ROBUST GRIDLESS SOUND FIELD DECOMPOSITION BASED ON STRUCTURED RECIPROCITY GAP FUNCTIONAL IN SPHERICAL HARMONIC DOMAIN

Yuhta Takida[†], *Shoichi Koyama*[†,‡], *Natsuki Ueno*[†], *and Hiroshi Saruwatari*[†]

[†]The University of Tokyo, Graduate School of Information Science and Technology,
7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan
[‡]JST, PRESTO, 4-1-8 Honcho, Kawaguchi, Saitama 332-0012, Japan

ABSTRACT

A sound field reconstruction method for a region including sources is proposed. Under the assumption of spatial sparsity of the sources, this reconstruction problem has been solved by using sparse decomposition algorithms with the discretization of the target region. Since this discretization leads to the off-grid problem, we previously proposed a gridless sound field decomposition method based on the reciprocity gap functional in the spherical harmonic domain. Even though this method allows efficient estimation with a closed-form solution while avoiding the off-grid problem, the estimation using a single time-frequency bin can be greatly affected by measurement errors. We formulate an optimization problem using the identical structure of source locations in multiple time-frequency bins and derive an algorithm based on an annihilating filter. Numerical simulation results indicated that robustness against noise can be improved by the proposed method.

Index Terms— reciprocity gap functional, annihilating filter, sound field decomposition, source identification, spherical harmonics

1. INTRODUCTION

Sound field reconstruction aims to estimate a continuous acoustic field inside a target region using measurements of multiple microphones. Such a reconstruction enables the realization of high-fidelity audio systems with arbitrarily selected listening positions by using multiple loudspeakers or a headphone. An efficient strategy in this reconstruction is to represent the target sound field as a sum of element solutions of the Helmholtz equation, such as plane waves [1–3], harmonic functions [1, 4–6], Green's functions [7, 8], and multipoles [9]. We here call this type of representation *sound field decomposition*. A sound field inside a region not including any sources can be reconstructed by decomposing the measurements into these element solutions. On the other hand, when the target region includes sources, this reconstruction problem becomes ill-posed since the source distribution can be any function.

To estimate a sound field inside a region including sources, it is necessary to impose some assumptions on the source distribution. An effective and practical assumption is the spatial sparsity of the sources, as proposed in [8], which also enables the high-resolution analysis of the sound field. The target region is discretized into a set of grid points, and the sound field is decomposed into Green's functions on these grids by using sparse-representation techniques (see Fig. 1(a)). To improve the robustness of the decomposition, the group-sparse structure in the time-frequency domain, i.e., the simultaneous activation of time-frequency bins on the same grids, is also exploited [9, 10]. However, the discretization of the target region causes the off-grid problem, i.e., the deterioration of the decomposition accuracy when the sources are off the grid points.

To overcome the above-mentioned off-grid problem, the authors previously proposed a gridless approach for decomposing the sound field based on the reciprocity gap functional (RGF) in the spherical harmonic domain [11]. The RGF was first proposed in the field of mathematical inverse problems, and has been applied to source localization and scattering problems [12–16]. The method based on the spherical-harmonic-domain RGF (SHD-RGF) enables the sound field to be decomposed into point sources in a gridless manner using multiple spherical microphone arrays or acoustic vector sensors with more flexible geometries (Fig. 1(b)). Even though the estimates of the locations and amplitudes of point sources inside the target region are obtained as a closed-form solution, the algorithm proposed in [11] cannot exploit the identical property of source locations in the time-frequency bins to improve robustness, as used in the group-sparse representation. Furthermore, the estimates in the form of a closed-form solution can be greatly affected by noise because the problem to be solved does not involve the modeling of measurement errors.

Here, we propose a gridless sound field decomposition algorithm that exploits the identical property of source locations in multiple time-frequency bins. The optimization problem based on structured SHD-RGF involving a noise model is formulated using an annihilating filter (AF) [17], and we derive an algorithm to find its valid solution. The use of an AF is known as a method for reconstructing the finite rate of innovation (FRI) signals [18]. Numerical experiments in a three-dimensional (3D) space are conducted to show the effect of the gridless property and the robustness of the proposed method against noise by comparing it with the methods using a single time-frequency bin [11] and group-sparse decomposition [9].

