Sound Field Estimation Method Robust to Microphone Position and Directivity Errors

Takumi Koga* and Natsuki Ueno*
* Kumamoto University, Kumamoto, Japan
E-mail: d5900@st.cs.kumamoto-u.ac.jp

***Abstract*—We propose a sound field estimation method that accounts for errors in microphone positions and directivities. First, the position and directivity of each microphone are incorporated into the modeling of an observation operator, and the associated errors are represented as perturbations in this operator. To improve robustness against these errors, we formulate the estimation as a joint optimization problem of the sound field and the perturbations in the observation operator. To solve this problem, we propose an iterative optimization algorithm that alternately estimates the sound field and the observation operator, deriving closed-form solutions for each iteration. Numerical experiments demonstrated that the proposed method achieved higher estimation accuracy in certain frequency bands than our previous method, which does not account for these errors.**

I. Introduction

The purpose of sound field estimation is to estimate the spatio-temporal distribution of sound pressure within a target region from signals obtained by multiple microphones. This technique has a wide range of fields, such as the sound field reproduction using loudspeakers or headphones [1]–[4] and the spatial active noise control [5]–[7].

A commonly used approach for sound field estimation is based on the Kirchhoff–Helmholtz integral and its related formulations [1], [8]–[10]. However, the microphone arrays employed in these methods must match the boundary of the target region and are typically restricted to simple geometries, such as spherical, planar, or cylindrical configurations. Consequently, designing such arrays is often impractical in real-world scenarios, particularly when the target region is large.

In contrast, methods based on discrete measurements [2], [10]–[13] impose no constraints on the positions or directivities of microphones. Laborie et al. proposed a sound field estimation method where the sound field is represented using a finite-dimensional expansion of the spherical harmonics and the corresponding coefficients are obtained by formulating the inverse problem [11]. Ueno et al. extended this approach to infinite dimensions using the theory of Hilbert spaces [12], [13]. These methods are established under the assumption that the positions and directional responses of the microphones are precisely given in advance. However, in practical situations, discrepancies between the assumed and actual microphone positions and directivity often arise due to several factors such as installation inaccuracies or modeling errors in directivity.

In this work, we propose a discrete-measurement-based sound field estimation method that is robust to errors in microphone positions and directivities. The proposed approach extends our prior works [12], [13] by explicitly modeling these errors. In the proposed method, the positions and directivities of microphones are incorporated into the formulation of observation operators, as in [10], [13]. Thus, errors in the microphone position and directivity are modeled as perturbations in the observation operators. We formulate a joint estimation problem of the sound field and the perturbations in the observation operators as a regularized least squares problem. To solve this problem, we propose an iterative algorithm and derive the update rule in closed form for each iteration. Numerical experiments demonstrated the effectiveness of the proposed method, showing that it achieved higher estimation accuracy than our previous method in the presence of errors in positions, orientations, and directivities of the microphones.

II. Previous Formulation

In this section, we briefly describe our previous method [10], [13], which does not take into account errors in microphone positions and directivities.

A. Representation of Sound Field

When there is no sound source within a simply connected region $\Omega \subset \mathbb{R}^3$, the sound pressure $u(\boldsymbol{r}, \omega)$ at position $\boldsymbol{r} \in \Omega$ and angular frequency $\omega \in \mathbb{R}$ satisfies the following homogeneous Helmholtz equation:

$$(\Delta + k^2)u(\boldsymbol{r}, \omega) = 0, \tag{1}$$

where Δ denotes the Laplace operator, $k \coloneqq \omega/c$ is the wave number, and c is the speed of sound. Hereafter, we omit ω and write $u(\boldsymbol{r})$ for notational simplicity. A typical representation of a sound field satisfying (1) is a superposition of plane-wave functions (also called the Herglotz wave function [14]) as follows:

$$u(\boldsymbol{r}) = \int_{\boldsymbol{x} \in \mathbb{S}_2} w(\boldsymbol{x}) \exp(-ik\boldsymbol{x} \cdot \boldsymbol{r}) dS. \tag{2}$$

Here, $\mathbb{S}_2$ denotes the unit sphere in $\mathbb{R}^3$, and $w \colon \mathbb{S}_2 \to \mathbb{C}$ represents the complex amplitude of plane waves arriving from each direction. Then, the sound field is modeled directly as an element of the infinite-dimensional function space $\mathcal{H}$ defined as

$$\mathcal{H} \coloneqq \left\{ u(\boldsymbol{r}) = \int_{\boldsymbol{x} \in \mathbb{S}_2} w(\boldsymbol{x}) \exp(-ik\boldsymbol{x} \cdot \boldsymbol{r}) dS \mid w \in L^2(\mathbb{S}_2) \right\}. \tag{3}$$

