Sparsity-based sound field reconstruction

Shoichi Koyama\(^1,2,*\)

\(^1\)Graduate School of Information Science and Technology, The University of Tokyo, 7–3–1 Hongo, Bunkyo-ku, Tokyo, 113–8656 Japan
\(^2\)JST, PRESTO

Abstract: Estimating and interpolating a sound field from measurements using multiple microphones are fundamental tasks in sound field analysis for sound field reconstruction. The sound field reconstruction inside a source-free region is achieved by decomposing the sound field into plane-wave or harmonic functions. When the target region includes sources, it is necessary to impose some assumptions on the sources. Recently, it has been increasingly popular to apply sparse representation algorithms to various sound field analysis methods. In this paper, we present an overview of sparsity-based sound field reconstruction methods and also demonstrate their application to sound field recording and reproduction.

Keywords: Acoustic holography, Compressed sensing, Interpolation, Sound field reconstruction, Sparse representation

PACS number: 43.60.+d [doi:10.1250/ast.41.269]

1. INTRODUCTION

Sound field reconstruction, which is a fundamental inverse problem in acoustics, aims to estimate and interpolate a sound field inside a target region by measurements using multiple microphones. This can be applied to various acoustic measurement tasks, e.g., visualization of an acoustic field [1,2], identification of sound sources [3,4], and capture of a sound field for reproduction using loudspeakers or headphones [5–7].

The sound field reconstruction problem can be classified into two cases: whether the target region includes sound sources or not. The sound field inside a source-free region can be reconstructed by decomposing the measurements using multiple microphones into element solutions of the Helmholtz equation such as plane-wave and harmonic functions, which is also well known as a theoretical foundation of Fourier acoustics [8]. It is also possible to apply the equivalent source method [9,10], which is based on the representation obtained using Green’s function on a fictitious boundary enclosing the target region. In contrast, the sound field reconstruction inside a region including sound sources is an ill-posed problem. This requires some assumptions on the sources since the source distribution can be any function. Similarly to the case of the source-free region, a key to solving this problem is to decompose the captured sound field into element solutions of the Helmholtz equation [11,12].

In recent years, sparse representation of acoustic fields has been proved to be effective in several applications, such as acoustic holography [13], source localization [14], estimation of room transfer functions [15], and sound field capturing [16,17], owing to the recent advancement of sparse decomposition algorithms in the context of compressed sensing [18]. These methods are based on the representation of acoustic fields by a limited number of element solutions of the Helmholtz equation, which is referred to as sparse sound field decomposition. Such decomposition practically improves the spatial resolution in the measurement and estimation. In this paper, we present an overview of the sparsity-based sound field reconstruction methods. In addition, an application of sparse sound field decomposition to recording and reproduction is demonstrated.

2. SOUND FIELD RECONSTRUCTION INSIDE SOURCE-FREE REGION

We here consider a three-dimensional region of interest \( \Omega \subseteq \mathbb{R}^3 \) not including any sources as shown in Fig. 1. The pressure field inside \( \Omega \) is denoted as \( u(r,k) \), where \( r \in \Omega \) is the position vector, \( k = \omega/c \) is the wave number, \( \omega \) is the angular frequency, and \( c \) is the sound velocity. Then, \( u(r,k) \) satisfies the following homogeneous Helmholtz equation [8]:

\[
(\nabla^2 + k^2)u(r,k) = 0, \tag{1}
\]

where \( \nabla^2 \) denotes the Laplacian. On the room surface outside \( \Omega \), some boundary conditions may also be imposed
on \(u(r,k)\). The objective of sound field reconstruction is to estimate \(u(r,k)\) for continuous \(r \in \Omega\) from pressure measurements using multiple microphones arranged in \(\Omega\). The positions of the microphones are denoted as \(r_m\) \((m \in \{1, \ldots, M\})\). The vector of the pressure measurements at \(r_m\) is denoted as \(y(k) \in \mathbb{C}^M\). We hereafter omit the argument \(k\) for notational simplicity.

