

FEBRUARY 21 2025

Sound field estimation for source-included region based on Gaussian process using prior source information

Ryo Matsuda  ; Makoto Otani 



J. Acoust. Soc. Am. 157, 1403–1417 (2025)

<https://doi.org/10.1121/10.0035941>



Articles You May Be Interested In

Using learned priors to regularize the Helmholtz equation least-squares method

J. Acoust. Soc. Am. (February 2024)

Instantaneous Bayesian regularization applied to real-time near-field acoustic holography

J. Acoust. Soc. Am. (August 2017)

Generative adversarial networks with physical sound field priors

J. Acoust. Soc. Am. (August 2023)



Advance your science and career as a member of the

Acoustical Society of America

[LEARN MORE](#)

Sound field estimation for source-included region based on Gaussian process using prior source information

Ryo Matsuda^{a)}  and Makoto Otani 

Department of Architecture and Architectural Engineering, Graduate School of Engineering, Kyoto University, Kyoto-daigaku-katsura, Nishikyo-ku, Kyoto, 615-8540, Japan

ABSTRACT:

Estimating a sound field in a region that includes sources (i.e., an inhomogeneous sound field) is challenging. This paper proposes the Gaussian process (GP) for estimating an inhomogeneous sound field in the case of anechoic condition. A kernel function is formulated as a weighted spatial correlation of free-field transfer functions in the modal domain. The weights for the kernel function are derived by introducing the probability distribution of source positions in spherical regions containing the sound sources. Here, a weight obtained by analytically solving the spherical integral with the probability distribution as Gaussian is proposed. Schemes of order truncation and hyperparameter optimization for the kernel function are also proposed. Compared with conventional methods, numerical experiments reveal that the proposed method achieves higher sound field estimation accuracy. In addition, Gaussian process regression, using the kernel function with the proposed weight, achieves higher estimation accuracy with lower computational cost than those using the kernel functions with other weights. Moreover, the advantages of the proposed method, which are obtained by treating the sound source as a distribution rather than a point source, are revealed. © 2025 Acoustical Society of America. <https://doi.org/10.1121/10.0035941>

(Received 9 September 2024; revised 24 December 2024; accepted 4 February 2025; published online 21 February 2025)

[Editor: Efren Fernandez-Grande]

Pages: 1403–1417

I. INTRODUCTION

Estimating a time-space continuous sound pressure distribution (i.e., a sound field) is essential for many acoustic applications, including visualization and auralization.^{1–6} Visualization helps to understand the acoustic phenomena, whereas auralization is necessary for virtual reality applications, immersive telecommunications, and feedback for architectural acoustic design.

The estimation process is generally described as an inverse problem of estimating the sound pressures in a target region from those measured by distributed microphones,^{7–12} a single microphone array,^{13–17} or multiple microphone arrays.^{18–20} This paper focuses on the estimation of the sound field in target areas that include sound sources, specifically in cases where the sound field is described by the inhomogeneous Helmholtz equation.²¹

Several methods have been proposed to estimate such sound fields. Sparse sound field decomposition^{12–14,16,20} reconstructs the sound field by decomposing it into sparse bases and their coefficients using sparse optimization.^{22–28} This method may improve the estimation accuracy when the sound field meets the sparsity (i.e., when the sound field is composed of a few monopoles or plane waves); however, there are several practical problems. First, the basis functions must be designed based on the location of the prior discretized target region. Therefore, the estimation accuracy over the target region may decrease if the true source positions are not one of the defined candidates. Second,

estimation accuracy has a trade-off relationship with memory and computational cost: higher accuracy increases cost, making it unsuitable for real-time processing. Finally, sparse sound field decomposition cannot be implemented as finite-impulse response (FIR) filters because of the nonlinearity of the decomposition process. An alternative method is spherical-harmonic-domain reciprocity gap functional (SHD-RGF),²⁹ which analytically estimates the positions of sound sources via test functions without any prior discretization. However, because the sound source is assumed to exist in the entire region of interest enclosed by the distributed microphones, spatial aliasing is determined by the region size and only effective at lower frequencies compared with sparse sound field decomposition.

Recently, a spatial-correlation-based method,^{30,31} known as Gaussian process regression (GPR) or kernel ridge regression (KRR), has been proposed. It enables the sound field to be estimated using a more flexible microphone arrangement, rather than a conventional spherical array or circular array, by projecting measured space into high-dimensional feature space using a kernel function. Thus, unlike spherical-wave-function (SWF) expansion²¹ or plane-wave-function expansion,³² GPR does not require direct consideration of the expansion basis functions. Previous studies have revealed that the kernel functions designed to satisfy the physical constraints of the Helmholtz equation are more effective than traditional kernels such as the radial basis function kernel.^{7,10} Studies have also revealed that the kernel function can adapt to the sound field, resulting in greater estimation accuracy than that achieved without adaptation.^{8,9,11,30,31} Furthermore, one of

^{a)}Email: matsuda.ryo.45s@st.kyoto-u.ac.jp

the advantages of the method is that it can be implemented as FIR filters; thus, no additional calculations are required if the sound field does not change due to movement of the sound sources or changes in room characteristics. However, these kernel functions are derived from the homogeneous Helmholtz equation, which assumes that no sound sources exist within the target region. Therefore, these functions cannot be directly applied when the sound sources are present within the target region.

In this paper, we propose a Gaussian process (GP) for estimating the sound pressures in a region, including sources, i.e., following the inhomogeneous Helmholtz equation, in the case of anechoic conditions. The kernel function is defined as the spherical integration of the spatial correlation of the transfer function from the sound source region to the microphones, weighted by the probability distribution of the presence of sound sources, resulting in a representation through weighted spatial correlation. This definition is similar to that in Ref. 33; however, our proposed kernel function is based on rigorous integration when a Gaussian distribution is used as the probability distribution. In addition, we make the following contributions: proposing the kernel design for the case of multiple sound sources and schemes for initializing the hyperparameters of the kernel function using prior source information and updating them. This work is the first attempt to estimate the sound field in the presence of sound sources using such a kernel function. The advantages of GPR using the proposed kernel function are threefold. (1) The proposed kernel function not only achieves higher sound field estimation accuracy compared to the conventional kernel function, but also requires lower computational cost; (2) because the proposed method treats a sound source as a distribution rather than a point source in the kernel function, it does not require obtaining strict source positions; and (3) when the source position undergoes small movements, the once-optimized FIR filters can continue to be used without any additional computation. These advantages of the proposed method are demonstrated through numerical experiments.

This paper is organized as follows. Section II defines the notation used throughout this paper and reviews proper complex GPR. Section III describes our GPR framework by proposing the kernel functions along with the scheme of truncating order and optimizing hyperparameters. Section IV reviews SWF expansion and the sparse decomposition method as the conventional methods. Section V presents the numerical experiments and their results. Section VI discusses the proposed method. Section VII concludes this study.

II. NOTATIONS AND PRELIMINARIES

A. Notations

The sets of real numbers, complex numbers, and positive integers are denoted by $\mathbb{R}$, $\mathbb{C}$, and $\mathbb{Z}_+$, respectively. The imaginary unit is denoted by j . Uppercase bold $\mathbf{A}$

denotes a matrix, whereas lowercase bold $\mathbf{a}$ represents a vector. $[\mathbf{A}]_{i,j}$ refers to the (i,j) th element of matrix $\mathbf{A}$, and $[\mathbf{a}]_i$ denotes the i th element of vector $\mathbf{a}$. In the spherical coordinate system (r, θ, ϕ) , θ represents the zenith angle measured from the z axis, and ϕ denotes the azimuthal angle measured counterclockwise from the x axis. The transpose is denoted by $(\cdot)^T$. The complex conjugate and conjugate transpose are denoted by $(\cdot)^*$ and $(\cdot)^H$, respectively. The wavenumber is denoted by $k = \omega/c$, where ω and c are the angular frequency and sound speed, respectively. Throughout this paper, sound field estimation is considered in the wavenumber domain, assuming that the time dependency is $\exp(-j\omega t)$, where t denotes time. Hereafter, k is omitted for simple notation.

B. Proper complex GPR

The framework of zero-mean proper complex GPR^{8,34} is briefly reviewed. Suppose sound pressures are measured by M microphones located at $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_M]^T \in \mathbb{C}^{M \times D}$, where D denotes the dimension of the measured field. The measured sound pressures $\mathbf{y} = [y_1, \dots, y_M]^T \in \mathbb{C}^M$ at these positions $\mathbf{X}$ are modeled by

$$\mathbf{y} = \mathbf{u} + \mathbf{e}, \quad (1)$$

where $\mathbf{u} = [u(\mathbf{x}_1), \dots, u(\mathbf{x}_M)]^T \in \mathbb{C}^M$ is the sound field vector obtained through the unknown function $u: \mathbb{R}^D \mapsto \mathbb{C}$, representing the sound field, and $\mathbf{e} \in \mathbb{C}^M$ represents the measurement noise. Assuming that the unknown function $u(\mathbf{x})$ is represented by the superposition of the basis function $\Phi(\mathbf{x}): \mathbb{R}^D \mapsto \mathbb{C}^\infty$ and its coefficients $\mathbf{w} \in \mathbb{C}^\infty$, $\mathbf{u}$ can be expressed as

$$\mathbf{u} = \Phi \mathbf{w}, \quad (2)$$

where $\Phi = [\Phi(\mathbf{x}_1), \dots, \Phi(\mathbf{x}_M)]^T \in \mathbb{C}^{M \times \infty}$. Here, each $\mathbf{w}$ and $\mathbf{e}$ are assumed to follow independent and identically distributed complex Gaussian distributions as

$$p(\mathbf{w}) = \mathcal{CN}(\mathbf{0}, \sigma_w^2 \mathbf{I}_M), \quad (3a)$$

$$p(\mathbf{e}) = \mathcal{CN}(\mathbf{0}, \sigma_e^2 \mathbf{I}_M), \quad (3b)$$

where $\mathbf{I}_M \in \mathbb{R}^{M \times M}$, σ_w , and σ_e denote the identity matrix of size M , the variance of $\mathbf{w}$, and the variance of $\mathbf{e}$, respectively. Because Φ is a linear operator and $\mathbf{w}$ follows the complex Gaussian [Eq. (3a)], the GP prior $p(\mathbf{u})$ also follows the complex Gaussian such that³⁴

$$p(\mathbf{u}) = \mathcal{CN}(\mu_u, \mathbf{K}), \quad (4)$$

where mean $\mu_u \in \mathbb{C}^M$ and covariance $\mathbf{K} \in \mathbb{C}^{M \times M}$ are

$$\mu_u = \Phi \mathbb{E}[\mathbf{w}] = \mathbf{0}, \quad (5a)$$

$$\mathbf{K} = \Phi \mathbb{E}[\mathbf{w} \mathbf{w}^H] \Phi^H = \sigma_w^2 \Phi \Phi^H. \quad (5b)$$

Here, $\mathbf{K}$ is the Gram matrix whose element is given by $[\mathbf{K}]_{i,j} = \kappa(\mathbf{x}_i, \mathbf{x}_j) = \sigma_w^2 \Phi(\mathbf{x}_j)^H \Phi(\mathbf{x}_i)$, where $\kappa(\cdot, \cdot): \mathbb{R}^D \times$