2. PROBLEM STATEMENT

Consider a spherical target region Ω with its boundary $\partial\Omega$ in the free field. The coordinate origin is set at the center of Ω and the pressure of the frequency ω at the position $\mathbf{r} = (r, \theta, \phi)$ in spherical coordinates is denoted as $u(\mathbf{r}, \omega)$. Since Ω is assumed to include sources, $u(\mathbf{r}, \omega)$ satisfies the following inhomogeneous Helmholtz equation:

$$(\nabla^2 + k^2)u(\mathbf{r}, \omega) = -Q(\mathbf{r}, \omega), \tag{1}$$

where k denotes the wave number, which is defined as $k = \omega/c$ with sound speed c, and $Q(\mathbf{r}, \omega)$ represents the source distribution inside Ω. Here, $u(\mathbf{r}, \omega)$ is also assumed to satisfy the Sommerfeld radiation condition [1]. We hereafter omit ω for notational simplicity. By

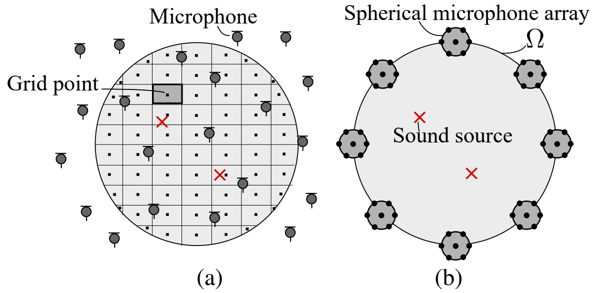


Fig. 1: Sound field models (a) in [9] and (b) in proposed method for decomposition.

assuming that Ω includes J point sources, $Q(\mathbf{r})$ is represented as a linear combination of J delta functions as

$$Q(\mathbf{r}) \approx \sum_{j=1}^{J} c_j \delta(\mathbf{r} - \mathbf{r}_j), \tag{2}$$

where $\mathbf{r}_j$ and c_j represent the jth source location and its amplitude, respectively. The solution of (1), $u(\mathbf{r})$, can be represented as the convolution of $Q(\mathbf{r})$ and the free-field Green's function $G(\mathbf{r})$ [1]. Then, $u(\mathbf{r})$ is approximated by using (2) as

$$u(\mathbf{r}) = \int_{\Omega} Q(\mathbf{r}')G(\mathbf{r}|\mathbf{r}')d\mathbf{r}' \approx \sum_{j=1}^{J} c_j G(\mathbf{r}|\mathbf{r}_j), \tag{3}$$

where $G(\cdot)$ is defined as

$$G(\mathbf{r}|\mathbf{r}') = \frac{e^{ik\|\mathbf{r}-\mathbf{r}'\|_2}}{4\pi\|\mathbf{r}-\mathbf{r}'\|_2}. \tag{4}$$

Therefore, the estimation of the source parameters $\mathbf{r}_j$ and c_j makes it possible to represent the sound field inside Ω by (3), which also enables the reconstruction of $u(\mathbf{r})$ for continuous $\mathbf{r}$. When the sound field inside Ω includes reverberation, i.e., homogeneous solutions to the Helmholtz equation, sound field separation techniques will be necessary as preprocessing [19].

Our measurement model is as follows. Suppose that Q spherical microphone arrays or acoustic vector sensors are arranged on $\partial\Omega$ as shown in Fig. 1(b). By using the qth array at $\mathbf{r}_q$, the spherical harmonic expansion coefficients of the sound field up to the N_mth order with the expansion center $\mathbf{r}_q$ are obtained, which are denoted as $\boldsymbol{\alpha}^{(q)} := [\alpha_{0,0}^{(q)}, \ldots, \alpha_{N_\mathrm{m},N_\mathrm{m}}^{(q)}]^\top$. On the other hand, the pressure field on $\partial\Omega$ can be represented in the spherical harmonic domain with the expansion center at the (global) origin as

$$u(\mathbf{r}) = \sum_{\nu,\mu} u_{\nu,\mu} h_\nu(kr) Y_{\nu,\mu}(\theta,\phi), \tag{5}$$