A typical strategy to estimate \(u(r)\) is to decompose it using plane-wave or harmonic functions since the reconstructed sound field can also satisfy the homogeneous Helmholtz equation [8]. The Herglotz wave function is the form employed to represent \(u(r)\) using the superposition of plane waves as [19]

\[
u(r) = \int_{S^2} \gamma(\eta)e^{ik\eta \cdot r} d\eta, \tag{2}\]

where the integral is over the unit sphere \(S^2\), \(\gamma(\cdot)\) is the density, and \((\cdot, \cdot)\) denotes the inner product. By discretizing the unit sphere, we can estimate the discrete form of \(\gamma(\eta)\) by solving the linear equation of \(\gamma(\eta)\) and \(y\) using the matrix of plane wave functions, which enables the reconstruction of the continuous \(u(r)\). Another example is the use of a spherical wave function. In spherical coordinates, \(u(r, \theta, \phi)\) of the interior field can be expanded around the expansion center set in \(\Omega\) as

\[
u(r) = \sum_{\nu=0}^{\infty} \sum_{\mu=-\nu}^{\nu} \alpha_{\nu,\mu} j_{\nu}(kr) Y_{\nu,\mu}(\theta, \phi), \tag{3}\]

where \(\alpha_{\nu,\mu}\) are the expansion coefficients, \(j_{\nu}(\cdot)\) is the \(\nu\)-th order spherical Bessel function of the first kind, and \(Y_{\nu,\mu}(\cdot)\) is the spherical harmonic function of order \(\nu\) and degree \(\mu\). For the exterior field, the spherical Bessel function is replaced by the spherical Hankel function. By constructing a matrix of the spherical wave functions with a truncation order of \(\nu\), we can estimate the coefficients \(\alpha_{\nu,\mu}\) by solving the linear equation using the pressure measurements \(y[20]\). Then, the continuous pressure field \(u(r)\) can be reconstructed by using the estimated expansion coefficients \(\alpha_{\nu,\mu}\).

Although this procedure depends on the truncation order of \(\nu\) and the predefined expansion center, Ueno et al. proposed an estimation method for \(\alpha_{\nu,\mu}\) independent of them by using spherical wave functions of infinite order [21]. It is worth noting that this method corresponds to the kernel ridge regression using the reproducing kernel of the zeroth-order spherical Bessel function of the first kind [22].

Another well-known method for reconstructing a homogeneous sound field is the equivalent source method [9,10]. The theoretical basis of the equivalent source method is founded on sound field representation using a single-layer potential [19]. We suppose a fictitious region \(D\) including \(\Omega\), i.e., \(D \supset \Omega\) (see Fig. 1). The sound field inside \(D\) can be represented by using the free-field Green’s function on the boundary of \(D\), i.e., \(\partial D\), as

\[
u(r) = \int_{\partial D} \phi(r') G(r'r) d\Gamma' \quad (r \in D), \tag{4}\]

where \(\phi(\cdot)\) is the potential density and \(G(\cdot)\) is the free-field Green’s function. Here, \(G(\cdot)\) is defined as

\[
G(r') = \frac{e^{ik\|r-r'\|}}{4\pi\|r-r'\|}, \quad (5)
\]

which corresponds to the transfer function of the monopole from \(r'\) to \(r\). Analytical solutions of \(\phi(\cdot)\) can be derived only when the array geometry of the microphones is a simple shape [8]. In the case that the microphones are not simply arranged, by discretizing \(\partial D\) into \(K\) monopoles, we can formulate a linear equation that relates the discrete potential density of the monopoles with the pressure measurements as

\[
y = Gq, \tag{6}\]

where \(G \in \mathbb{C}^{M \times K}\) is the matrix whose elements consist of Green’s function (Eq. (5)), and \(q \in \mathbb{C}^K\) is the potential density of the monopoles. The potential density \(q\) can be simply estimated by solving Eq. (6) using the Moore–Penrose pseudoinverse of \(G\). Then, the continuous \(u(r)\) can be reconstructed by using the estimate \(\hat{q}\).