$\mathbb{R}^D \mapsto \mathbb{C}$ represents the kernel function. The kernel function is indirectly determined by selecting basis functions $\Phi(\cdot)$ here. However, it can also be directly defined if the Gram matrix $\mathbf{K}$ is a positive semi-definite matrix.^{34–36} The likelihood $p(\mathbf{y}|\mathbf{u})$ can be obtained by combining Eq. (1) with Eq. (3b),

$$p(\mathbf{y}|\mathbf{u}) = \mathcal{CN}(\mathbf{u}, \sigma_e^2 \mathbf{I}_M). \quad (6)$$

Hence, by combining Eqs. (4) and (6), we obtain the marginal likelihood as

$$p(\mathbf{y}) = \int p(\mathbf{y}|\mathbf{u})p(\mathbf{u})d\mathbf{u} = \mathcal{CN}(\mathbf{0}, \mathbf{K} + \sigma_e^2 \mathbf{I}_M). \quad (7)$$

Consider estimating the sound pressure $u_\bullet$ at unobserved location $\mathbf{x}_\bullet$. Because u and $\mathbf{y}$ follow the complex Gaussian distribution, the joint distribution of $u_\bullet$ and $\mathbf{y}$ also follows the complex Gaussian distribution,³⁴ which is expressed as

$$p\left(\begin{bmatrix} \mathbf{y} \\ u_\bullet \end{bmatrix}\right) = \mathcal{CN}\left(\mathbf{0}, \begin{bmatrix} \mathbf{K} + \sigma_e^2 \mathbf{I} & \boldsymbol{\kappa}_\bullet \\ \boldsymbol{\kappa}_\bullet^H & \kappa_{\bullet\bullet} \end{bmatrix}\right), \quad (8)$$

where $\boldsymbol{\kappa}_\bullet = [\kappa(\mathbf{x}_1, \mathbf{x}_\bullet), \kappa(\mathbf{x}_2, \mathbf{x}_\bullet), \dots, \kappa(\mathbf{x}_M, \mathbf{x}_\bullet)]^T \in \mathbb{C}^M$ and $\kappa_{\bullet\bullet} = \kappa(\mathbf{x}_\bullet, \mathbf{x}_\bullet)$. Therefore, the estimated sound pressure $u_\bullet$ can be obtained as the conditional distribution of $u_\bullet$ given $\mathbf{y}$, which is expressed as

$$p(u_\bullet|\mathbf{y}) = \mathcal{CN}(\mu_{u_\bullet|\mathbf{y}}, \sigma_{u_\bullet|\mathbf{y}}), \quad (9)$$

where the posterior mean $\mu_{u_\bullet|\mathbf{y}}$ and variance $\sigma_{u_\bullet|\mathbf{y}}$ are

$$\mu_{u_\bullet|\mathbf{y}} = \boldsymbol{\kappa}_\bullet^H (\mathbf{K} + \sigma_e^2 \mathbf{I}_M)^{-1} \mathbf{y}, \quad (10a)$$

$$\sigma_{u_\bullet|\mathbf{y}} = \kappa_{\bullet\bullet} - \boldsymbol{\kappa}_\bullet^H (\mathbf{K} + \sigma_e^2 \mathbf{I}_M)^{-1} \boldsymbol{\kappa}_\bullet. \quad (10b)$$

III. GP FOR INHOMOGENEOUS FREE FIELD

A. Kernel function for a single source region

Assume that point sources exist within a spherical region $\Omega^{(s)}$ of radius $R^{(s)}$ centered at the origin $O^{(s)}$, where $s \in \{1, \dots, S\} =: \mathcal{S}$ denotes the index number of the regions, and the case of $s = 1$ is considered here. The kernel function can be defined as^{33,37}

$$\kappa(\mathbf{r}_i^{(s)}, \mathbf{r}_j^{(s)}) := \int_{\Omega^{(s)}} G(\mathbf{r}_i^{(s)}|\mathbf{r}^{(s)}) \overline{G(\mathbf{r}_j^{(s)}|\mathbf{r}^{(s)})} \eta(r^{(s)}) d\Omega^{(s)}, \quad (11)$$

where $\mathbf{r}^{(s)} = (r^{(s)}, \theta^{(s)}, \phi^{(s)})$ and $\mathbf{r}_i^{(s)} = (r_i^{(s)}, \theta_i^{(s)}, \phi_i^{(s)})$ denote the position of the sound source and the microphone relative to $O^{(s)}$ in the spherical coordinates, respectively; $\eta(r^{(s)})$ denotes the probability density function of the sound source, which only depends on the radius $r^{(s)}$ under the assumption of a spherical symmetric distribution; $G(\mathbf{r}_i^{(s)}|\mathbf{r}^{(s)})$ represents the transfer function between $\mathbf{r}_i^{(s)}$ and $\mathbf{r}^{(s)}$, which is also called the three-dimensional free-field Green's function.²¹ The integral expresses

$$\int_{\Omega^{(s)}} d\Omega^{(s)} = \int_0^\pi \int_0^{2\pi} \int_0^R r^{(s)2} \sin \theta^{(s)} dr^{(s)} d\phi^{(s)} d\theta^{(s)}. \quad (12)$$

To solve the integral in Eq. (11), we use the SWF expansion of Green's function,²¹ expressed as

$$G(\mathbf{r}_i^{(s)}|\mathbf{r}^{(s)}) = jk \sum_{n=0}^\infty \sum_{m=-n}^n j_n(kr^{(s)}) h_n(kr_i^{(s)}) \overline{Y_n^m(\theta^{(s)}, \phi^{(s)})} \times Y_n^m(\theta_i^{(s)}, \phi_i^{(s)}), \quad r_i^{(s)} > r^{(s)}, \quad (13)$$

where $Y_n^m(\cdot)$ is the spherical harmonic function of order n and degree m , $j_n(\cdot)$ denotes the n th-order spherical Bessel function of the first kind, and $h_n(\cdot)$ denotes the n th-order spherical Hankel function of the first kind.²¹ Using the orthogonality and addition theorem of spherical harmonics,²¹ Eq. (11) yields³³

$$\begin{aligned} \kappa(\mathbf{r}_i^{(s)}, \mathbf{r}_j^{(s)}) &= \int_{\Omega^{(s)}} \left(\sum_{n=0}^\infty \sum_{m=-n}^n j_k j_n(kr^{(s)}) h_n(kr_i^{(s)}) \overline{Y_n^m(\theta^{(s)}, \phi^{(s)})} Y_n^m(\theta_i^{(s)}, \phi_i^{(s)}) \right) \\ &\quad \times \left(\sum_{n'=0}^\infty \sum_{m'=-n'}^n -j_k j_{n'}(kr^{(s)}) \overline{h_{n'}(kr_j^{(s)})} Y_{n'}^{m'}(\theta^{(s)}, \phi^{(s)}) \overline{Y_{n'}^{m'}(\theta_j^{(s)}, \phi_j^{(s)})} \right) \eta(r^{(s)}) d\Omega^{(s)} \\ &= \sum_{n=0}^\infty \sum_{m=-n}^n h_n(kr_i^{(s)}) Y_n^m(\theta_i^{(s)}, \phi_i^{(s)}) \overline{h_n(kr_j^{(s)}) Y_n^m(\theta_j^{(s)}, \phi_j^{(s)})} \underbrace{k^2 \int_0^{R^{(s)}} j_n(kr^{(s)})^2 \eta(r^{(s)}) r^{(s)2} dr^{(s)}}_{\xi_n} \\ &= \sum_{n=0}^\infty \frac{2n+1}{4\pi} \xi_n P_n(\cos \Theta_{ij}^{(s)}) h_n(kr_i^{(s)}) \overline{h_n(kr_j^{(s)})}, \end{aligned} \quad (14)$$

where the integral term is denoted by the weight function ξ_n , $P_n(\cdot)$ denotes the Legendre polynomials of order n ,²¹

and $\Theta_{ij}^{(s)}$ denotes the angle between $\mathbf{r}_i$ and $\mathbf{r}_j$. The infinite series in terms of n in Eq. (14) must be truncated at a finite

order $N^{(s)} \in \mathbb{Z}_+$ in practice. The kernel function in the second line of Eq. (14), truncated at the $N^{(s)}$ th-order can also be described in a matrix formulation, which is expressed as

$$\kappa(\mathbf{r}_i^{(s)}, \mathbf{r}_j^{(s)}) = \mathbf{h}_j^{(s)\text{H}} \mathbf{\Xi}^{(s)} \mathbf{h}_i^{(s)}, \quad (15)$$

where $\mathbf{h}_i^{(s)} \in \mathbb{C}^{V^{(s)}}$ is the vector of a SWF whose element is $[\mathbf{h}_i^{(s)}]_{n^2+n+m+1} = h_n(kr_i^{(s)}) Y_n^m(\theta_i^{(s)}, \phi_i^{(s)})$, and $\mathbf{\Xi}^{(s)} \in \mathbb{R}^{V^{(s)} \times V^{(s)}}$ is the diagonal weighting matrix with $[\mathbf{\Xi}^{(s)}]_{n^2+n+m+1, n^2+n+m+1} = \xi_n$, with the number of modes $V^{(s)} = (N^{(s)} + 1)^2$. To clarify the methodology for later comparisons, we have explicitly separated the vectors $\mathbf{h}_i^{(s)}$ and $\mathbf{h}_j^{(s)}$, which depend on the microphone positions, from the weighting matrix $\mathbf{\Xi}^{(s)}$, which depends solely on the probability distribution of the sound sources. A strategy for determining the truncation order $N^{(s)}$ will be proposed in Sec. III C. In addition, the kernel function represented by Eq. (15) is proven to be a valid kernel in the Appendix.

Two probability density functions are considered here.³³ Uniform distribution η_{uni} , which assumes the uniform distribution of sound source inside $\Omega^{(s)}$, is defined by

$$\eta_{\text{uni}}(r^{(s)}) = \frac{3}{4\pi R^{(s)3}}, \quad (16)$$

and Gaussian distribution η_{Gs} , which estimates the higher probability closer to the origin, is defined by

$$\eta_{\text{Gs}}(r^{(s)}) = \frac{1}{\sqrt{8\pi^3} \sigma^{(s)3}} \exp\left(-\frac{r^{(s)2}}{2\sigma^{(s)2}}\right), \quad (17)$$

where $\sigma^{(s)}$ denotes the variance. Note that the Gaussian distribution is normalized within the range $r^{(s)} \in [0, \infty)$.

