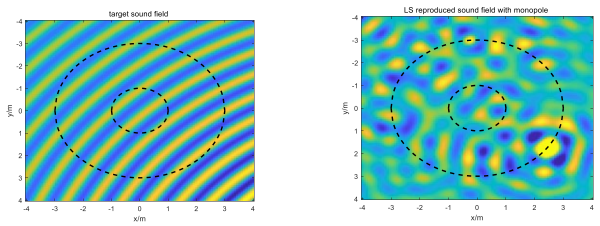
$$\mathbf{p}^{des} = \sum_{n=1}^{N} \mathbf{w}(n)\boldsymbol{\varphi}_n + e_{N+1} \tag{11}$$

In step 6, the parameter λ can be calculated using method proposed in [7] to limit the energy.

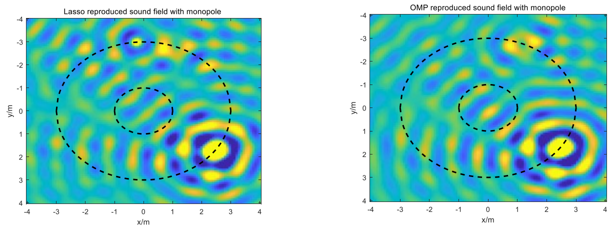
Simulation results

In this simulation, the target sound field is stirred by a primary point source locating at the position[10,10,0] in the free field. A spherical array with 36 microphones is used to sample the sound field, the radius of this microphone array is 1m. The minimum distance between two microphones in this array is 0.59m, corresponding to the frequency 566Hz, which means the highest reproduction frequency using the sound pressure data from this array without the spatial aliasing is 283Hz. On the other hand, a 3m spherical array with 64 loudspeakers which are all modeled as point sources is used to reproduce the target sound field.

The real part of target sound field and reproduced sound fields on the horizontal plane using aforementioned three methods at 390Hz are shown in figure 1. According to this figure, the field reproduced by those two sparse methods shows better reproduction performance than the LS method. Especially, the wave front of the sound field in the listening region is better reproduced by these two sparse methods.



(a) (b)



(c) (d)

Figure 1: The real part of (a) the target sound field. (b) the field reproduced by LS method. (c) the field reproduced by Lasso method. (d) the field reproduced by OMP method.

And we also investigate the approximation error of these three methods over a range of frequencies from 250Hz to 600Hz. The results are shown in figure 2. The relative error is defined as:

$$Error(dB) = 10\log_{10} \frac{\int_{\Omega} \| p_{re} - p_{pre} \|_2^2 dV}{\int_{\Omega} \| p_{pre} \|_2^2 dV} \tag{12}$$

In this simulation, the total energy of loudspeakers are constrained to be same for these three methods to ensure the reproductions are conducted under the same condition. According to this figure, the performances of all methods degrades as the frequency increase and the reproduction relative error of LS method is higher than errors of those two sparse methods. As mentioned before, the sound field measured by this microphone array is under-sampled and this means sparse methods show better performance than LS method in this case.

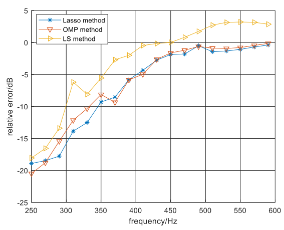
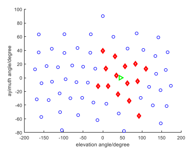
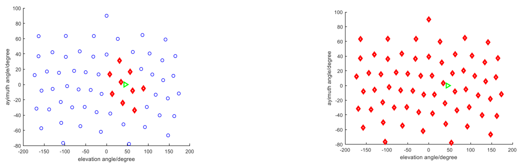


Figure 2: The reproduction relative error comparison of LS, Lasso and OMP method from 250Hz to 590Hz. The reproduction using these three methods are conducted under the same loudspeaker energy constrain.

The position of activated loudspeakers for these three methods are shown in figure 3. The green mark denotes the orientation of the primary source. Blue circles denote orientations of loudspeakers and red marks denote the orientations of activated loudspeakers. As shown in this figure, only loudspeakers around the orientation of green mark are activated for sparse methods and this reduces the 'blurry' caused by the application of LS method.