Several attempts have been made to reconstruct a sound field on the basis of the sparsity assumption [12,13,15,16,23]. In many works, it is assumed that a sound field inside a source-free region can be sparsely represented by plane waves. According to [24], a sound field in a certain star-shaped region can be well approximated by a limited number of plane waves. By using overcomplete plane-wave basis functions, we can represent \(u(r)\) as

\[
\text{Fig. 1 Reconstruction of sound field } u(r,k) \text{ inside region of interest } \Omega, \text{ which does not include any sources. Multiple microphones are placed inside } \Omega. \text{ Fictitious boundary } \partial D \text{ is set in equivalent source method.}
\]
where \( l \in \{1, \ldots, L\} \) is the index of plane waves, \( k_l \) is the wave vector of the \( l \)th plane wave, and \( \gamma_l \) is its amplitude. Here, \( L \gg M \) is assumed. A limited number of nonzero expansion coefficients \( \gamma_l \) will be sufficient to approximate \( u(r) \). By constructing the plane-wave dictionary matrix \( W \in \mathbb{C}^{M \times L} \), we find that the following relation holds:

\[
y = Wp + e,
\]

where \( p \in \mathbb{C}^L \) consists of \( \gamma_l \) and \( e \in \mathbb{C}^M \) is the approximation error. Then, the limited number of elements in \( p \) will have nonzero values. The solution of this underdetermined linear equation with the sparsity of \( p \) promoted can be obtained by using the \( \ell_p \)-(quasi) norm \((0 < p \leq 1)\)

penalty term as

\[
\text{minimize } \frac{1}{2} \| y - Wp \|_2^2 + \lambda \| p \|_p^p,
\]

where \( \| \cdot \|_p \) represents the \( \ell_p \)-norm and \( \lambda \) is a parameter for balancing the approximation error and the sparsity-inducing penalty. The optimization problem (Eq. (9)) corresponds to the maximum a posteriori estimation of \( p \) using the generalized Gaussian distribution as the prior probability distribution with the shape parameter \( p \) as [25].

The smaller the value of \( p \) set, the sparser and heavier-tailed the prior distribution assumed. Therefore, the sparsity of the solution \( p \) can be controlled by \( p \) of the \( \ell_p \)-norm. This sparsity-based representation of the homogeneous sound field has been successfully applied to improve the spatial resolution in the analysis.

### 3. SOUND FIELD RECONSTRUCTION INSIDE REGION INCLUDING SOUND SOURCES

As opposed to the reconstruction in the source-free region, reconstructing a sound field including sound sources is an ill-posed problem since the source distribution can be any function. We consider a region of interest \( \Omega \subseteq \mathbb{R}^3 \) that includes sources as in Fig. 2. The pressure field \( u(r) \) satisfies the inhomogeneous Helmholtz equation as [8]

\[
(\nabla^2 + k^2)u(r, k) = -Q(r, k),
\]

where \( Q(r) \) is the source distribution inside \( \Omega \). The solution \( u(r) \) of Eq. (10) becomes the sum of the particular and homogeneous solutions, \( u_p(r) \) and \( u_{h1}(r) \), respectively,

\[
u(r) = u_p(r) + u_{h1}(r).
\]

When all the sources are included in \( \Omega \), \( u_p(r) \) and \( u_{h1}(r) \) can be regarded as direct-source and reverberant components, respectively. The particular solution \( u_p(r) \) can be obtained by convolution of \( Q(r) \) and \( G(r|\Omega) \) as

\[
u_p(r) = \int \Omega Q(r')G(r'|\Omega)dr'.
\]

The homogeneous solution \( u_{h1}(r) \) is determined by the homogeneous Helmholtz equation with some boundary conditions and the particular solution \( u_p(r) \) on the room boundary. Therefore, \( u_{h1}(r) \) can be represented as a linear combination of plane waves, harmonic functions, or equivalent sources (Green’s functions) as discussed in Sect. 2. Thus, \( u(r) \) is represented as

\[
u(r) = \int \Omega Q(r')G(r'|\Omega)dr' + u_{h1}(r).
\]

The objective is to estimate \( Q(r) \) and \( u(r) \) from the pressure measurements inside \( \Omega \).

To reconstruct \( u(r) \) in Eq. (13), it is necessary to place some constraints on the source distribution \( Q(r) \) to make the problem tractable. We assume that the source distribution is spatially sparse as in [11]. First, the region \( \Omega \) is discretized into a set of small regions \( \Omega_n \) \((n \in \{1, \ldots, N\} \) ).