The weight function ξ_n can be obtained analytically from the above definition of probability distribution $\eta_n(r^{(s)})$. For a uniform distribution, the weight function $\xi_{\text{uni},n}$ can be derived by substituting Eq. (16) into Eq. (14) such that³³

$$\xi_{\text{uni},n} = \frac{3k^2}{8\pi} [j_n(kR^{(s)})^2 - j_{n-1}(kR^{(s)})j_{n+1}(kR^{(s)})]. \quad (18)$$

Similarly, the weight function for the Gaussian distribution ξ_{Gs} can be derived analytically by substituting Eq. (17) into Eq. (14) such that⁶

$$\begin{aligned} \xi_{\text{Gs},n} &= \frac{k^2 R^{(s)3}}{\sqrt{32\pi^3} \sigma^{(s)3}} \exp\left(-\frac{R^{(s)2}}{4\sigma^{(s)2}}\right) \\ &\times \sum_{\nu=0}^{\infty} (-1)^\nu i_\nu\left(\frac{R^{(s)2}}{4\sigma^{(s)2}}\right) [j_{n-\nu}(kR^{(s)})j_{n+\nu}(kR^{(s)}) \\ &- j_{n-\nu-1}(kR^{(s)})j_{n+\nu+1}(kR^{(s)})], \end{aligned} \quad (19)$$

where $i_\nu(\cdot)$ is the ν th-order modified spherical Bessel function of the first kind. Notably, the Gaussian weight function derived in Eq. (19) assumes an integration range of $[0, R^{(s)}]$,

thereby strictly considering the presence of sound sources only within $\Omega^{(s)}$; by contrast, the integration range for $r^{(s)}$ in the Gaussian weight function defined in Ref. 33, expressed as

$$\begin{aligned} \xi_{\text{Gs-Inf},n} &:= k^2 \int_0^\infty j_n(kr^{(s)})^2 \eta_{\text{Gs}}(r^{(s)}) r^{(s)2} dr^{(s)} \\ &= \frac{k^2}{4\pi} \exp(-k^2 \sigma^2) i_n(k^2 \sigma^2), \end{aligned} \quad (20)$$

is $[0, \infty)$. Thus, the weight function defined in Eq. (20) would result in regions where $r_i^{(s)} < r^{(s)}$, thereby violating the relationship in Eq. (13) in a strict sense.

In addition, the conventional SWF-expansion-based sound field estimation can be interpreted as a GPR with the kernel function $\kappa(\mathbf{r}_i^{(s)}, \mathbf{r}_j^{(s)}) = \mathbf{h}_j^{(s)\text{H}} \mathbf{h}_i^{(s)}$, as described in Secs. IV A and the Appendix. The kernel function is equivalent to Eq. (15) when the weight function is set as

$$\xi_{\text{SWF},n} := 1. \quad (21)$$

B. Kernel function for multiple source regions

Suppose sound sources exist within the multiple spherical regions $\{\Omega^{(s)}\}_{s \in \mathcal{S}}$. Then, the kernel function can be a summation of kernel functions for each source region $\Omega^{(s)}$ [Eq. (15)], expressed as

$$\kappa(\mathbf{r}_i, \mathbf{r}_j) = \sum_{s=1}^S \kappa(\mathbf{r}_i^{(s)}, \mathbf{r}_j^{(s)}). \quad (22)$$

Here, we used the property that the summation of valid kernel functions results in a valid kernel function.³⁵

C. Truncation order

We define the truncation order $N^{(s)}$ as the minimum order at which the ratio of the n th-order ξ_n to zeroth-order ξ_0 falls below predetermined ε ,

$$N^{(s)} := \min \left\{ n \mid \frac{\xi_n}{\xi_0} \leq \varepsilon \right\}. \quad (23)$$

Figure 1 shows an example of the weight functions $\xi_{\text{uni},n}$, $\xi_{\text{Gs},n}$, $\xi_{\text{Gs-Inf},n}$, and $\xi_{\text{SWF},n}$ in the case where $k = 30 \text{ rad/m}$, $R^{(s)} = 0.2 \text{ m}$, and $\sigma = R^{(s)}/3 \text{ m}$. Setting ε to 10^{-5} , we observe that the truncation orders for each integral term are $N_{\text{uni}} = 11$, $N_{\text{Gs}} = 9$, $N_{\text{Gs-Inf}} = 11$, and $N_{\text{SWF}} = \infty$ from Fig. 1. This observation suggests that the choice of weight function affects the computational cost of the kernel function defined as Eq. (14). In addition, N_{SWF} must be determined arbitrarily.

As an additional example, the required orders $N^{(s)}$ for the kernel functions in Eq. (15) are plotted as a function of frequency in Fig. 2 for the case of $R^{(s)} = 0.2 \text{ m}$, $\sigma = R^{(s)}/3 \text{ m}$, and $\varepsilon = 10^{-5}$. The case of SWF was omitted here because the truncation order is arbitrary. Figure 2 demonstrates that the required orders vary depending on the type

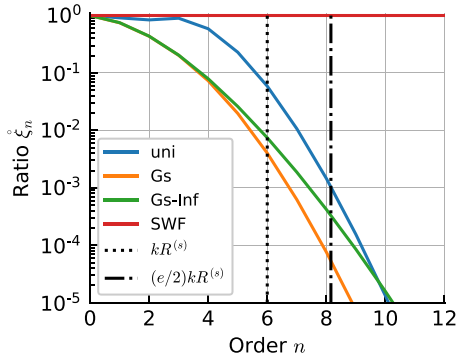


FIG. 1. An example of weight functions plotted against order in the case of $k = 30 \text{ rad/m}$, $R^{(s)} = 0.2 \text{ m}$, and $\sigma = R^{(s)}/3 \text{ m}$. The cases of weight functions $\xi_{\text{uni},n}$, $\xi_{\text{Gs},n}$, $\xi_{\text{Gs-Inf},n}$, and $\xi_{\text{SWF},n}$ are denoted as uni, Gs, Gs-Inf, and SWF, respectively.

of weight function selected and this difference becomes more pronounced with increasing frequency. In addition, we can observe that the required orders in the case of proposed weight $\xi_{\text{Gs},n}$ are smaller than those in the cases of other weights.

Furthermore, the required orders are plotted as a function of the radius of the source region $R^{(s)}$ in Fig. 3 for the case of $k = 30 \text{ rad/m}$, $\sigma = R^{(s)}/3 \text{ m}$, and $\varepsilon = 10^{-5}$. This example shows that the required order increases as $R^{(s)}$ increases. Among all of the cases using the kernel function, the case of proposed weight $\xi_{\text{Gs},n}$ shows the smallest order for all of the radii.

D. Hyperparameter optimization

The kernel function defined as Eq. (22) has multiple hyperparameters to be optimized: the number of source regions S , the location of origins $\{O^{(s)}\}_{s \in S}$, and the radius of the source regions $\{R^{(s)}\}_{s \in S}$. In contrast to Ref. 8, which employs plane-wave kernels and optimizes weights for a fixed set of plane waves, our approach uses a kernel based on point-source distributions, as introduced in Sec. III A. Because point sources behave as singularities at the source locations, proposed kernel function naturally introduces hyperparameters that must be tuned to accurately capture their spatial configuration and correlations in the case of reconstruction of the region containing the sound sources.

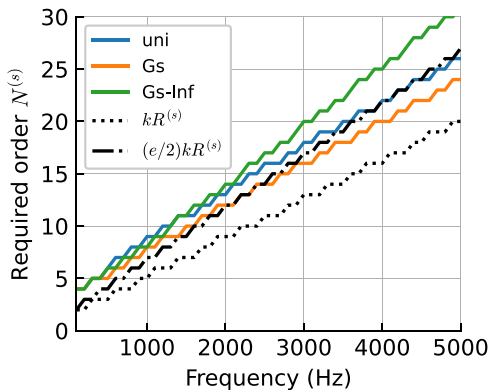


FIG. 2. The required order for kernel function plotted against frequency in the case of $R^{(s)} = 0.2 \text{ m}$, $\sigma = R^{(s)}/3 \text{ m}$, and $\varepsilon = 10^{-5}$.

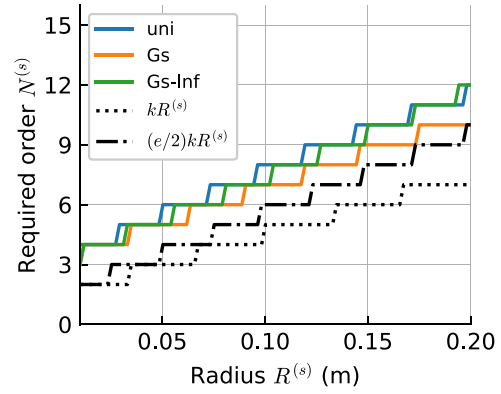


FIG. 3. The required order for kernel function plotted against the source region size $R^{(s)}$ in the case of $k = 30 \text{ rad/m}$, $\sigma = R^{(s)}/3 \text{ m}$, and $\varepsilon = 10^{-5}$.

First, the number of sound sources $Q \in \mathbb{Z}_+$ and their approximate locations are assumed to be known as prior information. It is not always necessary for the number of source regions S to correspond to the number of sound sources Q because nearby sound sources can be covered by a single source region $\Omega^{(s)}$, according to the definition of the kernel function expressed as Eq. (11). However, we assume that $S := Q$ for simplicity.

Second, hyperparameters $\{O^{(s)}\}_{s \in S}$, and $\{R^{(s)}\}_{s \in S}$ are optimized via the grid-search-based algorithm. The optimization aims to reduce the size of $\Omega^{(s)}$ by reducing $R^{(s)}$ and updating $O^{(s)}$ to be located near the sound source locations, thereby enabling more accurate sound field reconstruction in the vicinity of $\Omega^{(s)}$. Moreover, the required order decreases as $R^{(s)}$ decreases, which thereby reduces the computational cost (Fig. 3). These hyperparameters are initialized first such that each source region $\Omega^{(s)}$ constituted by $O^{(s)}$ and $R^{(s)}$ includes the corresponding sound source using prior information. In the grid search, the value that minimizes the objective function is selected through an exhaustive search over the predefined discrete sets of candidates $\mathcal{T}_{O^{(s)}}$ and $\mathcal{T}_{R^{(s)}}$ for each source region. The objective function uses leave-one-out cross-validation (LOO-CV),³⁸ which is expressed as

$$\mathcal{L}_{\text{LOO}} = \frac{1}{M} \sum_{m=1}^M \mathcal{L}(\tilde{u}(\mathbf{r}_m) - y_m), \quad (24)$$

where $\mathcal{L}(\cdot)$ denotes a criterion that measures the error, and $\tilde{u}(\mathbf{r}_m)$ denotes the estimated sound pressure at $\mathbf{r}_m$ of the partial GPR, excluding a set $\{\mathbf{r}_m, y_m\}$. The hyperparameter optimization scheme is summarized in Algorithm 1.

IV. CONVENTIONAL METHODS

Two conventional methods are briefly introduced in this section.