where $Y_{\nu,\mu}(\theta,\phi)$ is the spherical harmonic function, $h_\nu(\cdot)$ is the νth-order spherical Hankel function of the first kind, and $\sum_{\nu,\mu}$ is the abbreviated form of $\sum_{\nu=0}^{\infty}\sum_{\mu=-\nu}^{\nu}$. By using the truncated global coefficients $\mathbf{u} := [u_{0,0}, \ldots, u_{N,N}]^\top$ with the truncation order N, $\boldsymbol{\alpha}^{(q)}$ is approximately represented as

$$\boldsymbol{\alpha}^{(q)} = \mathbf{S}(\mathbf{r}_q)\mathbf{u}, \tag{6}$$

where $\mathbf{S}(\mathbf{r}_q) \in \mathbb{C}^{(N_\mathrm{m}+1)^2 \times (N+1)^2}$ is the translation matrix [20] used to relate expansion coefficients for different expansion centers [5, 21]. When the vector of the measured harmonic coefficients is defined as $\boldsymbol{\alpha} := [\boldsymbol{\alpha}^{(1)\top}, \ldots, \boldsymbol{\alpha}^{(Q)\top}]^\top \in \mathbb{C}^{Q(N_\mathrm{m}+1)^2}$, (6) can be reformulated as

$$\boldsymbol{\alpha} = \mathbf{T}\mathbf{u}, \tag{7}$$

where the matrix $\mathbf{T} \in \mathbb{C}^{Q(N_\mathrm{m}+1)^2 \times (N+1)^2}$ is defined as $\mathbf{T} := [\mathbf{S}(\mathbf{r}_1)^\top, \ldots, \mathbf{S}(\mathbf{r}_Q)^\top]^\top$. Our goal is to estimate the source parameters $\mathbf{r}_j$ and c_j from the measurements $\boldsymbol{\alpha}$.

3. RGF IN SPHERICAL HARMONIC DOMAIN

We here overview the gridless sound field decomposition method based on SHD-RGF [11]. First, we define the test function $w(\cdot)$ that satisfies the homogeneous Helmholtz equation and the functional $R(\cdot)$ with respect to $w(\cdot)$ as

$$R(w) = \int_{\partial\Omega} \left(u(\mathbf{r}) \frac{\partial w(\mathbf{r})}{\partial \mathbf{n}} - w(\mathbf{r}) \frac{\partial u(\mathbf{r})}{\partial \mathbf{n}} \right) dS. \tag{8}$$

From (2) and (8), the parameters to be estimated, c_j and $\mathbf{r}_j$, are related to the boundary values of the sound field as

$$\sum_{j=1}^{J} c_j w(\mathbf{r}_j) = \int_{\partial\Omega} \left(u(\mathbf{r}) \frac{\partial w(\mathbf{r})}{\partial \mathbf{n}} - w(\mathbf{r}) \frac{\partial u(\mathbf{r})}{\partial \mathbf{n}} \right) dS. \tag{9}$$

Several test functions $w(\cdot)$ have been proposed for solving (9) [12–14]. We use the test function $w_n(\cdot)$ proposed in [12], which is defined in Cartesian coordinates $\mathbf{r} = (x, y, z)$ as

$$w_n(\mathbf{r}) = p^n e^{-ikz}. \tag{10}$$

Here, n is an arbitrary positive integer and $p := x + iy$. The test function $w_n(\mathbf{r})$ can be represented in the spherical harmonic domain as

$$w_n(\mathbf{r}) = \sum_{\nu,\mu} w_{\nu,\mu}^{(n)} j_\nu(kr) Y_{\nu,\mu}(\theta,\phi), \tag{11}$$

where $j_\nu(\cdot)$ is the νth-order spherical Bessel function and the coefficients $w_{\nu,\mu}^{(n)}$ can be analytically obtained as

$$w_{\nu,\mu}^{(n)} = \frac{i^{n-\nu}}{k^n} \sqrt{\frac{4\pi(2\nu+1)(\nu+n)!}{(\nu-n)!}} \delta_{\mu,n}. \tag{12}$$