(a)



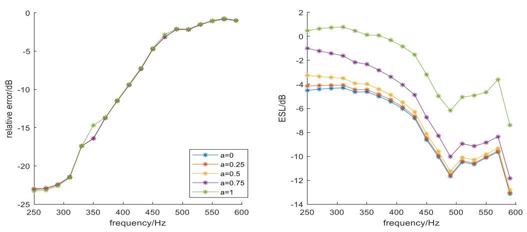
(b) (c)

Figure 3: The position of activated loudspeakers for (a) the Lasso method. (b) the OMP method. (c) the LS method.

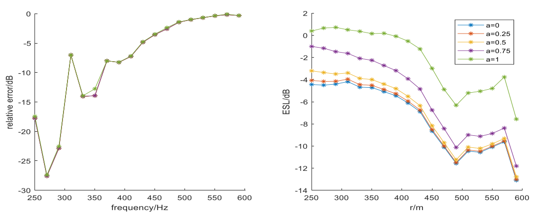
For a real reproduction system which is usually placed in 'listening room', the reproduction performance in the target region will be affected or even destroyed by the superimposing of reflection from walls. If reproduction loudspeakers radiate less energy to the space out of the array, the reproduction performance will be less affected by reflections. A metric to evaluate the energy radiates towards the space out of the array is defined as the the energy ratio of target and reproduced sound field energy on the same spherical surface (out of the loudspeaker array).

$$ESL = 10\log_{10}\frac{\int_{S_{r0}}\|p_{re}\|_2^2 dS}{\int_{S_{r0}}\|p_{pre}\|_2^2 dS} \tag{14}$$

It was proposed to use fixed directivity loudspeaker in spherical harmonic domain to achieve energy radiation reduction. The loudspeaker is modeled as a weighted combination of monopole and dipole. A parameter a is defined, which varies from 0 to 1, corresponding to the directivity from dipole to monopole [5]. In this paper, the effect of loudspeaker directivity on the energy outwards radiation when using Lasso and LS is compared, the result is shown in figure 4.



(a)



(b)

Figure 4: Reproduction relative error in the reproduction area and external sound level at the spherical surface r=3.2m for (a) the Lasso method (b) the LS method.

As shown in figure 4, for both Lasso and LS method, with the directivity parameter a varies from 0 to 1, the relative reproduction error is almost the same value for three methods on the investigated frequency range but the ESL increases. This means that a proper design of loudspeaker directivity can achieve reproduction performance and the control of outwards radiation energy to some extent at the same time.

Conclusion

In this paper, the OMP method is applied to sound field reproduction and its reproduction performance is compared with Lasso and LS method. At the same time, the effect of loudspeaker directivity on the reproduction performance and ESL is also investigated. Simulation results indicate sparse method reduce the reproduction 'blurry' caused by the LS method and reduce relative reproduction error especially when the target sound field (stirred by single source) is not sufficiently sampled. And the ESL can be controlled to some extent by using a properly designed fixed directivity loudspeaker.

References

[1] Boone M M, Verheijen E N G, Van Tol P F. Spatial sound-field reproduction by wave-field synthesis[J]. Journal of the Audio Engineering Society, 1995, 43(12): 1003-1012.

[2] Poletti M A. Three-dimensional surround sound systems based on spherical harmonics[J]. Journal of the Audio Engineering Society, 2005, 53(11): 1004-1025.

[3] Lilis G N, Angelosante D, Giannakis G B. Sound field reproduction using the lasso[J]. IEEE Transactions on Audio, Speech, and Language Processing, 2010, 18(8): 1902-1912.

[4] Needell D, Vershynin R. Uniform uncertainty principle and signal recovery via regularized orthogonal matching pursuit[J]. Foundations of computational mathematics, 2009, 9(3): 317-334.

[5] Poletti M, Fazi F M, Nelson P A. Sound-field reproduction systems using fixed-directivity loudspeakers[J]. The Journal of the Acoustical Society of America, 2010, 127(6): 3590-3601.

[6] Khalilian H, Bajic I V, Vaughan R G. Loudspeaker placement for sound field reproduction by constrained matching pursuit[C]//2013 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics. IEEE, 2013: 1-4.

[7] Betlehem T, Withers C. Sound field reproduction with energy constraint on loudspeaker weights[J]. IEEE Transactions on Audio, Speech, and Language Processing, 2012, 20(8): 2388-2392.