A. SWF expansion method

The SWF expansion method can interpolate or extrapolate the sound field using the coefficients expanded by the

ALGORITHM 1: Grid-search-based optimization algorithm for hyperparameters in a kernel function.

```

Initialize  $S$ ,  $\{O^{(s)}\}_{s \in S}$ , and  $\{R^{(s)}\}_{s \in S}$  using prior information.
Set  $\mathcal{T}_{R^{(s)}}$  and  $\mathcal{T}_{O^{(s)}}$  for each source region.
for  $s \in S$  do
    Initialize best score  $\mathcal{L}_{\text{LOO}}^{\text{best}} = \infty$  (or worst possible value).
     $\kappa'(\mathbf{r}_i, \mathbf{r}_j) \leftarrow \sum_{s' \in S \setminus \{s\}} \kappa(\mathbf{r}_i^{(s')}, \mathbf{r}_j^{(s')})$ 
    for  $O_{\text{cand}}^{(s)} \in \mathcal{T}_{O^{(s)}}$  do
        for  $R_{\text{cand}}^{(s)} \in \mathcal{T}_{R^{(s)}}$  do
             $\kappa(\mathbf{r}_i, \mathbf{r}_j) \leftarrow \kappa'(\mathbf{r}_i, \mathbf{r}_j) + \kappa(\mathbf{r}_i^{(s)}, \mathbf{r}_j^{(s)} | O_{\text{cand}}^{(s)}, R_{\text{cand}}^{(s)})$ 
             $\mathcal{L}_{\text{LOO}} = \frac{1}{M} \sum_{m=1}^M \mathcal{L}(\tilde{u}(\mathbf{r}_m) - y_m) // \text{Eq. (24)}$ 
            if  $\mathcal{L}_{\text{LOO}} < \mathcal{L}_{\text{LOO}}^{\text{best}}$  then
                 $\mathcal{L}_{\text{LOO}}^{\text{best}} \leftarrow \mathcal{L}_{\text{LOO}}$ 
                 $O^{(s)} \leftarrow O_{\text{cand}}$ 
                 $R^{(s)} \leftarrow R_{\text{cand}}$ 
            end
        end
    end
end

```

SWFs as bases.¹ The SWF expansion at the origin $O^{(s)}$ for the exterior problem²¹ is represented as

$$u(\mathbf{r}^{(s)}) \simeq \sum_{n=0}^{N^{(s)}} \sum_{m=-n}^n \beta_n^{m(s)} h_n(kr^{(s)}) Y_n^m(\theta^{(s)}, \phi^{(s)}) = \mathbf{h}_\bullet^{(s)\text{T}} \mathbf{b}^{(s)}, \quad (25)$$

where $\beta_n^{m(s)}$ denotes the coefficient of order n and degree m ; $\mathbf{b}^{(s)} \in \mathbb{C}^{V^{(s)}}$ denotes the vector of the coefficients with $[\mathbf{b}^{(s)}]_{n^2+n+m+1} = \beta_n^{m(s)}$; and $\mathbf{h}_\bullet^{(s)} \in \mathbb{C}^{V^{(s)}}$ is the vector of the basis function at the estimation position with $[\mathbf{h}_\bullet^{(s)}]_{n^2+n+m+1} = h_n(kr^{(s)}) Y_n^m(\theta^{(s)}, \phi^{(s)})$. Here, the infinite series is assumed to be truncated at $N^{(s)} \in \mathbb{Z}_+$. Similarly, measured sound pressures at the microphones can be expressed as

$$\mathbf{y} = \mathbf{\Psi}^{(s)} \mathbf{b}^{(s)} + \mathbf{e}, \quad (26)$$

where $\mathbf{\Psi}^{(s)} = [\mathbf{h}_1^{(s)}, \dots, \mathbf{h}_M^{(s)}]^\text{T} \in \mathbb{C}^{M \times V^{(s)}}$ denotes the basis matrix. First, the expansion coefficients can be obtained from the sound pressures observed by microphones by solving the following least squares problem with l_2 regularization such that

$$\underset{\mathbf{b}^{(s)} \in \mathbb{C}^{V^{(s)}}}{\text{minimize}} \mathcal{J}(\mathbf{b}^{(s)}) = \|\mathbf{y} - \mathbf{\Psi}^{(s)} \mathbf{b}^{(s)}\|_2^2 + \lambda \|\mathbf{b}^{(s)}\|_2^2, \quad (27)$$

where $\lambda \in \mathbb{Z}_+$ denotes the regularization parameter. The solution that minimizes the objective function can be obtained by solving $(\partial/\partial \mathbf{b}^{(s)\text{H}}) \mathcal{J}(\mathbf{b}^{(s)}) = \mathbf{0}$ and can be expressed as

$$\hat{\mathbf{b}}^{(s)} = (\mathbf{\Psi}^{(s)\text{H}} \mathbf{\Psi}^{(s)} + \lambda \mathbf{I}_{V^{(s)}})^{-1} \mathbf{\Psi}^{(s)\text{H}} \mathbf{y}. \quad (28)$$

Substituting Eq. (28) into Eq. (25), we obtain the estimated sound pressure at arbitrary position $\mathbf{r}_\bullet^{(s)}$ as

$$\hat{u}(\mathbf{r}_\bullet^{(s)}) = \mathbf{h}_\bullet^{(s)\text{T}} (\mathbf{\Psi}^{(s)\text{H}} \mathbf{\Psi}^{(s)} + \lambda \mathbf{I}_{V^{(s)}})^{-1} \mathbf{\Psi}^{(s)\text{H}} \mathbf{y}. \quad (29)$$

In addition, the sound field can also be described by superposition of SWF expansions at multiple origins in cases in which multiple sound sources exist, as shown in

$$u(\mathbf{r}) \simeq \sum_{s=1}^S \sum_{n=0}^{N^{(s)}} \sum_{m=-n}^n \beta_n^{m(s)} h_n(kr^{(s)}) Y_n^m(\theta^{(s)}, \phi^{(s)}) = \mathbf{h}_\bullet^\text{T} \mathbf{b}, \quad (30)$$

where $\mathbf{h} = [\mathbf{h}_\bullet^{(1)\text{T}}, \mathbf{h}_\bullet^{(2)\text{T}}, \dots, \mathbf{h}_\bullet^{(S)\text{T}}]^\text{T} \in \mathbb{C}^V$, and $\mathbf{b} = [\mathbf{b}^{(1)\text{T}}, \mathbf{b}^{(2)\text{T}}, \dots, \mathbf{b}^{(S)\text{T}}]^\text{T} \in \mathbb{C}^V$, with the total modes $V := \sum_{s=1}^S V^{(s)}$. The measured sound pressure can be formulated as

$$\mathbf{y} = \mathbf{\Psi} \mathbf{b} + \mathbf{e}, \quad (31)$$

where $\mathbf{\Psi} = [\mathbf{\Psi}^{(1)}, \mathbf{\Psi}^{(2)}, \dots, \mathbf{\Psi}^{(S)}] \in \mathbb{C}^{M \times V}$. Then, the estimation of SWF coefficients can be formulated as

$$\underset{\mathbf{b} \in \mathbb{C}^V}{\text{minimize}} \mathcal{J}(\mathbf{b}) = \|\mathbf{y} - \mathbf{\Psi} \mathbf{b}\|_2^2 + \lambda \|\mathbf{b}\|_2^2. \quad (32)$$

Again, the solution can be obtained by solving $(\partial/\partial \mathbf{b}^\text{H}) \mathcal{J}(\mathbf{b}) = \mathbf{0}$ and expressed as

$$\hat{\mathbf{b}} = (\mathbf{\Psi}^\text{H} \mathbf{\Psi} + \lambda \mathbf{I}_V)^{-1} \mathbf{\Psi}^\text{H} \mathbf{y}. \quad (33)$$

The sound pressure at an arbitrary position $\mathbf{r}$ can then be obtained by substituting Eq. (33) into Eq. (30) such that

$$\hat{u}(\mathbf{r}_\bullet) = \mathbf{h}_\bullet^\text{T} (\mathbf{\Psi}^\text{H} \mathbf{\Psi} + \lambda \mathbf{I}_V)^{-1} \mathbf{\Psi}^\text{H} \mathbf{y}. \quad (34)$$

The estimation process based on SWF expansion can be interpreted as a GPR with the kernel function

$$\kappa(\mathbf{r}_i^{(s)}, \mathbf{r}_j^{(s)}) = \begin{cases} \mathbf{h}_j^{(s)\text{H}} \mathbf{h}_i^{(s)} & \text{for single origin,} \\ \sum_{s=1}^S \mathbf{h}_j^{(s)\text{H}} \mathbf{h}_i^{(s)} & \text{for multiple origins.} \end{cases} \quad (35)$$

The derivation is detailed in the [Appendix](#).

B. Sparse decomposition method

Sparse sound field decomposition takes advantage of the property that the sound field can be approximated as a superposition of a small number of plane waves or point sources.^{12–14,16,20} In particular, in nonhomogeneous sound fields, where there is no reflected sound in the target area, the sound field can be described by a superposition of point sources. This superposition can be formulated as

$$u(\mathbf{r}) \simeq \sum_{l=1}^L a(\mathbf{r}_l) G(\mathbf{r}_l | \mathbf{r}), \quad (36)$$

where L denotes the number of point sources, $a(\mathbf{r}_l)$ denotes the amplitude of the point source located at $\mathbf{r}_l$, and $G(\mathbf{r}_l | \mathbf{r})$ denotes Green's function²¹ in spatial domain, expressed as

$$G(\mathbf{r}_l|\mathbf{r}) = \frac{\exp(jk\|\mathbf{r}_l - \mathbf{r}\|_2)}{4\pi\|\mathbf{r}_l - \mathbf{r}\|_2}. \quad (37)$$

The measured sound pressures at the microphones can be represented in a matrix form as

$$\mathbf{y} \simeq \mathbf{G}\mathbf{a}, \quad (38)$$

where the elements of $\mathbf{G} \in \mathbb{C}^{M \times L}$ and $\mathbf{a} \in \mathbb{C}^L$ are given by $[\mathbf{G}]_{m,l} = G(\mathbf{r}_m|\mathbf{r}_l)$ and $[\mathbf{a}]_l = a(\mathbf{r}_l)$, respectively. Here, the candidates of source positions $\{\mathbf{r}_l\}_{l \in \mathcal{S}_L}$, where $\mathcal{S}_L := \{1, \dots, L\}$, are assumed to be discretized in the target region beforehand. In general, setting candidates such that $L \gg M$ to cover the entire target region results in an underdetermined system; however, this problem can be solved by imposing a sparsity constraint, which ensures that only a small number of candidates are activated. The optimization problem can be formulated as

$$\underset{\mathbf{a} \in \mathbb{C}^L}{\text{minimize}} \mathcal{J}(\mathbf{x}) = \|\mathbf{y} - \mathbf{G}\mathbf{a}\|_2^2 + \gamma\varphi_p(\mathbf{a}), \quad (39)$$

where γ represents the regularization parameter, and $\varphi_p(\mathbf{a})$ denotes an l_p -norm for inducing the sparsity with

$$\varphi_p(\mathbf{a}) := \begin{cases} \mathcal{I}(\mathbf{a}), & p = 0, \\ \left(\sum_{l=1}^L |[\mathbf{a}]_l|^p \right)^{1/p}, & 0 < p \leq 1, \end{cases} \quad (40)$$

where $\mathcal{I}(\cdot)$ denotes a function that returns the number of nonzero elements.¹³

The objective function in Eq. (39) can be solved using orthogonal matching pursuit (OMP)^{22,23} for $p = 0$, the focal underdetermined system solver (FOCUSS)²⁴ for $0 < p < 1$, and either the fast iterative shrinkage-thresholding algorithm (FISTA)^{25,26} or the alternating direction method of multipliers (ADMM)^{27,28} for $p = 1$.