979-8-3315-7206-8/25/$31.00

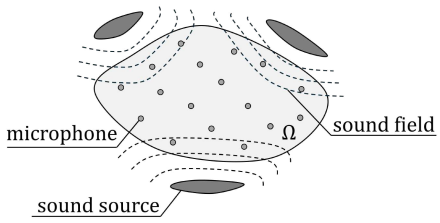


Fig. 1: The target sound field

Here, $L^2(\mathbb{S}_2)$ denotes the set of square-integrable functions from $\mathbb{S}_2$ to $\mathbb{C}$. As it will be essential for the subsequent sections, we define an inner product $\langle \cdot, \cdot \rangle_{\mathcal{H}} \colon \mathcal{H} \times \mathcal{H} \to \mathbb{C}$ and the norm $\| \cdot \|_{\mathcal{H}} \colon \mathcal{H} \to \mathbb{R}$ as follows:

$$\langle u_1, u_2 \rangle_{\mathcal{H}} \coloneqq 4\pi \int_{\boldsymbol{x} \in \mathbb{S}_2} w_1(\boldsymbol{x})^* w_2(\boldsymbol{x}) dS \quad (u_1, u_2 \in \mathcal{H}), \tag{4}$$

$$\|u\|_{\mathcal{H}} \coloneqq \sqrt{\langle u, u \rangle_{\mathcal{H}}} \quad (u \in \mathcal{H}). \tag{5}$$

Here, $(\cdot)^*$ denotes the complex conjugate, and $w_1, w_2 \in L^2(\mathbb{S}_2)$ are the amplitude functions associated with u_1 and u_2, respectively, in the expansion given by (2). Under this definition, the function space $(\mathcal{H}, \langle \cdot, \cdot \rangle_{\mathcal{H}})$ forms a Hilbert space [15].

B. Modeling Observed Signals and Problem Formulation

As shown in Fig. 1, we assume that M microphone are placed within Ω. Then, the observed signal of the mth microphone is modeled as follows:

$$s_m = \int_{\boldsymbol{x} \in \mathbb{S}_2} \gamma_m(\boldsymbol{x}) w(\boldsymbol{x}) \exp(-ik\boldsymbol{x} \cdot \boldsymbol{r}_m) dS + \epsilon_m. \tag{6}$$

Here, $\epsilon_m \in \mathbb{C}$ denotes the observation noise associated with the mth microphone, and $\gamma_m(\cdot) \in L^2(\mathbb{S}_2)$ denotes the directivity function, representing its sensitivity to incoming plane waves from each direction. For many commonly used microphones, such as omnidirectional and bidirectional types, $\gamma_m(\cdot) \in L^2(\mathbb{S}_2)$ can be well represented in terms of finite-order spherical harmonics, yielding the representation:

$$\gamma_m(\boldsymbol{x})^* = \sum_{\nu,\mu}^{N} c_{m,\nu,\mu} \hat{Y}_{\nu,\mu}(\boldsymbol{x}) \quad (\boldsymbol{x} \in \mathbb{S}_2), \tag{7}$$

where $c_{m,\nu,\mu} \in \mathbb{C}$ is the spherical harmonic coefficient of the mth microphone, N is the maximum order, $\sum_{\nu,\mu}^{N}$ is the abbreviation of $\sum_{\nu=0}^{N} \sum_{\mu=-\nu}^{\nu}$, and $\hat{Y}_{\nu,\mu}(\cdot) \colon \mathbb{S}_2 \to \mathbb{C}$ is the unnormalized spherical harmonic function [10] of order ν and degree μ, defined using the normalized spherical harmonic function $Y_{\nu,\mu}(\cdot) \colon \mathbb{S}_2 \to \mathbb{C}$ [16] as

$$\hat{Y}_{\nu,\mu}(\boldsymbol{x}) \coloneqq \sqrt{4\pi} Y_{\nu,\mu}(\boldsymbol{x}) \quad (\boldsymbol{x} \in \mathbb{S}_2). \tag{8}$$

For simplicity, (6) is hereafter rewritten as

$$s_m = \tilde{\mathcal{F}}_m u + \epsilon_m. \tag{9}$$

Here, $\tilde{\mathcal{F}}_m \colon \mathcal{H} \to \mathbb{C}$ is defined as

$$\tilde{\mathcal{F}}_m u \coloneqq \int_{\boldsymbol{x} \in \mathbb{S}_2} \gamma_m(\boldsymbol{x}) w(\boldsymbol{x}) \exp(-ik\boldsymbol{x} \cdot \boldsymbol{r}_m) dS \quad (u \in \mathcal{H}). \tag{10}$$

That is, $\tilde{\mathcal{F}}_m$ denotes the observation operator that maps the sound field u to the observed signal s_m. Since $\gamma_m(\cdot) \in L^2(\mathbb{S}_2)$, $\tilde{\mathcal{F}}_m$ is a bounded linear functional on $\mathcal{H}$.

On the basis of the above definitions, the sound field estimation problem is formulated as a regularized least-squares problem:

$$\underset{u \in \mathcal{H}}{\text{minimize}} \quad \sum_{m=1}^{M} |\tilde{\mathcal{F}}_m u - s_m|^2 + \lambda \|u\|_{\mathcal{H}}^2, \tag{11}$$

where $\lambda \in (0, \infty)$ is a regularization parameter. It has been shown that this problem admits a closed-form solution [10], [12], [13].

III. Proposed Method

In Section II, the sound field estimation problem is formulated under the assumption that the microphone positions and directivities are given precisely in advance. However, such ideal conditions are rarely achieved in practice, as various factors can introduce errors in both position and directivity. For example, the microphone positions may deviate during installation, and their directivities may differ from the nominal values due to individual variations. To address these issues, this section proposes a sound field estimation method that explicitly accounts for errors in microphone positions and directivities.

A. Formulation

In this and the following sections, we refer to the ideal values of microphone positions and directivities given in advance as the *nominal* values, and to those including errors as the *actual* values.