The representative point of \( \Omega_n \) is defined as a grid point and its location is denoted as \( r_n \). Then, \( u_p(r) \) is approximated as

\[
u_p(r) = \sum_{n=1}^N \int_{r \in \Omega_n} Q(r')G(r'|\Omega)dr'
\]

\[
\approx \sum_{n=1}^N G(r_n) \int_{r \in \Omega_n} Q(r')dr',
\]

where the approximation in the second line can be regarded as the zeroth-order Taylor expansion of \( G(r|\Omega_0) \) around \( r_n \). We assume \( N \gg M \) because the grid points should entirely and densely cover the region \( \Omega \). The dictionary matrix consisting of \( G(r_n|\Omega_0) \) and the vector of the source distribution (direct-source component) inside \( \Omega_n \) are defined as \( D \in \mathbb{C}^{M \times N} \) and \( x \in \mathbb{C}^N \), respectively. Thus, the
discrete form of Eq. (13) is described as the following linear equation:

\[ y = Dx + z, \]  

(15)

where \( z \in \mathbb{C}^M \) is the homogeneous solution \( u_h(r) \) (reverberant component) at the microphone positions. A small number of elements in \( x \) will have nonzero values because of the sparsity assumption of the sources. We here also assume that the reverberant component \( z \) is spatially uncorrelated. Then, similarly to Eq. (9), the direct-source component \( x \) can be obtained by solving the following optimization problem:

\[
\min_x \frac{1}{2} \| y - Dx \|_2^2 + \lambda \| x \|_p^p.
\]  

(16)

Again, \( \| x \|_p \) represents the \( \ell_p \)-norm of \( x \) and \( \lambda \) is the balancing parameter. It will also be possible to incorporate the modeling of the homogeneous sound field, which is discussed in Sect. 2, into Eq. (15) \([12,26]\). Although the above procedure requires the discretization of \( \Omega \) as the set of grid points, a decomposition method without discretization by assuming the source distribution as the sum of delta functions, i.e., point sources, is proposed in \([27,28]\). The source positions and their amplitudes can be estimated by a closed-form solution with multiple spherical microphone arrays near \( \partial \Omega \) \([27]\).

**4. SPARSE SOUND FIELD DECOMPOSITION ALGORITHM USING MULTIDIMENSIONAL MIXED-NORM PENALTY**

As in Sects. 2 and 3, the sound field reconstruction in both the source-free region and the region including sources is achieved by solving a similar sparse decomposition problem, i.e., Eqs. (9) and (16) (sparse sound field decomposition).

Although Eqs. (9) and (16) are formulated for a single time-frequency bin, it is possible to exploit a physical property of the sound field for multiple time-frequency bins, i.e., the group sparse structure of the solution vector. We explain this group sparse model using Eq. (15) with the subscripts of the time frame \( t \in \{1, \ldots, T\} \) and frequency bin \( f \in \{1, \ldots, F\} \). When the sound sources are static for several time frames, each \( x_{t,f} \) will have the same sparsity pattern for \( t \). In addition, since many types of acoustic source signal have a broad frequency band, each \( x_{t,f} \) will also have the same sparsity pattern for \( f \). Figure 3 illustrates the conceptual diagram of this group sparse model. Here, \( Y \in \mathbb{C}^{M \times T \times F} \), \( D \in \mathbb{C}^{M \times N \times F} \), \( X \in \mathbb{C}^{N \times T \times F} \), and \( Z \in \mathbb{C}^{M \times T \times F} \) denote the third-order tensors consisting of \( y_{t,f}, D_{t,f}, x_{t,f} \), and \( z_{t,f} \), respectively. This group sparse decomposition can be achieved by solving the following optimization problem:

\[
\min_X \frac{1}{2} \sum_{t,f} \| y_{t,f} - D_{t,f}x_{t,f} \|_2^2 + \lambda J_{p,q}(X),
\]  

(17)

where \( J_{p,q}(X) \) is the multidimensional mixed-norm penalty term defined as

\[
J_{p,q}(X) = \sum_n \left( \sum_{t,f} |X(n,t,f)|^q \right)^{\frac{p}{q}}.
\]  

(18)

The parameter \( q \) controls the sparsity inside the groups. In general, the group sparse decomposition algorithms are derived for \( q = 2 \) \([29]\); therefore, the \( \ell_2 \)-norm, which will not induce sparsity, is applied to each group. However, acoustic source signals such as speech and music signals also have a sparse structure in the time-frequency domain. Nevertheless, their sparsity strength will be smaller than the spatial sparsity. To incorporate this property, we derived an algorithm for solving Eq. (17) with the sparsity-controlling parameters \( 0 < p \leq q \leq 1 \) \([30]\). Since the optimization problem (Eq. (17)) for \( 0 < p \leq q \leq 1 \) is nonconvex, we derived the algorithm on the basis of the majorization–minimization (MM) method \([31,32]\). First, a surrogate function of Eq. (17) is developed using an auxiliary variable. The monotonic nonincrease in the original objective function is guaranteed by iteratively updating the optimization problem using the surrogate function for the parameter to be solved and the auxiliary variable.