V. NUMERICAL EXPERIMENTS

A. Experimental conditions

Numerical experiments of sound field estimation were conducted to evaluate the proposed method's performance. As conventional methods, the SWF expansion method,^{21,39} sparse decomposition method,^{12–14,16,20} and SHD-RGF²⁹ were used for comparison. Hereafter, the proposed method, SWF expansion method, sparse decomposition method, and SHD-RGF are denoted as **Proposed**, **SWF**, **Sparse**, and **SHD-RGF**, respectively.

The anechoic condition was assumed here, and the sound speed was set to 340 m/s. The target region to be estimated was set to a sphere region Ω with a radius of 1.0 m. Microphones were distributed on the sphere surface $\partial\Omega$ with $M = 64$ using spherical t -design.⁴⁰ The true sound field $u_{\text{true}}^{(1)}$ was composed of two point sources located at $\mathbf{r}_{\text{src}}^{(1)} = (0.72, \pi/2, 0)$ and $\mathbf{r}_{\text{src}}^{(2)} = (0.57, \pi/2, 2\pi/3)$ in spherical coordinates, with amplitudes $a_{\text{src}}^{(1)} = 1$ and $a_{\text{src}}^{(2)} = 0.5j$, respectively, and expressed using Green's function in Eq. (37) as

$$u_{\text{true}}(\mathbf{r}) = \sum_{q=1}^Q a_{\text{src}}^{(q)} G(\mathbf{r}_{\text{src}}^{(q)}|\mathbf{r}), \quad (41)$$

where $Q = 2$ denotes the number of point sources. The real part of the true sound field at 2500 Hz in the xy -plane is plotted in Fig. 4, where the black circle expresses $\partial\Omega$. The measured sound pressure at each microphone was generated by

$$u(\mathbf{r}_m) = \sum_{q=1}^Q a_{\text{src}}^{(q)} G(\mathbf{r}_{\text{src}}^{(q)}|\mathbf{r}_m) + e_m, \quad (42)$$

where the noise term e_m was independently sampled from a zero-mean complex Gaussian distribution and added to the signals observed by all microphones to achieve a signal-to-noise ratio (SNR) of 20 dB.

Proposed employed GPR using the proposed kernel functions [Eq. (15)] associated with three different weight functions: ξ_{uni} , ξ_{Gs} , and $\xi_{\text{Gs-Inf}}$. The assumed prior information was that (1) the approximate sphere regions containing the sound sources, denoted as $\{\Omega_{\text{src}}^{(s)}\}_{s \in \mathcal{S}}$, were given as shown in Fig. 5, and (2) $O^{(s)}$ and $R^{(s)}$ were initialized using the center $O_{\text{src}}^{(s)}$ and radius $R_{\text{src}}^{(s)}$ of the approximate region $\Omega_{\text{src}}^{(s)}$ for each s . We set $O_{\text{src}}^{(1)} = (0.6, \pi/2, \pi/24)$, $O_{\text{src}}^{(2)} = (0.55, 35\pi/72, 7\pi/12)$, and $R_{\text{src}}^{(1)} = R_{\text{src}}^{(2)} = 0.2$ here. These hyperparameters were optimized by the grid search expressed in Algorithm 1. Here, the loss function in Eq. (24) was set to $\mathcal{L}(\cdot) = \|\cdot\|_2^2$, the update candidates for $\{O^{(s)}\}_{s \in \mathcal{S}}$ were set to $\mathcal{T}_{O^{(s)}} = \mathcal{D}_{\Omega_{\text{src}}^{(s)}}(0.05)$, where $\mathcal{D}_{\Omega_{\text{src}}^{(s)}}(\Delta)$ represents the set of points obtained by discretizing the region $\Omega_{\text{src}}^{(s)}$ into a three-dimensional lattice at intervals of Δ , and the update candidates for $\{R^{(s)}\}_{s \in \mathcal{S}}$ were set to $\mathcal{T}_{R^{(s)}} = \{0.05, 0.1, 0.15, 0.2\}$. $N^{(s)}$ was set using Eq. (23) with $\varepsilon = 10^{-5}$ in each update step. The estimated sound field was obtained by Eq. (10a). The regularization parameter σ_e^2 in Eq. (10a) was set to $\sigma_e^2 = \sigma_y^2 / (10^{\text{SNR}/10} + 1)$, where σ_y denotes the variance of measured signals $\mathbf{y}$ under the assumptions that the SNR was known and the noise-free signal $\mathbf{u}$ and noise components $\mathbf{e}$ are uncorrelated. Hereafter, **Proposed** using initialized hyperparameters with each weight function ξ_{uni} , ξ_{Gs} , and $\xi_{\text{Gs-Inf}}$ are denoted as **Proposed (uni, Init)**, **Proposed (Gs, Init)**, and **Proposed (Gs-Inf, Init)**, respectively. **Proposed** using optimized hyperparameters with each weight function are also denoted as **Proposed (uni, Opt)**, **Proposed (Gs, Opt)**, and **Proposed (Gs-Inf, Opt)**, respectively.

GPR with the kernel function defined in Eq. (35) (for multiple origins) was employed as **SWF**. The initialization and optimization scheme were the same as those for **Proposed**. The truncation order $N^{(s)}$ of the kernel was set to $N^{(s)} = \lceil kR^{(s)} \rceil$ in each optimization step. Hereafter, **SWF** using initialized and optimized hyperparameters are denoted as **SWF (Init)**, and **SWF (Opt)**, respectively.

In **Sparse**, two approaches for obtaining the candidates of active source positions, $\{\mathbf{r}_l\}_{l \in \mathcal{S}_L}$, were used. First, the candidates were obtained by discretizing whole Ω by the grids with an interval of 0.1 m [i.e., $\{\mathbf{r}_l\}_{l \in \mathcal{S}_L} = \mathcal{D}_\Omega(0.1)$]. This discretization results in the number of candidates being $L = 4169$. Second, the candidates were obtained using prior information about source positions, as in the case of **Proposed**. In this case, which is denoted as **Sparse (SI)** hereafter, the candidates were set to $\{\mathbf{r}_l\}_{l \in \mathcal{S}_L} = \mathcal{D}_{\{\Omega_{\text{src}}^{(s)}\}_{\mathcal{S}}}(0.025)$ such that the resultant number of candidates $L = 4218$ approached that for **Sparse**. As the sparse optimization, OMP^{22,23} was used to solve the objective function defined by Eq. (39) under the assumption of the $p = 0$ case. The implementation of OMP followed Algorithm 3 (*OMP for Signal Recovery*) in Ref. 23. The sound field was then reconstructed using Eq. (36).

In **SHD-RGF**, the positions and amplitudes of the sound sources were estimated using the algorithm described in Sec. 4.2 of Ref. 29, under the assumption that the number of sound sources is known. Similar to the **Sparse** case, the sound field was reconstructed by substituting the estimated amplitudes and positions of the sound sources into Eq. (36). In the algorithm, the order of SWF was chosen from $\mathbb{Z}_+([1, 7])$ such that it yielded the highest estimation accuracy [i.e., the lowest normalized mean squared error (NMSE)], as introduced in Eq. (43).

B. Evaluation measure

As an evaluation measure, the NMSE was defined to evaluate the accuracy of sound field estimation as^{9,20,30,31}

$$\text{NMSE} := 10 \log_{10} \frac{\int_{\Omega} |u_{\text{true}}(\mathbf{r}) - \hat{u}(\mathbf{r})|^2 d\mathbf{r}}{\int_{\Omega} |u_{\text{true}}(\mathbf{r})|^2 d\mathbf{r}} \quad (\text{dB}), \quad (43)$$

where the evaluation points were sampled randomly from a uniform distribution in Ω . The number of samples was set to 500. In addition, the root mean square error (RMSE) was defined to measure the distance between the true source position and estimated source position in Cartesian coordinates as follows:²⁹

$$\text{RMSE} := \sqrt{\frac{1}{Q} \sum_{q=1}^Q \|\mathbf{r}_{\text{src}}^{(q)} - \hat{\mathbf{r}}^{(q)}\|_2} \quad (\text{m}). \quad (44)$$

In **Proposed**, the centers $\{\Omega^{(s)}\}$ were used as the estimated source positions to evaluate the RMSE. Furthermore, the normalized squared error (NSE) was defined to visualize the error distribution of the estimated sound field as³¹

$$\text{NSE}(\mathbf{r}) := 10 \log_{10} \frac{|u_{\text{true}}(\mathbf{r}) - \hat{u}(\mathbf{r})|^2}{|u_{\text{true}}(\mathbf{r})|^2} \quad (\text{dB}). \quad (45)$$

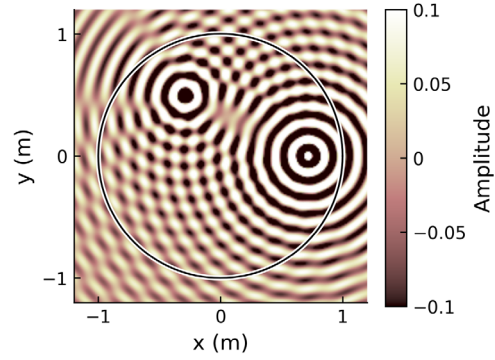


FIG. 4. The real part of the true sound field at 2500 Hz plotted in the xy -plane. Two point sources are located at $(0.72, \pi/2, 0)$ and $(0.57, \pi/2, 2\pi/3)$ in spherical coordinates.

C. Results

The estimated sound fields and NSE distributions at 2500 Hz are shown in Figs. 6 and 7, respectively. The estimation accuracies of the GPR-based methods (i.e., **Proposed** and **SWF**) were improved through the grid-search optimization, as revealed by comparing Figs. 7(a)–7(d) with Figs. 7(e)–7(h). However, **SWF** shows a higher error than **Proposed** after optimization, likely because the optimization did not perform as well as **Proposed**. The estimation errors of **Sparse (SI)** were lower than those of **Sparse**, suggesting that the prior information helps sparse optimization when compared at approximately the same computational cost [Figs. 7(a) and 7(b)]. Notably, a different trend is observed between **Proposed** and **Sparse**: **Proposed** can suppress the errors over the entire region Ω but show relatively high errors inside the source regions $\{\Omega_{\text{src}}^{(s)}\}_{\mathcal{S}}$ [shown as black dark spots in Figs. 7(a)–7(c)], whereas **Sparse** does not show such high-error spots near the source positions but shows relatively high estimation errors in the entire target

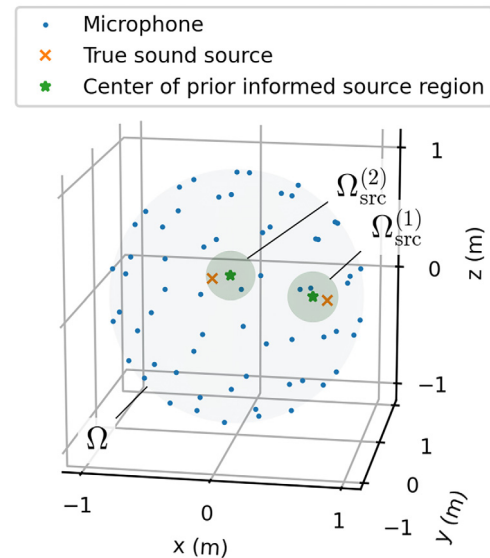


FIG. 5. Geometry of the experiments system; the blue sphere represents the target region Ω , and each green sphere represents the source-included region $\Omega_{\text{src}}^{(s)}$, given as prior information.