Let $\mathcal{H}^*$ denote the set of all bounded linear functionals from $\mathcal{H}$ to $\mathbb{C}$, that is, the continuous dual space of $\mathcal{H}$. Then, $\tilde{\mathcal{F}}_1, \ldots, \tilde{\mathcal{F}}_M \in \mathcal{H}^*$, as defined in (10). The observation operators $\tilde{\mathcal{F}}_1, \ldots, \tilde{\mathcal{F}}_M$ are given on the basis of the nominal positions and directivities of the microphones. On the other hand, the actual observation operators $\mathcal{F}_1, \ldots, \mathcal{F}_M$ include errors and are modeled as

$$\mathcal{F}_m = \tilde{\mathcal{F}}_m + \mathcal{E}_m \quad (m = 1, \ldots, M), \tag{12}$$

where $\mathcal{E}_m \in \mathcal{H}^*$ represents the perturbation of the observation operator of the mth microphone caused by errors in its position and directivity. To account for these errors, we formulate the problem of the joint estimation of the sound field and the observation operators as the following optimization problem:

$$\begin{aligned} \underset{u \in \mathcal{H}, \mathcal{F}_1, \ldots, \mathcal{F}_M \in \mathcal{H}^*}{\text{minimize}} \quad & \sum_{m=1}^{M} |\mathcal{F}_m u - s_m|^2 \\ & + \lambda_1 \|u\|_{\mathcal{H}}^2 + \lambda_2 \sum_{m=1}^{M} \|\mathcal{F}_m - \tilde{\mathcal{F}}_m\|_{\mathcal{H}^*}^2, \end{aligned} \tag{13}$$

where $\lambda_1, \lambda_2 \in (0, \infty)$ are regularization parameters, and $\|\cdot\|_{\mathcal{H}^*}$ denotes the operator norm on $\mathcal{H}^*$. The third term penalizes the deviations of the actual observation operators from the nominal ones.

Since $\mathcal{H}$ is a Hilbert space, the Riesz representation theorem [17] can be applied. Thus, for each linear functional $\mathcal{F}_m, \tilde{\mathcal{F}}_m, \mathcal{E}_m \in \mathcal{H}^*$, there exist unique functions $v_m, \tilde{v}_m, \zeta_m \in \mathcal{H}$ satisfying

$$\mathcal{F}_m u = \langle v_m, u \rangle_{\mathcal{H}}, \tag{14}$$

$$\tilde{\mathcal{F}}_m u = \langle \tilde{v}_m, u \rangle_{\mathcal{H}}, \tag{15}$$

$$\mathcal{E}_m u = \langle \zeta_m, u \rangle_{\mathcal{H}} \tag{16}$$

for all $u \in \mathcal{H}$. From (12), we obtain

$$v_m = \tilde{v}_m + \zeta_m. \tag{17}$$

Furthermore, using (4) and (10), we have

$$\tilde{v}_m(\boldsymbol{r}) = \frac{1}{4\pi} \int_{\boldsymbol{x} \in \mathbb{S}_2} \gamma_m(\boldsymbol{x})^* \exp(-ik\boldsymbol{x} \cdot (\boldsymbol{r} - \boldsymbol{r}_m)) dS. \tag{18}$$

Then, by substituting (7) into (18) and applying the Funk–Hecke formula [16], we obtain

$$\tilde{v}_m(\boldsymbol{r}) = \sum_{\nu,\mu}^{N} c_{m,\nu,\mu} \varphi_{\nu,\mu}(\boldsymbol{r} - \boldsymbol{r}_m), \tag{19}$$

where $\varphi(\cdot) \colon \mathbb{R}^3 \to \mathbb{C}$ is the regular spherical wave function [10] defined as

$$\varphi_{\nu,\mu}(\boldsymbol{r}) \coloneqq \frac{1}{i^\nu} j_\nu(k\|\boldsymbol{r}\|) \hat{Y}_{\nu,\mu}\left(\frac{\boldsymbol{r}}{\|\boldsymbol{r}\|}\right). \tag{20}$$

Here, $j_\nu(\cdot) \coloneqq \mathbb{R} \to \mathbb{R}$ is the νth order spherical Bessel function of the first kind, and i is the imaginary unit. Therefore, the problem (13) is reduced to the following optimization problem over $\mathcal{H}$:

$$\begin{aligned} \underset{u,\zeta_1,\ldots,\zeta_M \in \mathcal{H}}{\text{minimize}} \quad & \sum_{m=1}^{M} |\langle \tilde{v}_m + \zeta_m, u \rangle_{\mathcal{H}} - s_m|^2 \\ & + \lambda_1 \|u\|_{\mathcal{H}}^2 + \sum_{m=1}^{M} \lambda_2 \|\zeta_m\|_{\mathcal{H}}^2. \end{aligned} \tag{21}$$

This reformulation shows that the estimation problem (13) can be fully expressed within the Hilbert space $\mathcal{H}$ by applying the Riesz representation theorem [17]. In particular, the estimation of the observation operators $\mathcal{F}_m \in \mathcal{H}^*$ is reduced to the estimation of $\zeta_m \in \mathcal{H}$.

B. Optimization Algorithm

The optimization problem (21) is difficult to solve directly for a global optimum. Therefore, we propose an iterative optimization algorithm in which the objective function of (21) is alternately minimized with respect to u and $\zeta_1, \ldots, \zeta_M$.