For the penalty term (Eq. (18)), the following inequality holds for \( 0 < p \leq q \leq 1 \):

\[
J_{p,q}(X) = \sum_n \left( \sum_{t,f} |X(n,t,f)|^q \right)^{\frac{p}{q}} \leq \sum_{n,t,f} \eta \eta_{n,t,f} \sum_{t,f} |X(n,t,f)|^q + C
\]  

(19)

where \( C \) is the parameter not related to \( X \), \( \eta_n \) and \( \eta_{n,t,f} \) are parameters defined as

\[
\eta_n = \sum_{t,f} |X(n,t,f)|^q
\]  

(20)
The variable of the surrogate function $\mathcal{Z}$ is simply updated as $X^{(0)} = X^{(i)}$ from the equality condition in Eq. (21). The proposed algorithm is summarized in Algorithm 1, which can be regarded as a type of iteratively reweighted least-squares algorithm [33]. The stopping rule can be set on the basis of the variation of $X$ or the objective function in an iteration, or the maximum number of iterations. This algorithm can be extended to multidimensional sparse decomposition of tensor signals of more than third order [30]. For example, in [11], the group sparse property for a multipole dictionary on the grid points is also introduced, which enables more accurate decomposition when the grid points are coarsely distributed and the sources are away from the grid points.

### 5. APPLICATION TO SUPER-RESOLUTION

The sparse representation enables the spatial resolution limit in the sound field reconstruction to be overcome. We here demonstrate sound field recording and reproduction using sparse decomposition as an example. The sound field recording and reproduction methods are targeted at high-fidelity audio systems, i.e., acoustic virtual reality (VR) systems, which typically involve decomposition of the captured sound field into plane-wave or harmonic functions to obtain driving signals of the loudspeakers to synthesize the captured sound field [5,6]. These methods suffer from artifacts originating from spatial aliasing artifacts; these artifacts depend on the intervals between microphones and loudspeakers. Owing to the significant effect of the spatial aliasing artifacts, listeners will not clearly localize reproduced sound images. Additionally, the frequency characteristics of the reproduced direct sound are greatly affected, which is called the coloration effect.

In [11], the authors applied the sound field model of Eq. (13) to sound field capturing to separate the near-field and far-field sound field components. Since the direct sources in the near field will be sparsely decomposed into Green’s functions, the spatial aliasing artifacts generated owing to the microphone intervals can be significantly reduced compared with the methods based on plane-wave and harmonic decomposition.

We show results of experiments comparing the recording and reproduction performance of the method proposed in [11] and that of the plane-wave-decomposition-based method [6]. The impulse response from a sound source to a microphone array was measured in the room shown in Fig. 4. The reverberation time ($T_{60}$) was 1.14 s. A linear microphone array was set along the x-axis with the center at the origin in the recording area. There were 32

---

**Algorithm 1** Sparse sound field decomposition algorithm using $\ell_{p,q}$-norm penalty.

Initialize $X^{(0)}$, $i = 0$

while loop $\neq 0$ do

$X^{(i)} = X^{(i)}$

$\eta_{n,i,f} = \sum_{t,f} |\mathcal{Z}(n,t,f)|^q$ for $\forall n$

for $t = 1$ to $T$ do

for $f = 1$ to $F$ do

$\eta_{n,i,f}^{(i)} = (\mathcal{Z}^{(i)}(n,t,f))^2$ for $\forall n$

$W_{i,t}^{(i)} = \text{diag} \left( \sqrt{p^{-1}(\eta_{n,i,f}^{(i)})^{1-p/q}(\eta_{n,i,f}^{(i)})^{1-q/2}} \right)$