By calculating the surface integral on the right side of (9) in the spherical harmonic domain, $s_n := R(w_n)$ can be analytically obtained as

$$s_n = \frac{i}{kR^2} \sum_{\nu,\mu} (-1)^{\mu+1} w_{\nu,-\mu}^{(n)} u_{\nu,\mu}, \tag{13}$$

where R is the radius of Ω. From (7), the estimate of $\mathbf{u}$, $\hat{\mathbf{u}}$, can be obtained as

$$\hat{\mathbf{u}} = \mathbf{T}^\dagger \boldsymbol{\alpha}, \tag{14}$$

where $(\cdot)^\dagger$ represents the Moore–Penrose pseudoinverse matrix. The estimate of s_n is obtained by truncating the order ν up to N in (13) and substituting $\hat{u}_{\nu,\mu}$ obtained using (14) and $w_{\nu,\mu}^{(n)}$ given in (12) into (13). The parameters including the source locations p_j can be estimated by the singular value decomposition of Hankel matrices composed of s_n [11, 12].

4. STRUCTURED RGF AND ALGORITHM USING ANNIHILATING FILTER

As discussed in Sect. 3, the method proposed in [11] enables us to estimate source parameters using the closed-form algorithm. However, such a solution does not allow accurate estimation in a noisy situation, where the measurements are modeled with the measurement errors $\boldsymbol{\epsilon}$ as

$$\boldsymbol{\alpha} = \mathbf{T}\mathbf{u} + \boldsymbol{\epsilon}. \tag{15}$$

Moreover, the estimation using $\boldsymbol{\alpha}$ for a single time-frequency bin will be strongly affected by $\boldsymbol{\epsilon}$. To improve robustness against measurement errors, we formulate an optimization problem for (15) by using the identical property of source locations for multiple time-frequency bins. However, it is difficult to solve such a problem by composing the Hankel matrices as in [11]. Therefore, we consider the application of an AF, which is a parameter estimation technique using a filter represented by a polynomial function whose roots correspond to the parameters to be estimated, to the problem formulation and algorithm design. The proposed AF-based decomposition algorithm enables the robust estimation of p_j from s_n using multiple time-frequency bins.

4.1. Construction of AF

First, we construct the AF represented by a polynomial function whose roots correspond to p_j as

$$H(q) = \prod_{j=0}^{J}(1 - p_j q^{-j}) = \sum_{j=1}^{J} h_j q^{-j}. \tag{16}$$

The convolution of the coefficient sequence h_n and the elements s_n for $n \in \mathbb{N}$ is represented as

$$h_{n+J} * s_{n+J} = \sum_{j=0}^{J} h_j s_{n+J-j} \tag{17}$$

$$= \sum_{j'=1}^{J} c_{j'} e^{-ikz_{j'}} p_{j'}^{n+J} \sum_{j=0}^{J} h_j p_{j'}^{-j}. \tag{18}$$

On the basis of the property of the AF, $H(p_j)$ becomes 0 for $j = 1, \ldots, J$. Therefore, the convolution (18) satisfies

$$h_{n+J} * s_{n+J} = 0 \quad (n = 1, \ldots, J). \tag{19}$$

For simplicity, (19) for $n = 1, \ldots, J$ is represented as $\mathbf{h} * \mathbf{s} = \mathbf{0}$. By estimating the filter coefficients $(h_j)_{j=0}^{J}$, the estimate p_j, which includes the source parameters, can be obtained. Since the measurements $\boldsymbol{\alpha}$ can be obtained for multiple time-frequency bins by using a short-time Fourier transform, the common AF can be used for each time-frequency bin. Here, the indices of the frequency bins and time frames are denoted as f ($\in \{1, \ldots, F\}$) and t ($\in \{1, \ldots, T\}$), respectively. When the source locations are assumed to be static for T time frames, the AF satisfies $\mathbf{h} * \mathbf{s}_{f,t} = \mathbf{0}$ for all f and t.

4.2. Algorithm using multiple time-frequency bins

The optimization problem for SHD-RGF using the AF can be formulated as

$$\begin{aligned} \underset{\mathbf{u}_{f,t},\mathbf{h}}{\text{minimize}} \quad & \sum_{f,t} \|\boldsymbol{\alpha}_{f,t} - \mathbf{T}_f \mathbf{u}_{f,t}\|_2^2 \\ \text{such that} \quad & \mathbf{h} * \mathbf{s}_{f,t} = \mathbf{0} \\ & \mathbf{h}^{\mathsf{H}}\mathbf{h} = 1, \end{aligned} \tag{20}$$