First, we divide the objective function of (21) into sub-objective functions: Q_1, which contains only the terms dependent on u, and $Q_{2,1}, \ldots, Q_{2,M}$, which contain only the terms dependent on $\zeta_1, \ldots, \zeta_M$, respectively. They are defined as follows:

$$Q_1(u; \hat{\zeta}_1, \ldots, \hat{\zeta}_M) \coloneqq \sum_{m=1}^{M} |\langle \tilde{v}_m + \hat{\zeta}_m, u \rangle_{\mathcal{H}} - s_m|^2 + \lambda_1 \|u\|_{\mathcal{H}}^2, \tag{22}$$

$$Q_{2,m}(\zeta_m; \hat{u}) \coloneqq |\langle \tilde{v}_m + \zeta_m, \hat{u} \rangle_{\mathcal{H}} - s_m|^2 + \lambda_2 \|\zeta_m\|_{\mathcal{H}}^2. \tag{23}$$

By alternately minimizing Q_1 and $Q_{2,1}, \ldots, Q_{2,M}$, the update rules are obtained as

$$\hat{u}^{(n)} = \underset{u \in \mathcal{H}}{\text{argmin}} \quad Q_1(u; \hat{\zeta}_1^{(n-1)}, \ldots, \hat{\zeta}_M^{(n-1)}), \tag{24}$$

$$\hat{\zeta}_m^{(n)} = \underset{\zeta_m \in \mathcal{H}}{\text{argmin}} \quad Q_{2,m}(\zeta_m; \hat{u}^{(n)}). \tag{25}$$

Under the assumption that the initial functions $\zeta_1^{(0)}, \ldots, \zeta_M^{(0)}$ in (24) are all zero, it can be shown inductively, by applying the representer theorem [18], [19] to (24) and (25), that the optimal solutions can be expressed in the following finite-dimensional forms:

$$\hat{u}^{(n)}(\boldsymbol{r}) = \sum_{m=1}^{M} \hat{\alpha}_m^{(n)} \tilde{v}_m(\boldsymbol{r}), \tag{26}$$

$$\hat{\zeta}_m^{(n)}(\boldsymbol{r}) = \sum_{m'=1}^{M} \hat{\beta}_{m,m'}^{(n)} \tilde{v}_{m'}(\boldsymbol{r}). \tag{27}$$

Here, $\hat{\boldsymbol{\alpha}}^{(n)} \coloneqq [\hat{\alpha}_1^{(n)}, \ldots, \hat{\alpha}_M^{(n)}]^{\mathsf{T}} \in \mathbb{C}^M$ is the coefficient vector representing $\hat{u}^{(n)}$, and $\hat{\boldsymbol{\beta}}_m^{(n)} \coloneqq [\hat{\beta}_{m,1}^{(n)}, \ldots, \hat{\beta}_{m,M}^{(n)}]^{\mathsf{T}} \in \mathbb{C}^M$ is the coefficient vector representing $\hat{\zeta}_m^{(n)}$. Therefore, instead of solving for $u, \zeta_1, \ldots, \zeta_M$ in the infinite-dimensional linear space $\mathcal{H}$, we need only to seek the optimal coefficient vectors $\hat{\boldsymbol{\alpha}}^{(n)}, \hat{\boldsymbol{\beta}}_1^{(n)}, \ldots, \hat{\boldsymbol{\beta}}_M^{(n)}$ in the finite dimensional space $\mathbb{C}^M$. By substituting (26) and (27) into (22) and (23), the optimal coefficients $\hat{\boldsymbol{\alpha}}^{(n)}, \hat{\boldsymbol{\beta}}_m^{(n)}$ can obtained by solving the following optimization problem:

$$\underset{\boldsymbol{\alpha} \in \mathbb{C}^M}{\text{minimize}} \quad \left\| (\hat{\boldsymbol{B}}^{(n-1)} + \boldsymbol{I}_M)^{\mathsf{H}} \boldsymbol{K} \boldsymbol{\alpha} - \boldsymbol{s} \right\|^2 + \lambda_1 \boldsymbol{\alpha}^{\mathsf{H}} \boldsymbol{K} \boldsymbol{\alpha}, \tag{28}$$

$$\underset{\boldsymbol{\beta}_m \in \mathbb{C}^M}{\text{minimize}} \quad \left| \boldsymbol{\beta}_m^{\mathsf{H}} \boldsymbol{K} \hat{\boldsymbol{\alpha}}^{(n)} - (s_m - \boldsymbol{k}_m^{\mathsf{H}} \hat{\boldsymbol{\alpha}}^{(n)}) \right|^2 + \lambda_2 \boldsymbol{\beta}_m^{\mathsf{H}} \boldsymbol{K} \boldsymbol{\beta}_m, \tag{29}$$

respectively, where

$$\hat{\boldsymbol{B}}^{(n)} \coloneqq [\hat{\boldsymbol{\beta}}_1^{(n)}, \ldots, \hat{\boldsymbol{\beta}}_M^{(n)}], \tag{30}$$

$$\begin{aligned} \boldsymbol{K} &\coloneqq [\boldsymbol{k}_1, \ldots, \boldsymbol{k}_M] \\ &= \begin{bmatrix} \langle \tilde{v}_1, \tilde{v}_1 \rangle_{\mathcal{H}} & \cdots & \langle \tilde{v}_M, \tilde{v}_1 \rangle_{\mathcal{H}} \\ \vdots & \ddots & \vdots \\ \langle \tilde{v}_1, \tilde{v}_M \rangle_{\mathcal{H}} & \cdots & \langle \tilde{v}_M, \tilde{v}_M \rangle_{\mathcal{H}} \end{bmatrix}, \end{aligned} \tag{31}$$