$A_{i,t}^{(i)} = D_f W_{i,t}^{(i)}$

$x_{i,f}^{(i+1)} = W_{i,t}^{(i)}(A_{i,t}^{(i)})^H(A_{i,t}^{(i)}A_{i,t}^{(i)})^H + \lambda I)^{-1} y_{i,f}^{(i)}$

end for

end for

if stopping condition is satisfied, then

loop $= 0$

end if

end while

$$\eta_{n,i,f} = |\mathcal{Z}(n,t,f)|^2,$$  \(21\)

and $\mathcal{Z}$ is the third-order tensor whose elements consist of the variables of the surrogate function. The equality holds for $X = \mathcal{Z}$. Therefore, the objective function (Eq. (17)) can be upper-bounded by

$$\frac{1}{2} \sum_{i,f} \|y_{i,f} - D_f x_{i,f}\|_2^2 + \lambda \mathcal{J}_{p,q}^r(X|\mathcal{Z}).$$  \(22\)

The objective function (Eq. (22)) is monotonically non-increasing upon alternately minimizing Eq. (22) with respect to $X$ and $\mathcal{Z}$.

By denoting the index of the iteration as $i$, we can derive the update rule for $X$ by solving the following problem for each $t$ and $f$:

$$x_{i,f}^{(i+1)} = \arg \min_{x_{i,f}} \frac{1}{2} \sum_{i,f} \|y_{i,f} - D_f x_{i,f}\|_2^2 + \frac{1}{2} \|Ax_{i,f}\|_2^2 + \frac{1}{2} \|P_{i,f}^{(i)} x_{i,f}\|_2^2,$$  \(23\)

where $P_{i,f}^{(i)} \in \mathbb{R}^{N \times N}$ is a diagonal matrix whose $(n,n')$th element is represented as

$$P_{i,f}^{(i)}_{n,n'} = \begin{cases} p(\eta_{n,i,f}^{(i)})^{q-1}(\eta_{n,i,f}^{(i)})^{q-1} & n = n' \ \text{and} \ \eta_{n,i,f}^{(i)} \neq 0, \\ 0 & \text{otherwise}. \end{cases}$$  \(24\)

Equation (23) can be simply minimized as

$$x_{i,f}^{(i)} = (D_f^H D_f + \lambda P_{i,f}^{(i)})^{-1} D_f^H y_{i,f}.$$  \(25\)

By defining $W_{i,t}^{(i)}$ and $A_{i,t}^{(i)}$ as $P_{i,t}^{(i)} = (W_{i,t}^{(i)})^{-2}$ and $A_{i,t}^{(i)} = D_f W_{i,t}^{(i)}$, respectively, we can obtain the closed-form update rule of $X$ for each $t$ and $f$ as

$$x_{i,f}^{(i)} = W_{i,t}^{(i)}(A_{i,t}^{(i)})^H(A_{i,t}^{(i)}A_{i,t}^{(i)})^H + \lambda I)^{-1} y_{i,f}.$$  \(26\)
microphones set at intervals of 0.06 m. The two-dimensional region \( \Omega \) was set to be a square region of \( 2.4 \times 2.4 \text{ m}^2 \) centered at \((0.0, -2.0, 0.0)\) m on the \(x\)-\(y\) plane at \( z = 0\). The number of grid points was \( 25 \times 13 \) set at intervals of 0.1 m for \(x\) and 0.2 m for \(y\). The primary sound source was an ordinary enclosed loudspeaker at \((-0.5, -1.0, 0.0)\) m. The source signal was a speech signal of a female utterance. The target area was simulated as an anechoic environment. A linear loudspeaker was set along the \(x\)-axis with the center at the origin in the target area. There were 48 loudspeakers set at intervals of 0.04 m. Therefore, the spatial Nyquist frequency of the microphone array was about 2.8 kHz and that of the loudspeaker array was about 4.3 kHz. The sampling frequency was 16 kHz. The frame length and shift length of the short-time Fourier transform (STFT) were 256 and 128 samples, respectively. The Hanning window function was applied for each time frame.

The parameters \( p \) and \( q \) for Algorithm 1 were set to 0.5 and 0.8, respectively. The regularization parameter \( \lambda \) was manually chosen as \( 5.0 \times 10^2 \). We also used the dipole dictionary at each grid point to reduce the off-grid effect [11]. The sparsity-control parameter for the monopole and dipole dictionaries was set to 2.0. The numbers of time frames \( T \) and frequency bins \( F \) were set at 8 and 128, respectively.