Algorithm 1 Proposed AF-based algorithm for SHD-RGF

Input: Local spherical harmonic coefficients $\boldsymbol{\alpha}_{f,t}$, translation matrices $\mathbf{T}_f$, number of sources J, and threshold η.
Output: Global spherical harmonic coefficients $\mathbf{u}_{f,t}$, and AF $\mathbf{h}$.
for $i \leftarrow 1$ **to** I **do**
  Compute $\boldsymbol{\Lambda}(\mathbf{h})$ with $\mathbf{h} = \mathbf{h}^{(i-1)}$ and update $\mathbf{h}^{(i)}$ by solving (24)
  Update $\mathbf{u}_{f,t}^{(i)}$ with the updated AF $\mathbf{h} = \mathbf{h}^{(i)}$ by (23)
  if $\sum_{f,t} \|\boldsymbol{\alpha}_{f,t} - \mathbf{T}_f \mathbf{u}_{f,t}^{(i)}\|_2^2 \leq \eta$ **then**
    Terminate loop
  end if
end for
$\mathbf{u}_{f,t} \leftarrow \mathbf{u}_{f,t}^{(i)}$, $\mathbf{h} \leftarrow \mathbf{h}^{(i)}$.

where the constraint $\mathbf{h}^{\mathsf{H}}\mathbf{h} = 1$ is imposed to avoid the trivial solution $\mathbf{h} = \mathbf{0}$. To represent the convolution $\mathbf{h} * \mathbf{s}_{f,t}$ as a linear form with respect to $\mathbf{u}_{f,t}$ and $\mathbf{h}$, the matrices $\mathbf{W}_f(\mathbf{h}) \in \mathbb{C}^{J \times (N+1)^2}$ and $\mathbf{V}_f(\mathbf{u}_{f,t}) \in \mathbb{C}^{J \times (J+1)}$ are defined so that they satisfy

$$\mathbf{h} * \mathbf{s}_{f,t} = \mathbf{W}_f(\mathbf{h})\mathbf{u}_{f,t} = \mathbf{V}_f(\mathbf{u}_{f,t})\mathbf{h}. \tag{21}$$

By using this linear representation, an equivalent optimization problem to (20) can be formulated by substituting the constraint condition into the objective function [22, 23]. When $\mathbf{h}$ is fixed, the augmented Lagrangian function $\mathcal{L}(\mathbf{u}_{f,t}, \boldsymbol{\theta}_{f,t})$ is defined as

$$\begin{aligned} \mathcal{L}(\mathbf{u}_{f,t}, \boldsymbol{\theta}_{f,t}) := \frac{1}{2} \sum_{f,t} \|\boldsymbol{\alpha}_{f,t} - \mathbf{T}_f \mathbf{u}_{f,t}\|_2^2 \\ + \sum_{f,t} \boldsymbol{\theta}_{f,t}^{\mathsf{H}} \mathbf{W}_f(\mathbf{h})\mathbf{u}_{f,t}, \end{aligned} \tag{22}$$

where $\boldsymbol{\theta}_{f,t}$ is the Lagrangian multiplier. From the stationary condition of (22), the solution $\mathbf{u}_{f,t}$ can be obtained as

$$\begin{aligned} \mathbf{u}_{f,t}(\mathbf{h}) = \mathbf{v}_{f,t} - (\mathbf{T}_f^{\mathsf{H}}\mathbf{T}_f)^{-1} \\ \cdot \mathbf{W}_f(\mathbf{h})^{\mathsf{H}} \boldsymbol{\Sigma}(\mathbf{h})^{-1} \mathbf{W}_f(\mathbf{h}) \mathbf{v}_{f,t}, \end{aligned} \tag{23}$$

where $\mathbf{v}_{f,t} := (\mathbf{T}_f^{\mathsf{H}}\mathbf{T}_f)^{-1}\mathbf{T}_f^{\mathsf{H}}\boldsymbol{\alpha}_{f,t}$. By substituting (23) into the objective function (20), the following optimization problem, which is equivalent to (20), can be derived:

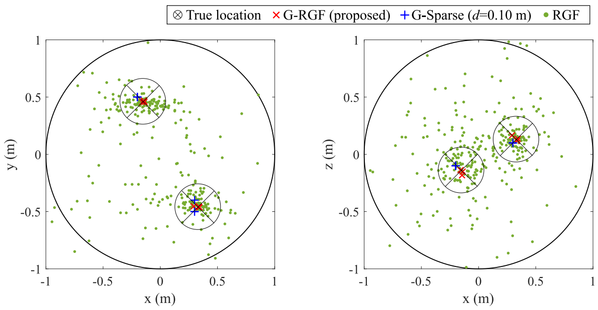
$$\begin{aligned} \underset{\mathbf{h}}{\text{minimize}} \quad & \mathbf{h}^{\mathsf{H}}\boldsymbol{\Lambda}(\mathbf{h})\mathbf{h} \\ \text{such that} \quad & \mathbf{h}^{\mathsf{H}}\mathbf{h} = 1, \end{aligned} \tag{24}$$

where the matrix $\boldsymbol{\Lambda}(\mathbf{h}) \in \mathbb{C}^{(J+1)\times(J+1)}$ is defined as

$$\boldsymbol{\Lambda}(\mathbf{h}) := \sum_{f,t} \mathbf{V}_f(\mathbf{v}_{f,t})^{\mathsf{H}} \boldsymbol{\Sigma}(\mathbf{h})^{-1} \mathbf{V}_f(\mathbf{v}_{f,t}) \tag{25}$$

$$\boldsymbol{\Sigma}(\mathbf{h}) := \mathbf{W}_f(\mathbf{h})(\mathbf{T}_f^{\mathsf{H}}\mathbf{T}_f)^{-1}\mathbf{W}_f(\mathbf{h})^{\mathsf{H}}. \tag{26}$$

Since it is difficult to solve (24) directly, we apply an iterative algorithm to find a valid solution [22]. Starting with an initial value $\mathbf{h}^{(0)}$, $\mathbf{h}$ is updated by solving (24). Then, $\boldsymbol{\Lambda}(\mathbf{h})$ is computed with the reconstructed $\mathbf{h}$ from the previous iteration. At each iteration, $\mathbf{u}_{f,t}$ is updated by (23), which is used for the stopping rule of the iterations. The proposed algorithm is summarized in Algorithm 1.



(a) (b)

Fig. 2: Scatter plots of true and estimated locations. Marks "⊗" in black, "×" in red, "+" in blue, and "·" in green represent the true locations and the locations estimated by **G-RGF**, **G-Sparse**, and **RGF**, respectively. (a) x-y-plane and (b) x-z-plane.

5. EXPERIMENTS

Numerical simulations were conducted to evaluate the proposed method under the free-field condition in 3D space. We compared the proposed method using the structured (group) SHD-RGF (**G-RGF**) with the method based on SHD-RGF using the Hankel matrix (**RGF**) [11] and the method based on group-sparse decomposition (**G-Sparse**) [9]. The sound field decomposition performance of each method was evaluated in terms of the source localization accuracy.

The radius of the spherical target region Ω was 1.0 m. Spherical microphone arrays (or acoustic vector sensors) that can measure the spherical harmonic coefficients up to the second order were used. The arrays were uniformly arranged on four horizontal rings on the surface $\partial\Omega$. Nine arrays were set on each circular ring and the rings were set at zenith angles of 40°, 75°, 105°, and 140°. For **G-Sparse**, the grid points were set inside Ω. Three different intervals between the grid points d were investigated, $d = 0.10$, 0.15, and 0.20 m. Orthogonal matching pursuit (OMP) [24] was applied as the sparse decomposition algorithm. Two point sources were set at $(0.33, -0.46, 0.13)$ m and $(-0.15, 0.46, -0.14)$ m. Multiple sine waves with amplitudes generated by a complex Gaussian distribution were used as source signals. Three frequency bands (groups) of the source signals were assumed to be obtained: 200–400 Hz, 400–600 Hz, and 600–800 Hz. The number of time frames, T, was 10. Gaussian noise was also added so that the signal-to-noise ratio (SNR) was 20 dB. In **RGF**, the signal of the single frequency bin at each time frame was used for the estimation. Each group of time-frequency bins was used in **G-RGF** and **G-Sparse**.

For evaluation, we define the root-mean-square error (RMSE) of the estimated source locations as

$$\mathrm{RMSE} = \sqrt{\frac{1}{J}\sum_{j=1}^{J} \|\mathbf{r}_{j,\mathrm{true}} - \hat{\mathbf{r}}_j\|_2^2}, \tag{27}$$

where $\mathbf{r}_{j,\mathrm{true}}$ and $\hat{\mathbf{r}}_j$ are the true and estimated locations of the jth source, respectively.