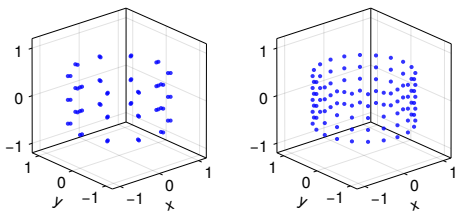
$$\boldsymbol{s} \coloneqq [s_1, \ldots, s_M]^{\mathsf{T}}. \tag{32}$$

Then, by substituting (4) and (19) into (31), we obtain the following expression of each element of $\boldsymbol{K}$, denoted by $K_{m_1,m_2} \coloneqq \langle \tilde{v}_{m_1}, \tilde{v}_{m_2} \rangle_{\mathcal{H}}$, in accordance with [10], [13]:

$$K_{m_1,m_2} = \sum_{\nu_1,\mu_1}^{N_{m_1}} \sum_{\nu_2,\mu_2}^{N_{m_2}} c_{m_1,\nu_1,\mu_1}^* c_{m_2,\nu_2,\mu_2} T_{\nu_1,\mu_1}^{\nu_2,\mu_2}(\boldsymbol{r}_{m_1} - \boldsymbol{r}_{m_2}) \tag{33}$$

where $T_{\nu_1,\mu_1}^{\nu_2,\mu_2}(\cdot)$ is the translation operator [16] given by

$$T_{\nu_1,\mu_1}^{\nu_2,\mu_2}(\boldsymbol{r}) \coloneqq \sum_{\nu_3,\mu_3}^{\nu_1+\nu_2} \mathcal{G}(\nu_1,\mu_1;\nu_2,\mu_2;\nu_3,\mu_3) \varphi_{\nu_3,\mu_3}(\boldsymbol{r}) \tag{34}$$



(a) (b)

Fig. 2: Configuration microphone arrays. (a) a double-layered omnidirectional microphone array; (b) a single-layered cardioid microphone array.

with the modified Gaunt coefficient $\mathcal{G}(\cdot,\cdot;\cdot,\cdot;\cdot,\cdot)$, defined as

$$\mathcal{G}(\nu_1,\mu_1;\nu_2,\mu_2;\nu_3,\mu_3) \coloneqq \frac{1}{4\pi}\int_{\boldsymbol{x}\in\mathbb{S}_2} \hat{Y}_{\nu_1,\mu_1}(\boldsymbol{x})^* \hat{Y}_{\nu_2,\mu_2}(\boldsymbol{x}) \hat{Y}_{\nu_3,\mu_3}(\boldsymbol{x})^* dS. \tag{35}$$

Finally, by defining $\boldsymbol{L}^{(n)} \coloneqq \boldsymbol{B}^{(n)} + \boldsymbol{I}_M$, we obtain the following update rules by minimizing the objective functions of (28) and (29):

$$\hat{\boldsymbol{\alpha}}^{(n)} = \boldsymbol{L}^{(n-1)} \left(\boldsymbol{L}^{(n-1)\mathsf{H}} \boldsymbol{K} \boldsymbol{L}^{(n-1)} + \lambda_1 \boldsymbol{I}_M\right)^{-1} \boldsymbol{s}, \tag{36}$$

$$\hat{\boldsymbol{\beta}}_m^{(n)} = \left(\hat{\boldsymbol{\alpha}}^{(n)\mathsf{H}} \boldsymbol{K} \hat{\boldsymbol{\alpha}}^{(n)} + \lambda_2\right)^{-1} \hat{\boldsymbol{\alpha}}^{(n)} \left(s_m - \boldsymbol{k}_m^{\mathsf{H}} \hat{\boldsymbol{\alpha}}^{(n)}\right)^*. \tag{37}$$

In particular, by collecting (37) for $m = 1, \ldots, M$, we obtain

$$\hat{\boldsymbol{B}}^{(n)} = \left(\hat{\boldsymbol{\alpha}}^{(n)\mathsf{H}} \boldsymbol{K} \hat{\boldsymbol{\alpha}}^{(n)} + \lambda_2\right)^{-1} \hat{\boldsymbol{\alpha}}^{(n)} \left(\boldsymbol{s} - \boldsymbol{K} \hat{\boldsymbol{\alpha}}^{(n)}\right)^{\mathsf{H}}. \tag{38}$$

In conclusion, the optimization algorithm of the proposed method is given as follows:

1: Initialize $\hat{\boldsymbol{B}}^{(0)} = \boldsymbol{O}$

2: Update $\hat{\boldsymbol{\alpha}}^{(n)}$ using (36); Update $\hat{\boldsymbol{B}}^{(n)}$ using (38) } Repeat for N_i iterations

3: $\hat{u}(\boldsymbol{r}) = \sum_{m=1}^{M} \hat{\alpha}_m^{(n)} \tilde{v}_m(\boldsymbol{r})$

Here, N_i denotes the number of iterations. This algorithm does not guarantee convergence to the global optimum of (21), but ensures that its objective function is monotonically non-increasing at each iteration. Moreover, if $\hat{\boldsymbol{B}}^{(n)}$ is not updated, the proposed method is equivalent to our previous sound field estimation method described in Section II.