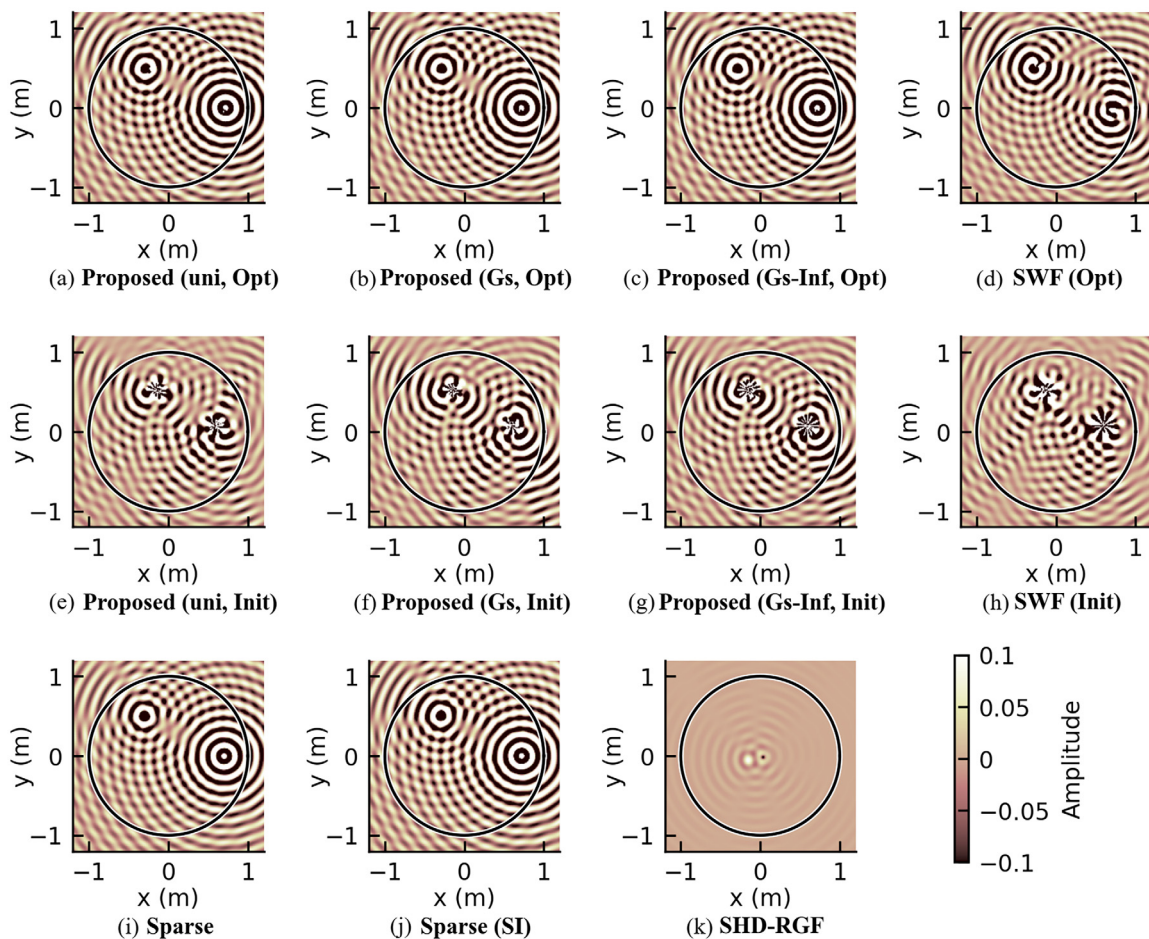


FIG. 6. Estimated sound fields at 2500 Hz plotted in the xy -plane.

region Ω . Figures 6(k) and 7(k) show that **SHD-RGF** cannot localize the positions of sound sources; thus, the estimation accuracy was lower in all of the other investigated methods. This inability to localize the positions of sound sources might be attributable to the microphones being located on a single layer and the region size being relatively large with respect to frequency.

Figures 8 and 9 show the NMSEs and RMSEs, respectively, at intervals of 100 Hz in the range from 100 to 5000 Hz. Hereafter, the general notations of **Proposed** with and without optimization are **Proposed** (*, **Opt**) and **Proposed** (*, **Init**), respectively. First, the NMSEs decreased through optimization, as revealed by comparing **Proposed** (*, **Opt**) and **Proposed** (*, **Init**) for all of the frequencies in Fig. 8. This result suggests that grid-search-based optimization is effective. The suggestion is also supported by the RMSE results in Fig. 9, where the RMSEs of **Proposed** (*, **Opt**) are lower than those of **Proposed** (*, **Init**) for all of the investigated frequencies. By contrast, the NMSEs of **SWF** (**Opt**) increased above 2000 Hz, even though the optimization scheme was the same as that **Proposed** (*, **Opt**). The difference between **Proposed** (*, **Opt**) and **SWF** (**Opt**) is attributed to the weight function (Fig. 1): **Proposed** weights the significant orders as a result of setting the source-included region in Eq. (11), whereas

SWF does not restrict any region that results in weighting the same weight even for insignificant higher orders. Compared with the conventional methods, **Proposed** (*, **Opt**) achieved the lowest NMSEs among all frequencies except for those below 500 Hz. Compared with **Proposed** (**uni**, **Opt**) and **Proposed** (**Gs-Inf**, **Opt**), **Proposed** (**Gs**, **Opt**) achieved slightly lower NMSEs across almost all frequencies despite calculating up to the smallest order for kernel function and, thus, the lowest computational cost among **Proposed** (*, **Opt**) (Fig. 2). **Sparse** (**SI**) shows the lower NMSEs compared with **Sparse** among all of the investigated frequencies because the discretization of candidates was denser in **Sparse** (**SI**) than in **Sparse**. In addition, **Sparse** (**SI**) shows the lowest NMSEs at frequencies below 500 Hz, possibly because the wavelengths are much larger than the distances between the positions of actual sources and activated candidates for the frequencies. Although **Sparse** (**SI**) shows smaller RMSEs than **Proposed** (*, **Opt**) for all of the investigated frequencies (Fig. 9), the NMSEs with **Proposed** (*, **Opt**) were lower for frequencies above 500 Hz, as mentioned above. This result demonstrates that **Proposed** (*, **Opt**) can estimate the sound field of the target region with higher accuracy than the other methods, provided that the sound source region is set to include the source without the need for precise source locations.

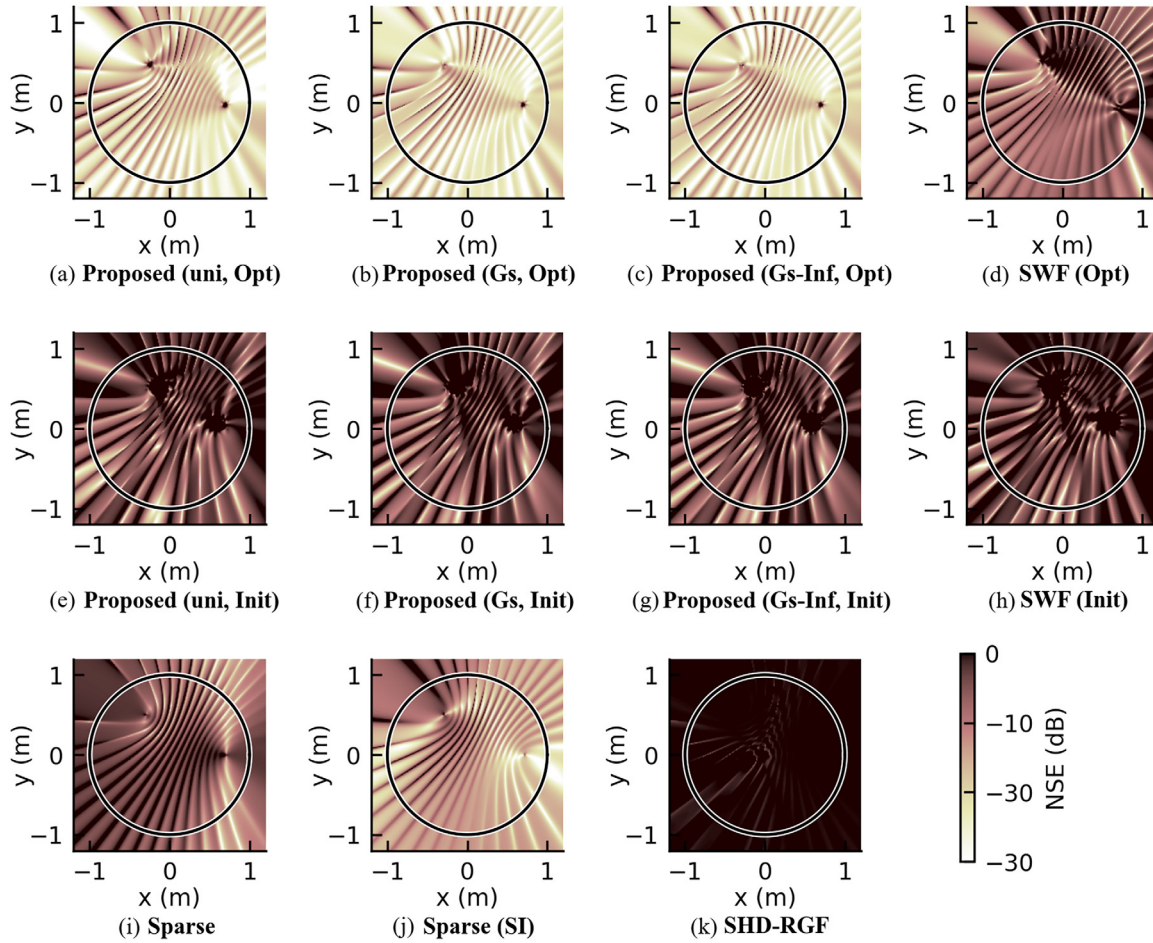


FIG. 7. NSE distribution at 2500 Hz plotted in the xy -plane. The NMSEs are (a) -18.7 dB, (b) -21.7 dB, (c) -21.3 dB, (d) -6.0 dB, (e) 19.0 dB, (f) 7.8 dB, (g) 47.9 dB, (h) 27.6 dB, (i) -5.5 dB, (j) -14.8 dB, and (k) 0.5 dB.

VI. DISCUSSION

A. Effects of discretization

In this paper, we demonstrate experimentally that the proposed method does not require the precise location of sound sources; as long as the number of sources and their approximate positions are known, the hyperparameters of the kernel function can be optimized using grid search. Compared with other methods, the numerical experiments reveal that this approach achieves higher sound field estimation accuracy across a broad frequency range.