Fig. 2 shows the scatter plot of the estimated locations for each method, where the results for $d = 0.1$ m are only plotted for **G-Sparse**. For **RGF**, the estimated locations widely varied owing to the measurement noise. On the other hand, **G-RGF** and **G-Sparse** achieved relatively accurate localization with small variances owing to the use of multiple time-frequency bins. Note that the estimated locations for **G-Sparse** cannot be closer to the true locations than the grid points, whereas this off-grid problem was avoided for **G-RGF**. The RMSE of each method is also shown in Fig. 3, where

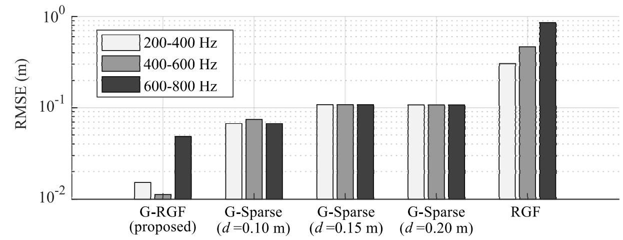


Fig. 3: RMSE for each group of frequency bins.

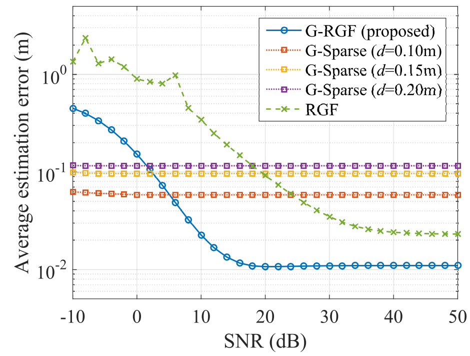


Fig. 4: Relationship between SNR and average estimation error.

the results for each group of frequency bins are separately plotted. For **RGF**, the RMSEs were averaged over the time-frequency bins of each group. The RMSE of **G-RGF** was smaller than those of **G-Sparse** and **RGF** in all cases.

Fig. 4 shows the relationship between the SNR and the expected estimation error, i.e., $\mathbb{E}[\|\hat{\mathbf{r}} - \mathbf{r}_{\mathrm{true}}\|_2]$, when a single point source was located at $(0.33, -0.46, 0.13)$ m. The computation of $\mathbb{E}[\|\hat{\mathbf{r}} - \mathbf{r}_{\mathrm{true}}\|_2]$ was performed for 200 noise patterns. The group of frequency bins of 400–600 Hz was used for **G-RGF** and **G-Sparse**. The single frequency bin at 400 Hz was used for **RGF**. The highest estimation accuracy was achieved by the proposed **G-RGF** above 6 dB. Although the estimation error of **G-Sparse** was almost constant owing to the off-grid effect, **G-Sparse** exhibited better performance than **G-RGF** at low SNRs.

6. CONCLUSION

We proposed a gridless sound field decomposition method based on the structured RGF in the spherical harmonic domain (SHD-RGF). Since it is difficult to solve the optimization problem including a noise model at multiple time-frequency bins with the previous Hankel-matrix-based schemes, we derived an iterative decomposition algorithm using an AF. The proposed method makes it possible to improve robustness against noise. Furthermore, the gridless property allows the accurate estimation of source parameters, especially at high SNRs.

7. ACKNOWLEDGEMENTS

This work was supported by JSPS KAKENHI Grant Number JP15H05312, JST, PRESTO Grant Number JPMJPR18J4, and SECOM Science and Technology Foundation.

8. REFERENCES

[1] E. G. Williams, *Fourier Acoustics: Sound Radiation and Nearfield Acoustical Holography*, Academic Press, London, 1999.

[2] G. Chardon, L. Daudet, A. Peillot, F. Ollivier, N. Bertin, and R. Gribonval, "Near-field acoustic holography using sparsity and compressive sampling principles," *J. Acoust. Soc. Am.*, vol. 132, no. 3, pp. 1521–1534, 2012.

[3] 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.*, vol. 21, no. 4, pp. 685–696, 2013.