IV. Numerical Experiments

A. Experimental Conditions

To evaluate the effectiveness of the proposed method, we conducted numerical experiments. We compared the proposed method (denoted by **Proposed**) with our previous method described in Section II (denoted by **Previous**).

As the nominal configuration, two types of cylindrical arrays, each consisting of 120 microphones, were used. As shown in Fig. 2 (a), the first configuration consists of two concentric cylindrical layers of omnidirectional microphones. The outer and inner layers have radii of $1.00\,\mathrm{m}$ and $0.95\,\mathrm{m}$, respectively. On each layer, 20 microphones are arranged in three vertical tiers at $z = 0, 0.5$, and $-0.5\,\mathrm{m}$. The second configuration, illustrated in Fig. 2 (b), consists of a single cylindrical layer of outward-facing cardioid microphones. This layer has a radius of $1.00\,\mathrm{m}$, with microphones arranged in five vertical tiers at $z = 0, -0.25, 0.25, -0.5$, and $0.5\,\mathrm{m}$.

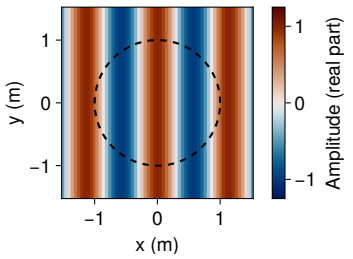


Fig. 3: The true sound field at $300\,\mathrm{Hz}$ plotted in xy-plane.

The true sound field u_{true} was set as a single plane wave propagating from the direction $\boldsymbol{d} = [1,0,0]^{\mathsf{T}}$, and is represented as

$$u_{\text{true}}(\boldsymbol{r}) \coloneqq \exp(-ik\boldsymbol{d}\cdot\boldsymbol{r}). \tag{39}$$

The speed of sound was set to $c = 340\,\mathrm{m/s}$. An example of the true sound field is shown in Fig. 3, where the black line denotes the boundary of the microphone array, i.e., a circle with a radius of $1\,\mathrm{m}$ As an evaluation metric, we used the normalized mean squared error (NMSE), defined as follows:

$$\text{NMSE} \coloneqq 10\log_{10} \frac{\sum_{i\in\mathrm{I}_{\text{eval}}} |u_{\text{true}}(\boldsymbol{r}_{\text{eval}}^{(i)}) - u_{\text{est}}(\boldsymbol{r}_{\text{eval}}^{(i)})|^2}{\sum_{i\in\mathrm{I}_{\text{eval}}} |u_{\text{true}}(\boldsymbol{r}_{\text{eval}}^{(i)})|^2} (\text{dB}). \tag{40}$$

Here, u_{est} denotes the estimated sound field, and $\{\boldsymbol{r}_{\text{eval}}^{(i)}\}_{i\in\mathrm{I}_{\text{eval}}}$ represents the set of evaluation grid points spaced at $0.1\,\mathrm{m}$ within a cylindrical region of radius $1\,\mathrm{m}$ and height ranging from $z = -0.5\,\mathrm{m}$ to $z = 0.5\,\mathrm{m}$. A smaller NMSE value indicates higher estimation accuracy.

B. Estimation Using a Double-Layered Omnidirectional Microphone Array

First, we evaluated estimation performance with a double-layered omnidirectional microphone array in the presence of discrepancies between the nominal and actual microphone positions. The position error was modeled as

$$\boldsymbol{r}_m^{\text{act}} = \tilde{\boldsymbol{r}}_m + \boldsymbol{e}_m, \tag{41}$$

where $\boldsymbol{r}_m^{\text{act}}$ and $\tilde{\boldsymbol{r}}_m$ denote respectively the actual and nominal positions, and $\boldsymbol{e}_m$ was sampled from the isotropic three-dimensional normal distribution with standard deviation σ_r in each coordinate.

The observed signal $s_1, \ldots, s_M$ for a sound field u were given by

$$s_m = u(\boldsymbol{r}_m^{\text{act}}) + \epsilon_m, \tag{42}$$

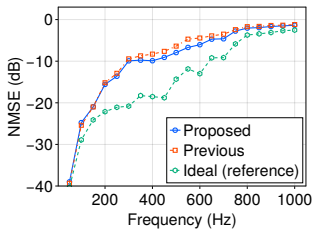


Fig. 4: NMSE plotted against frequency for a double-layered omnidirectional microphone array.

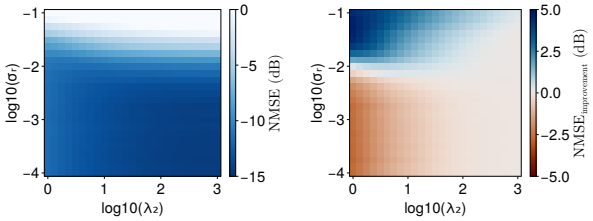


Fig. 5: NMSE with respect to λ_2 and σ_r at 500 Hz. The left panel shows **Proposed**; the right panel shows the NMSE improvement.

where ϵ_m denotes the observation noise, which was sampled from an independent and identically distributed (i.i.d.) complex normal distribution so that the signal to noise ratio (SNR) was 30 dB. The nominal directivity function and its corresponding spherical harmonic coefficients of each microphone was given as

$$\gamma_m(\boldsymbol{x}) = 1 \quad (\boldsymbol{x} \in \mathbb{S}_2), \tag{43}$$

$$c_{m,\nu,\mu} = \begin{cases} 1 & (\nu = 0) \\ 0 & (\text{otherwise}) \end{cases}. \tag{44}$$