In the grid-search-based proposed method, it is necessary to discretize the prior informed region in which sound sources may exist to determine candidate central positions for the source distribution. Unlike the sparse decomposition approach, the proposed method considers source positions as regions rather than individual points. Therefore, instead of precisely determining the exact source location, it is sufficient to optimize a region that fully encompasses the source, which constitutes a major advantage of the proposed method. In contrast, sparse decomposition methods inherently assume that the sound source is located at one of the discrete grid points. This can lead to positional errors if the actual source does not

coincide exactly with a grid point, resulting in estimation errors that propagate throughout the reconstructed sound field.

B. Prior information

We employed some prior knowledge of sound sources in the experiment; however, several approaches can be used to obtain the prior information. One approach is to use other acoustic signal processing methods, such as sparse sound field decomposition introduced in Sec. IV B, to determine the number of sources and their approximate locations. In our numerical experiments, in sparse decomposition, the estimation accuracy of source positions at certain frequencies is extremely poor (Fig. 9). However, more robust estimation can be achieved by introducing constraints in which source positions remain consistent across multiple time frames or frequencies.¹² In addition, source positions can be estimated in the angular-space domain as described in Ref. 41 to obtain the source information. Alternatively, using multiple distributed microphone arrays rather than a single layer array allows for source-position estimation using beamforming-based methods.⁴²

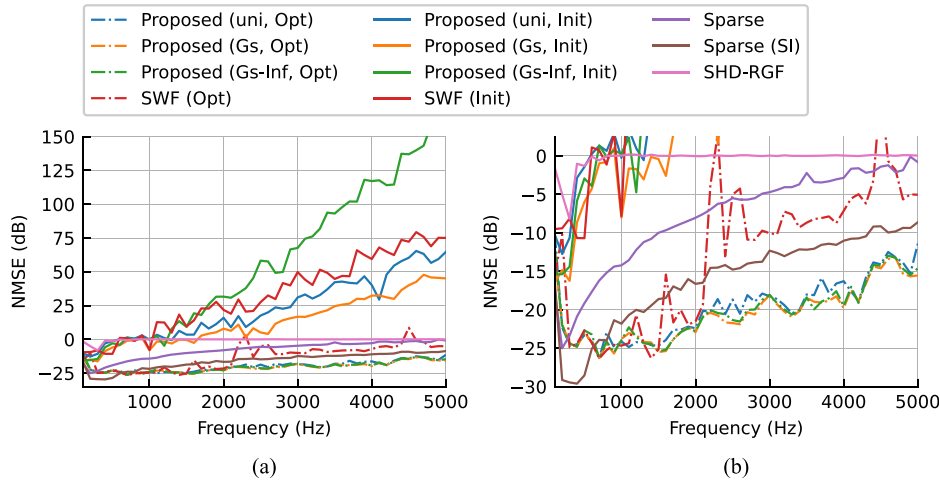


FIG. 8. NMSE (dB) against frequency plotted in (a) a wide NMSE range (from -30 to 150 dB) and (b) a narrow NMSE range (from -30 to 2.5 dB).

Another approach could involve using cameras or other sensors. For example, objects can be detected through image recognition with cameras, or they can be captured on the basis of depth with infrared cameras such as Kinect (Microsoft, Redmond, WA) cameras.⁴³ Then, the number and positions of these detected objects can be used as prior information to set the source region, enabling the proposed method to be used.

C. Kernel function updates

Even when approximate source information (e.g., rough location) is known in advance, if the source position changes, it may be necessary to re-optimize the hyperparameters of the kernel function and update the FIR filters. This introduces the potential need for additional computations. Here, we examine this issue by comparing the proposed method [**Proposed (Gs, Opt)**] with the sparse decomposition method (**Sparse**).

When the source positions are time-invariant, the sparse decomposition method requires optimization for each new input signal, whereas the proposed method can employ a common set of FIR filters without the need for re-optimization. In this case, the computational cost of the proposed method and sparse decomposition method for each frequency bin are $\mathcal{O}(EM)$ and $\mathcal{O}(EQ + QML)$, respectively.⁹ Here, $E \in \mathbb{Z}_+$ is the number of estimation points. Therefore, the proposed method can be considered to be more efficient than the sparse decomposition method, especially in the case in which E is small.

Next, we investigated the case where the sources move slightly. Specifically, we simulated minor source position variations as adding multivariate Gaussian with standard deviations of one-third of $\{0.05, 0.1, 0.15, 0.2\}$ m to each dimension (x, y, z) of the source positions used in the experiments of Sec. V. The source amplitude, microphone arrangement, and SNR are the same as those used in Sec. V. Figure 10 presents the results of sound field estimation accuracy evaluated by the NMSE averaged over 100 trials. For the proposed method, we used the hyperparameters obtained through the optimization in Sec. V without modification,

thus, incurring no additional computational cost beyond FIR filtering. In contrast, the sparse decomposition method recomputed the optimization for each observed signal to estimate the sound field.

As can be observed from Fig. 10, while the estimation accuracy of the proposed method decreases as the distance of source positions variation increase, it still achieves superior performance compared to the sparse decomposition method. In other words, for minor fluctuations in source positions, it is possible to continue using the FIR filters obtained from the initial optimization without updating.

Yet, if the source positions change significantly, the proposed method, like other methods, would need to perform the grid-search-based optimization again, incurring additional computational costs. Developing optimization and FIR filter update procedures suitable for real-time implementation remains an important topic for future work.

D. Reverberant condition

In the present paper, we assume an anechoic environment (i.e., a setting without incidental reflections). However, it is necessary to also consider reverberant environments in some practical situations. In such environments, the sound pressure at any position can be represented as the superposition of sound pressures produced by the sources within the target region and those incoming from outside the target region. In this context, it is necessary to redefine the kernel function when considering sound field estimation using GPR. For instance, a newly defined kernel function that is the sum of the proposed kernel $\kappa_{\text{prop}}(\mathbf{r}_i, \mathbf{r}_j)$ and a kernel corresponding to the incident sound field $\kappa_{\text{inc}}(\mathbf{r}_i, \mathbf{r}_j)$,

$$\kappa_{\text{mix}}(\mathbf{r}_i, \mathbf{r}_j) := \kappa_{\text{prop}}(\mathbf{r}_i, \mathbf{r}_j) + \kappa_{\text{inc}}(\mathbf{r}_i, \mathbf{r}_j), \quad (46)$$

could be used to achieve sound field estimation. The kernel functions corresponding to the incident sound field can be one corresponding to diffuse sound fields^{7,10} or one

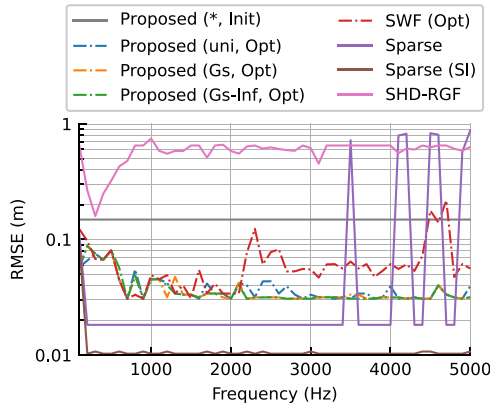


FIG. 9. RMSE (m) against frequency. Because the RMSEs of **Proposed (uni, Init)**, **Proposed (Gs, Init)**, and **Proposed (Gs-Inf, Init)** are identical, they are collectively referred to here as **Proposed (*, Init)**.

considering the room characteristics.^{9,30} In a future study, we will consider the case where reflections exist.

E. Multiple sources in a source region

The present paper also considers a scenario in which each region contains a single sound source; nevertheless, the definition of the kernel function does not limit the number of sources within a region, as shown in Eq. (11). Therefore, when closely spaced sources exist, treating them collectively as a single source region and designing the kernel function accordingly can reduce computational costs. Notably, however, this approach might enable the design of a sphere region with a relatively large radius, which could potentially decrease the estimation accuracy of the near sound field around the source region.

Even sources have directivities, and directional sources can be well approximated by the superposition of point sources.²¹ Therefore, they can be approximated by distributed point sources over the boundary surface surrounding the

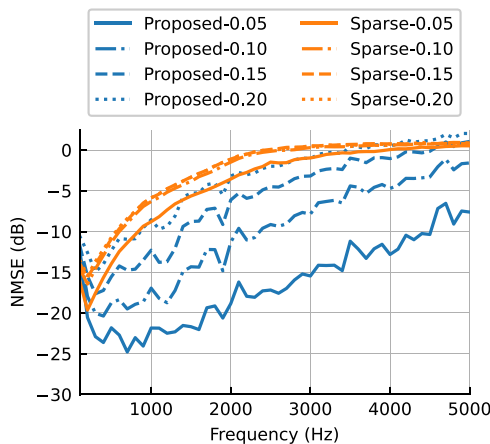


FIG. 10. NMSE (dB) against frequency in the case where the source positions move slightly. Proposed shows the results of **Proposed (Gs, Opt)** without re-optimization while sparse shows the results of **Sparse** with re-optimization. One-third of the numbers displayed after each method represent the standard deviations (m) of the multivariate Gaussian distribution used to simulate minor variations in the source positions.

source or within the bounded area based on the equivalent source method.^{44,45} Moreover, they can also be approximated on the basis of multipole expansion by regularly arranging point sources at close positions.⁴⁶ Thus, conceivably, the proposed method can perform sound field estimation by configuring the source region to encompass directional sources. These details are left as subjects for future research.

VII. CONCLUSION

GP for an inhomogeneous sound field was proposed. Kernel functions were formulated as the weighted spatial correlation derived by the existence probability distribution of sound sources. A weight for the kernel function derived by rigorous integration using a Gaussian distribution, the scheme of order truncation in the kernel function and hyperparameter optimization were proposed. Numerical experiments showed not only that the proposed GPR achieved greater estimation accuracy than the conventional methods but also that the kernel function with the proposed weight function achieved a higher estimation accuracy for a wide frequency range with lower computational cost compared with the other kernels. Furthermore, an additional advantage of the proposed GP is that it treats the sound source as a distribution rather than a single point. Consequently, when optimizing the center position of the source distribution, the discretized candidate positions need not perfectly align with the actual source location. In addition, if the source positions change slightly over time, the once-optimized FIR filters can be used without re-optimization.

ACKNOWLEDGMENTS

This research was partially supported by JSPS Grant No. JP22H00523.

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request.