The instantaneous reproduced pressure distribution is shown in Fig. 5. Although the ideal pressure distribution cannot be defined, the reproduced pressure for the plane-wave-decomposition-based method was clearly distorted by spatial aliasing artifacts. Note that this distortion did not originate from the effect of reverberation since the initial part, i.e., direct component, of the time-domain synthesized sound field is plotted in Fig. 5. In contrast, the pressure distribution from the single loudspeaker was accurately reproduced by the proposed method.

6. CONCLUSION

An overview of sound field reconstruction and its recent advancement based on sparse representation was presented. The difficulty of the problem differs depending on whether the target region includes sound sources or not. We also demonstrated an application of sparse sound field reconstruction to recording and reproduction for acoustic VR systems. The benefit of the sparsity assumption is basically the improvement of the spatial resolution in the analysis. The algorithm using a multidimensional mixed-norm penalty is useful for achieving such a sparse decomposition for acoustic array signals.

ACKNOWLEDGMENT

This work was supported by JSPS KAKENHI Grant No. JP15H05312 and JST, PRESTO Grant Number JPMJPR18J4, Japan.
REFERENCES

[1] J. D. Maynard, E. G. Williams and Y. Lee, “Nearfield acoustic holography: I. Theory of generalized holography and the development of NAH,” J. Acoust. Soc. Am., 78, 1395–1413 (1985).

[2] Z. Wang and S. F. Wu, “Helmholtz equation-least-squares method for reconstructing the acoustic pressure field,” J. Acoust. Soc. Am., 102, 2020–2032 (1997).

[3] P. A. Nelson and S. H. Yoon, “Estimation of acoustic source strength by inverse methods: Part I, conditioning of the inverse problems,” J. Sound Vib., 233, 639–664 (2000).

[4] H. Teutsch and W. Kellermann, “Acoustic source detection and localization based on wavefield decomposition using circular microphone arrays,” J. Acoust. Soc. Am., 120, 2724–2736 (2006).

[5] M. Poletti, “Three-dimensional surround sound systems based on spherical harmonics,” J. Audio Eng. Soc., 53, 1004–1025 (2005).

[6] S. Koyama, K. Furuya, Y. Hiwasaki and Y. Haneda, “Analytical approach to wave field reconstruction filtering in spatio-temporal frequency domain,” IEEE Trans. Audio Speech Lang. Process., 21, 685–696 (2013).

[7] P. Samarasinghe, T. Abhara-pala and M. Poletti, “Wavefield analysis over large areas using distributed higher order microphones,” IEEE/ACM Trans. Audio Speech Lang. Process., 22, 647–658 (2014).

[8] E. G. Williams, Fourier Acoustics: Sound Radiation and Nearfield Acoustic Holography (Academic Press, London, 1999).

[9] G. H. Koopmann, L. Song and J. B. Fahnline, “A method for computing acoustic fields based on the principle of wave superposition,” J. Acoust. Soc. Am., 86, 2433–2438 (1989).

[10] M. E. Johnson, E. G. Williams, K. H. Baek and J. Garcia-Bontio, “A method for estimating the acoustic field using a gridded sound field superposition,” J. Acoust. Soc. Am., 104, 1221–1231 (1998).

[11] S. Koyama, N. Murata and H. Saruwatari, “Sparse sound field decomposition for super-resolution in recording and reproduction,” J. Acoust. Soc. Am., 143, 3780–3795 (2019).

[12] S. Koyama and L. Daudet, “Sparse representation of a spatial sound field in a reverberant environment,” J. Acoust. Soc. Am., 138, 172–184 (2018).

[13] G. Chardon, L. Daudet, A. Peillot, F. Ollivier, N. Bertin and R. Girbonval, “Near-field acoustic holography using sparsity and compressive sampling principles,” J. Acoust. Soc. Am., 132, 1521–1534 (2012).

[14] A. Asael, H. Bourlard, M. Taghi-zadeh and V. Cevher, “Model-based sparse component analysis for reverberant speech localization,” Proc. ICASSP 2014, pp. 1453–1457, Florence (2014).

[15] R. Mignot, L. Daudet and F. Ollivier, “Room reverberation reconstruction: Interpolation of the early part using compressed sensing,” IEEE Trans. Audio Speech Lang. Process., 21, 2301–2312 (2013).

[16] A. Wabnitz, N. Epain, A. McEwan and C. Jin, “Upscaling ambisonics sound scenes using compressed sensing techniques,” Proc. IEEE Workshop Appl. Signal Process. Audio Acoust. (WASPAA), pp. 1–4, New Paltz (2011).