Fig. 4 shows the relationship between frequency and NMSE. Here, position error was set to $\sigma_r = 0.02\,\mathrm{m}$, and the parameters were set to $\lambda_1 = 10^{-3}$, $\lambda_2 = 10$, and $N_i = 100$ in **Proposed** and $\lambda = 10^{-3}$ in **Previous**. The label **Ideal (reference)** denotes **Previous** applied under the assumption of perfect knowledge of microphone positions and directivities, serving as an upper bound on the performance of methods that compensate for such errors. **Proposed** exhibited a lower NMSE than **Previous** in the frequency range of approximately 200 to 800 Hz, while showing comparable NMSE values outside this range. It is likely that, at low frequencies, the relatively long wavelengths reduced the impact of position errors, leading the proposed method to apply only minimal correction to the observation operator. For further investigation, the NMSE with respect to λ_2 and σ_r were plotted at 500 Hz in Fig. 5, with the other parameters set as previously described. The left panel presents the NMSE of **Proposed**, and the right panel shows the NMSE improvement, calculated as

$$\mathrm{NMSE}_{\mathrm{improvement}} := \mathrm{NMSE}_{\textbf{Previous}} - \mathrm{NMSE}_{\textbf{Proposed}}. \tag{45}$$

When σ_r ranged from 10^{-2} to 10^{-1}, **Proposed** achieved a lower NMSE than **Previous** across nearly all values of λ_2. In contrast, when σ_r was smaller than 10^{-2}, little to no improvement was observed, and in the range of $\lambda_2 = 0$ to 2 dB, performance even deteriorated. In both cases, the difference between **Proposed** and **Previous** diminished as λ_2 increased, which aligns which aligns with the theoretical expectation that **Proposed** converges to **Previous** as $\lambda_2 \to \infty$. This suggests that although the optimal value of λ_2 varies depending on the position error, the proposed method is effective across a wide range of λ_2 when the error is large. The development of a method to optimally select λ_2 remains a topic for future research.

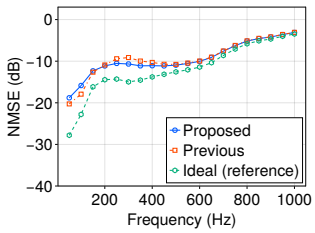


Fig. 6: NMSE plotted against frequency for a single-layered cardioid microphone array.

C. Estimation Using a Single-Layered Cardioid Microphone Array

Next, we evaluated the estimation performance using a single-layered cardioid microphone array in the presence of discrepancies in both microphone positions and directivities. The observed signals $s_1, \ldots, s_M$ were given by

$$s_m = a_m u(\boldsymbol{r}_m^{\mathrm{act}}) + \frac{1 - a_m}{ik} \frac{\partial u(\boldsymbol{r}_m^{\mathrm{act}})}{\partial \boldsymbol{y}_m^{\mathrm{act}}} + \epsilon_m, \tag{46}$$

where a_m denotes the directivity parameter, $\boldsymbol{r}_m^{\mathrm{act}}, \boldsymbol{y}_m^{\mathrm{act}}$ are the actual position and orientation, $\partial/\partial \boldsymbol{y}_m^{\mathrm{act}}$ denotes the directional derivative along the direction $\boldsymbol{y}_m^{\mathrm{act}}$, and ϵ_m is the observation noise. The actual orientation $\boldsymbol{y}_m^{\mathrm{act}}$ was modeled by rotating the nominal orientation $\tilde{\boldsymbol{y}}_m$ by a random angle $\theta' \sim \mathcal{N}(0, \sigma_{\theta'})$. The directivity parameters $a_1, \ldots, a_M$ were perturbed as

$$a_m = \frac{1}{2} + e_m \quad (e_m \sim \mathcal{N}(0, \sigma_a)), \tag{47}$$

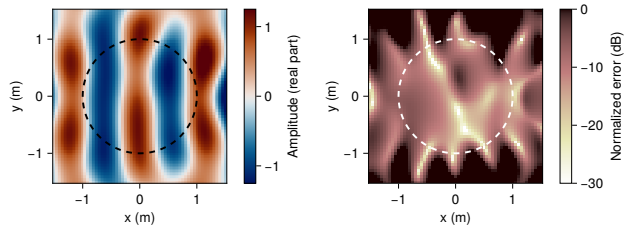
where $a_m = \frac{1}{2}$ corresponds to ideal cardioid characteristics. We set $\sigma_r = 0.02\,\mathrm{m}, \sigma_{\theta'} = 0.01\,\mathrm{rad}$, and $\sigma_a = 0.05$. The nominal directivity function and its corresponding spherical harmonic coefficients of each microphone are given by

$$\gamma_m(\boldsymbol{x}) = \frac{1}{2} + \frac{1}{2}\tilde{\boldsymbol{y}}_m \cdot \boldsymbol{x} \quad (\boldsymbol{x} \in \mathbb{S}_2), \tag{48}$$