[4] M. A. Poletti, "Three-dimensional surround sound systems based on spherical harmonics," *J. Audio Eng. Soc.*, vol. 53, no. 11, pp. 1004–1025, 2005.

[5] P. N. Samarasinghe, T. D. Abhayapala, and M. A. Poletti, "Wavefield analysis over large areas using distributed higher order microphones," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 22, no. 3, pp. 647–658, 2014.

[6] S. Koyama, K. Furuya, K. Wakayama, S. Shimauchi, and H. Saruwatari, "Analytical approach to transforming filter design for sound field recording and reproduction using circular arrays with a spherical baffle," *J. Acoust. Soc. Am.*, vol. 139, no. 3, pp. 1024–1036, 2016.

[7] 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.*, vol. 86, no. 5, pp. 2433–2438, 1998.

[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. (ICASSP)*, Florence, May 2014, pp. 4443–4447.

[9] S. Koyama, N. Murata, and H. Saruwatari, "Sparse sound field decomposition for super-resolution in recording and reproduction," *J. Acoust. Soc. Am.*, vol. 143, no. 6, pp. 3780–3895, 2018.

[10] 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.*, vol. 66, no. 12, pp. 3327–3338, 2018.

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

[12] A. El Badia and T. Nara, "An inverse source problem for Helmholtz's equation from the Cauchy data with a single wave number," *Inverse Prob.*, vol. 27, no. 105001, 2011.

[13] Z. Dogan, V. Tsiminaki, I. Jovanovic, T. Blu, and D. Van De Ville, "Localization of point sources for systems governed by the wave equation," in *Proc. SPIE, Wavelets and Sparsity XIV*, San Diego, Aug. 2011, vol. 8138.

[14] Z. Dogan, I. Jovanovic, T. Blu, and D. Van De Ville, "3D reconstruction of wave-propagated point sources from boundary measurements using joint sparsity and finite rate of innovation," in *Proc. IEEE Int. Symp. Biomed. Imaging (ISBI)*, Barcelona, May 2012, pp. 1575–1578.

[15] C. Alves, R. Kress, and P. Serranho, "Iterative and range test methods for an inverse source problem for acoustic waves," *Inverse Prob.*, vol. 25, no. 055005, 2009.

[16] D. Colton and H. Haddar, "An application of the reciprocity gap functional to inverse scattering theory," *Inverse Prob.*, vol. 21, no. 1, pp. 383–398, 2005.

[17] T. Blu, P. Dragotti, M. Vetterli, P. Marziliano, and L. Coulot, "Sparse sampling of signal innovations," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 31–40, 2008.

[18] M. Vetterli, P. Marziliano, and T. Blu, "Sampling signals with finite rate of innovation," *IEEE Trans. Signal Process.*, vol. 50, no. 6, pp. 1417–1428, 2002.

[19] Y. Takida, S. Koyama, and H. Saruwatari, "Exterior and interior sound field separation using convex optimization: comparison of signal models," in *Proc. European Signal Process. Conf. (EUSIPCO)*, Rome, Sep. 2018, pp. 2567–2571.

[20] P. A. Martin, *Multiple Scattering: Interaction of Time-Harmonic Waves with N Obstacles*, Cambridge University Press, New York, 2006.

[21] P. N. Samarasinghe, T. D. Abhayapala, and M. A. Poletti, "3D spatial soundfield recording over large regions," in *Proc. Int. Workshop Acoust. Signal Enhancement (IWAENC)*, Aachen, Sep. 2012, pp. 1–4.

[22] H. Pan, T. Blu, and M. Vetterli, "Towards generalized FRI sampling with an application to source resolution in radioastronomy," *IEEE Trans. Signal Process.*, vol. 65, no. 4, pp. 821–835, 2017.

[23] H. Pan, R. Scheibler, E. Bezzam, I. Dokmanic, and M. Vetterli, "FRIDA: FRI-based DOA estimation for arbitrary array layouts," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, New Orleans, Mar. 2017, pp. 3186–3190.

[24] Y. Pati, R. Rezaiifar, and P. S. Krishnaprasad, "Orthogonal matching pursuit: recursive function approximation with application to wavelet decomposition," in *Proc. 27th Annual Asilomar Conf. Signals, Systems, Computers*, Pacific Grove, Nov. 1993, pp. 40–44.