[17] S. Koyama, S. Shimauchi and H. Ohmuro, “Sparse sound field representation in recording and reproduction for reducing spatial aliasing artifacts,” Proc. ICASSP 2014, pp. 4443–4447, Florence (2014).

[18] D. L. Donoho, “Compressed sensing,” IEEE Trans. Inf. Theory, 52, 1289–1306 (2006).

[19] D. Colton and R. Kress, Inverse Acoustic and Electromagnetic Scattering Theory (Springer, New York, 2013).

[20] A. Laborie, R. Bruno and S. Montoya, “A new comprehensive approach of sound field recording,” Proc. 114th AES Conv., Amsterdam (2003).

[21] N. Ueno, S. Koyama and H. Saruwatari, “Sound field recording using distributed microphones based on harmonic analysis of infinite order,” IEEE Signal Process. Lett., 25, 135–139 (2018).

[22] N. Ueno, S. Koyama and H. Saruwatari, “Kernel ridge regression with constraint of Helmholtz equation for sound field interpolation,” Proc. Int. Workshop Acoust. Signal Enhancement (IWAENC), pp. 436–440, Tokyo (2018).

[23] Y. Takida, S. Koyama and H. Saruwatari, “Exterior and interior sound field separation using convex optimization: Comparison of signal models,” Proc. European Signal Process. Conf. (EUSIPCO), Rome, pp. 2567–2571 (2018).

[24] A. Moiola, R. Hiptmair and I. Perugia, “Plane wave approximation of homogeneous Helmholtz solutions,” Z. Angew. Math. Phys., 62, 809–837 (2011).

[25] B. D. Rao, K. Engan, S. F. Cotter, J. Palmer and K. Kreutz-Delgado, “Subset selection in noise based on diversity measure minimization,” IEEE Trans. Signal Process., 51, 760–770 (2003).

[26] S. Koyama and L. Daudet, “Comparison of reverberation models for sparse sound field decomposition,” Proc. IEEE Workshop Appl. Signal Process. Audio Acoust. (WASPAA), New Paltz, pp. 501–505 (2017).

[27] Y. Takida, S. Koyama, N. Ueno and H. Saruwatari, “Gridless sound field decomposition based on reciprocity gap functional in spherical harmonic domain,” Proc. IEEE Sensor Array Multichannel Signal Process. Workshop (SAM), pp. 627–631, Sheffield (2018).

[28] Y. Takida, S. Koyama, N. Ueno and H. Saruwatari, “Robust gridless sound field decomposition based on structured reciprocity gap functional in spherical harmonic domain,” Proc. ICASSP 2019, pp. 581–585, Brighton (2019).

[29] S. F. Cotter, D. Rao, K. Engan and K. Kreutz-Delgado, “Sparse solutions to linear inverse problems with multiple measurement vectors,” IEEE Trans. Signal Process., 53, 2477–2488 (2005).

[30] N. Murata, S. Koyama, N. Takamune and H. Saruwatari, “Sparse representation using multidimensional mixed-norm penalty with application to sound field decomposition,” IEEE Trans. Signal Process., 66, 3327–3338 (2018).

[31] D. R. Hunter and K. Lange, “A tutorial on MM algorithms,” Am. Stat., 58, 30–37 (2004).

[32] Y. Sun, P. Babu and D. P. Palomar, “Majorization-minimization algorithms in signal processing, communications, and machine learning,” IEEE Trans. Signal Process., 65, 794–816 (2016).

[33] R. Chartrand and W. Yin, “Iteratively reweighted algorithms for compressive sensing,” Proc. ICASSP 2008, pp. 3869–3872, Las Vegas (2008).

Shoichi Koyama received his B.E., M.S., and Ph.D. degrees from the University of Tokyo, Tokyo, Japan, in 2007, 2009, and 2014, respectively. He joined Nippon Telegraph and Telephone Corporation (NTT) in 2009. He is currently a lecturer at the Graduate School of Information Science and Technology, the University of Tokyo. From 2016 to 2018, he was a visiting researcher at Paris Diderot University (Paris7)/Institut Langevin, Paris, France. His research interests include sound field analysis and synthesis, sparse signal representation, and acoustic inverse problems. He is a member of IEEE, IEICE, the Acoustical Society of America, the Audio Engineering Society, and ASJ.