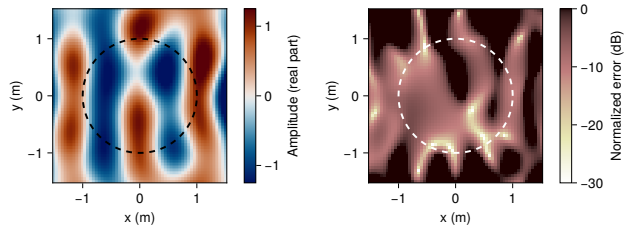
$$c_{m,\nu,\mu} = \begin{cases} 1/2 & (\nu = 0) \\ \hat{Y}_{1,\mu}(\tilde{\boldsymbol{y}}_m)^*/6 & (\nu = 1) \\ 0 & (\text{otherwise}) \end{cases}. \tag{49}$$

The remaining parameters, $\boldsymbol{r}_m^{\mathrm{act}}$ and ϵ_m, were configured in the same manner as in the experiment described in Section IV-B.

Fig. 6 shows the relationship between frequencies and NMSEs, under the same parameter settings as in Fig. 4. Between



(a) Estimated sound field (**Proposed**)



(b) Estimated sound field (**Previous**)

Fig. 7: Results of sound field estimation using a single-layer cardioid microphone array at 300 Hz in the xy-plane. The left panels show the sound pressure distributions, and the right panels show the corresponding normalized errors.

approximately 200 Hz and 500 Hz, **Proposed** achieved lower NMSE than **Previous**. Although the improvement was smaller compared to Fig. 4, a generally similar trend was observed.

Fig. 7 shows an example of the estimated sound fields and the pointwise normalized error distribution for both **Proposed** and **Previous** methods. In this example, the NMSEs were (a) $-9.80\,\mathrm{dB}$ and (b) $-6.83\,\mathrm{dB}$. As seen in this figures, **Proposed** achieved higher estimation accuracy than **Previous**.

V. Conclusions

We proposed a sound field estimation method that is robust to errors in microphone positions and directivities by formulating a joint estimation problem of the sound field and the observation operator. To solve this joint problem, we developed an alternation optimization approach, and derived closed-form solutions for each iteration. Numerical experiments demonstrated that the proposed method achieves higher estimation accuracy than our previous method in the presence of errors in microphone positions and directivities.

Acknowledgment

This work was supported by JSPS KAKENHI Grant Number JP23K24864.

References

[1] A. J. Berkhout, D. de Vries, and P.Vogel, "Acoustic control by wave field synthesis," *J. Acoust. Soc. Am.*, vol. 93, no. 5, pp. 2764–2778, 1993.

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

[3] D. Menzies and M. Al–Skaidi, "Nearfield binaural synthesis and ambisonics," *J. Acoust. Soc. Am.*, vol. 121, no. 3, pp. 1559–1563, 2007.

[4] S. Koyama, K. Kimura, and N. Ueno, "Weighted pressure and mode matching for sound field reproduction: Theoretical and experimental comparisons," *J. Audio Eng. Soc.*, vol. 71, no. 4, pp. 173–185, 2023.

[5] S. J. Elliott and P. A. Nelson, "The active control of sound," *Electronics & Communication Engineering Journal*, vol. 2, no. 4, pp. 127–136, 1990.

[6] S. Koyama, J. Brunnström, H. Ito, N. Ueno, and H. Saruwatari, "Spatial active noise control based on kernel interpolation of sound field," *IEEE/ACM Trans. Audio Speech Lang. Proc.*, vol. 29, pp. 3052–3063, 2021.

[7] H. Ito, S. Koyama, N. Ueno, and H. Saruwatari, "Feedforward spatial active noise control based on kernel interpolation of sound field," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2019, pp. 511–515.

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

[9] J. Daniel, R. Nicol, and S. Moureau, "Further investigations of high order ambisonics and wavefield synthesis for holophonic sound imaging," *Audio Engineering Society Convention 114*, vol. 51, 2003.

[10] N. Ueno and S. Koyama, "Sound field estimation: Theories and applications," *Foundations and Trends in Signal Processing*, vol. 19, no. 1, pp. 1–98, 2025.

[11] A. Laborie, R. Bruno, and S. Montoya, "A new comprehensive approach of surround sound recording," in *Audio Engineering Society Convention 114*, 2003.

[12] N. Ueno, S. Koyama, and H. Saruwatari, "Sound field recording using distributed microphones based on harmonic analysis of infinite order," *IEEE Signal Processing Letters*, vol. 25, no. 1, pp. 135–139, 2018.

[13] N. Ueno, S. Koyama, and H. Saruwatari, "Directionally weighted wave field estimation exploiting prior information on source direction," *IEEE Trans. Signal Process.*, vol. 69, pp. 2383–2395, 2021.

[14] D. Colton and R. Kress, *Inverse acoustic and electromagnetic scattering theory*. Springer, 1998.

[15] E. M. Stein and R. Shakarchi, *Real Analysis. Measure Theory, Integration and Hilbert Spaces*. Princeton University Press, 2005.

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

[17] J. B. Conway, *A Course in Functional Analysis*. Springer, 2007.

[18] B. Schölkopf, R. Herbrich, and A. J. Smola, "A generalized representer theorem," in *Proc. Comput. Learn. Theory*, 2001, pp. 416–426.

[19] F. Dinuzzo and B. Schölkopf, "The representer theorem for Hilbert spaces: A necessary and sufficient condition," in *Advance in Neural Information Processing Systems*, 2012, pp. 189–196.