APPENDIX A: PROOF OF THE VALIDITY OF THE PROPOSED KERNEL FUNCTION

If the Gram matrix $\mathbf{K}$ formed by the kernel function in Eq. (15) is a positive semi-definite matrix, then the kernel function is valid.³⁴ We demonstrate that the Gram matrix is a positive semi-definite matrix by showing that the condition

$$\mathbf{c}^H \mathbf{K} \mathbf{c} \geq 0 \quad (\text{A1})$$

holds for any arbitrary vector $\mathbf{c} \in \mathbb{C}^M$. The left-hand side of Eq. (A1) yields

$$\begin{aligned} \mathbf{c}^H \mathbf{K} \mathbf{c} &= \sum_{i=1}^M \sum_{j=1}^M c_i \bar{c}_j \kappa(\mathbf{r}_i, \mathbf{r}_j) \\ &= \sum_{i=1}^M \sum_{j=1}^M c_i \bar{c}_j \mathbf{h}_j^H \mathbf{\Xi} \mathbf{h}_i \\ &= \left(\sum_{j=1}^M c_j \mathbf{\Xi}^{1/2} \mathbf{h}_j \right)^H \left(\sum_{i=1}^M c_i \mathbf{\Xi}^{1/2} \mathbf{h}_i \right) \\ &= \left| \sum_{i=1}^M c_i \mathbf{\Xi}^{1/2} \mathbf{h}_i \right|^2 \geq 0. \end{aligned} \quad (\text{A2})$$

In the derivation of the third line, we used the fact that each element of $\mathbf{\Xi}$ is nonnegative, which is demonstrated by the definition of ξ_n in Eq. (14) and the conditions $k^2 \geq 0$, $j_n(kr^{(s)})^2 \geq 0$, $\eta(r^{(s)}) \geq 0$, and $r^{(s)^2} \geq 0$.

APPENDIX B: SOUND FIELD ESTIMATION BASED ON SWF EXPANSION AS GPR

First, SWF expansion at single origin is considered. When the matrix inversion lemma is used,⁴⁷ Eq. (29) yields

$$\begin{aligned} \hat{u}(\mathbf{r}^{(s)}) &= \mathbf{h}_\bullet^{(s)T} (\mathbf{\Psi}^{(s)H} \mathbf{\Psi}^{(s)} + \lambda \mathbf{I}_{V(s)})^{-1} \mathbf{\Psi}^{(s)H} \mathbf{y} \\ &= \mathbf{h}_\bullet^{(s)T} \mathbf{\Psi}^{(s)H} (\mathbf{\Psi}^{(s)} \mathbf{\Psi}^{(s)H} + \lambda \mathbf{I}_M)^{-1} \mathbf{y}, \end{aligned} \quad (\text{B1})$$

where

$$\begin{aligned} \mathbf{\Psi}^{(s)} \mathbf{\Psi}^{(s)H} &= \begin{bmatrix} \mathbf{h}_1^{(s)T} \\ \vdots \\ \mathbf{h}_M^{(s)T} \end{bmatrix} \begin{bmatrix} \overline{\mathbf{h}_1^{(s)}} & \dots & \overline{\mathbf{h}_M^{(s)}} \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{h}_1^{(s)H} \mathbf{h}_1^{(s)} & \dots & \mathbf{h}_M^{(s)H} \mathbf{h}_1^{(s)} \\ \vdots & \ddots & \vdots \\ \mathbf{h}_1^{(s)H} \mathbf{h}_M^{(s)} & \dots & \mathbf{h}_M^{(s)H} \mathbf{h}_M^{(s)} \end{bmatrix} \end{aligned} \quad (\text{B2})$$

and

$$\begin{aligned} \mathbf{h}_\bullet^{(s)T} \mathbf{\Psi}^{(s)H} &= \begin{bmatrix} \mathbf{h}_1^{(s)H} \mathbf{h}_\bullet^{(s)} & \dots & \mathbf{h}_M^{(s)H} \mathbf{h}_\bullet^{(s)} \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{h}_\bullet^{(s)H} \mathbf{h}_1^{(s)} \\ \vdots \\ \mathbf{h}_\bullet^{(s)H} \mathbf{h}_M^{(s)} \end{bmatrix}^H. \end{aligned} \quad (\text{B3})$$

Assuming that the kernel function is defined as

$$\kappa(\mathbf{r}_i^{(s)}, \mathbf{r}_j^{(s)}) = \mathbf{h}_j^{(s)H} \mathbf{h}_i^{(s)}, \quad (\text{B4})$$

Eqs. (B2) and (B3) can be, respectively, expressed as

$$\mathbf{\Psi}^{(s)} \mathbf{\Psi}^{(s)H} = \begin{bmatrix} \kappa(\mathbf{r}_1^{(s)}, \mathbf{r}_1^{(s)}) & \dots & \kappa(\mathbf{r}_1^{(s)}, \mathbf{r}_M^{(s)}) \\ \vdots & \ddots & \vdots \\ \kappa(\mathbf{r}_M^{(s)}, \mathbf{r}_1^{(s)}) & \dots & \kappa(\mathbf{r}_M^{(s)}, \mathbf{r}_M^{(s)}) \end{bmatrix} =: \mathbf{K}, \quad (\text{B5})$$

and

$$\mathbf{h}_\bullet^{(s)T} \mathbf{\Psi}^{(s)H} = \begin{bmatrix} \kappa(\mathbf{r}_1^{(s)}, \mathbf{r}_\bullet^{(s)}) \\ \vdots \\ \kappa(\mathbf{r}_M^{(s)}, \mathbf{r}_\bullet^{(s)}) \end{bmatrix}^H =: \boldsymbol{\kappa}_\bullet^H. \quad (\text{B6})$$

Substituting Eqs. (B5) and (B6) into Eq. (A2), we obtain

$$\hat{u}(\mathbf{r}^{(s)}) = \boldsymbol{\kappa}_\bullet^H (\mathbf{K} + \lambda \mathbf{I}_M)^{-1} \mathbf{y}, \quad (\text{B7})$$

which is the same expression as Eq. (10a) when $\lambda = \sigma_n^2$. Thus, SWF is demonstrated to be equivalent to GPR with the kernel function defined as Eq. (B4).

Next, we consider SWF expansion at multiple origins is considered. When the matrix inversion lemma⁴⁷ is used, Eq. (34) yields

$$\begin{aligned} \hat{u}(\mathbf{r}) &= \mathbf{h}_\bullet^T (\mathbf{\Psi}^H \mathbf{\Psi} + \lambda \mathbf{I}_V)^{-1} \mathbf{\Psi}^H \mathbf{y} \\ &= \mathbf{h}_\bullet^T \mathbf{\Psi}^H (\mathbf{\Psi} \mathbf{\Psi}^H + \lambda \mathbf{I}_M)^{-1} \mathbf{y}, \end{aligned} \quad (\text{B8})$$

where

$$\begin{aligned} \mathbf{\Psi} \mathbf{\Psi}^H &= \sum_{s=1}^S \mathbf{\Psi}^{(s)} \mathbf{\Psi}^{(s)H} \\ &= \begin{bmatrix} \sum_{s=1}^S \kappa(\mathbf{r}_1^{(s)}, \mathbf{r}_1^{(s)}) & \dots & \sum_{s=1}^S \kappa(\mathbf{r}_1^{(s)}, \mathbf{r}_M^{(s)}) \\ \vdots & \ddots & \vdots \\ \sum_{s=1}^S \kappa(\mathbf{r}_M^{(s)}, \mathbf{r}_1^{(s)}) & \dots & \sum_{s=1}^S \kappa(\mathbf{r}_M^{(s)}, \mathbf{r}_M^{(s)}) \end{bmatrix} \\ &=: \mathbf{K} \end{aligned} \quad (\text{B9})$$

and

$$\begin{aligned} \mathbf{h}_\bullet^T \mathbf{\Psi}^H &= \sum_{s=1}^S \mathbf{h}_\bullet^{(s)T} \mathbf{\Psi}^{(s)H} \\ &= \begin{bmatrix} \sum_{s=1}^S \kappa(\mathbf{r}_1^{(s)}, \mathbf{r}_\bullet^{(s)}) \\ \vdots \\ \sum_{s=1}^S \kappa(\mathbf{r}_M^{(s)}, \mathbf{r}_\bullet^{(s)}) \end{bmatrix}^H \\ &=: \boldsymbol{\kappa}_\bullet^H. \end{aligned} \quad (\text{B10})$$

When Eqs. (B9) and (B10) are substituted into Eq. (B8), SWF expansion at multiple origins is demonstrated also to be interpreted as GPR with mean

$$\hat{u}(\mathbf{r}_s) = \kappa_s^H (\mathbf{K} + \lambda \mathbf{I}_M)^{-1} \mathbf{y}, \quad (\text{B11})$$

and the kernel function

$$\kappa(\mathbf{r}_i, \mathbf{r}_j) = \sum_{s=1}^S \kappa(\mathbf{r}_i^{(s)}, \mathbf{r}_j^{(s)}) = \sum_{s=1}^S \mathbf{h}_j^{(s)H} \mathbf{h}_i^{(s)}. \quad (\text{B12})$$

- ¹M. A. Poletti, "Three-dimensional surround sound systems based on spherical harmonics," *J. Audio Eng. Soc.* **53**(11), 1004–1025 (2005).
- ²D. B. Ward and T. D. Abhayapala, "Reproduction of a plane-wave sound field using an array of loudspeakers," *IEEE Trans. Speech Audio Process.* **9**(6), 697–707 (2001).
- ³M. Otani, H. Shigetani, M. Mitsuishi, and R. Matsuda, "Binaural ambisonics: Its optimization and applications for auralization," *Acoust. Sci. Technol.* **41**(1), 142–150 (2020).
- ⁴M. Poletti, F. Fazi, and P. Nelson, "Sound-field reproduction systems using fixed-directivity loudspeakers," *J. Acoust. Soc. Am.* **127**(6), 3590–3601 (2010).
- ⁵R. Matsuda and M. Otani, "Binaural-centered mode-matching method for enhanced reproduction accuracy at listener's both ears in sound field reproduction," *J. Acoust. Soc. Am.* **150**(5), 3838–3851 (2021).
- ⁶N. Ueno, S. Koyama, and H. Saruwatari, "Three-dimensional sound field reproduction based on weighted mode-matching method," *IEEE/ACM Trans. Audio Speech Lang. Process.* **27**(12), 1852–1867 (2019).
- ⁷N. Ueno, S. Koyama, and H. Saruwatari, "Sound field recording using distributed microphones based on harmonic analysis of infinite order," *IEEE Signal Process. Lett.* **25**(1), 135–139 (2018).
- ⁸D. Caviedes-Nozal, N. A. Riis, F. M. Heuchel, J. Brunsog, P. Gerstoft, and E. Fernandez-Grande, "Gaussian processes for sound field reconstruction," *J. Acoust. Soc. Am.* **149**(2), 1107–1119 (2021).
- ⁹N. Ueno, S. Koyama, and H. Saruwatari, "Directionally weighted wave field estimation exploiting prior information on source direction," *IEEE Trans. Signal Process.* **69**, 2383–2395 (2021).
- ¹⁰N. Ueno, S. Koyama, and H. Saruwatari, "Kernel ridge regression with constraint of Helmholtz equation for sound field interpolation," in *2018 16th International Workshop on Acoustic Signal Enhancement (IWAENC)*, Tokyo, Japan (September 17–20) (IEEE, New York, 2018), pp. 1–440.
- ¹¹Z. Liang, W. Zhang, and T. D. Abhayapala, "Sound field reconstruction using neural processes with dynamic kernels," *EURASIP J. Audio Speech Music Process.* **2024**(1), 13.
- ¹²S. Koyama and L. Daudet, "Sparse representation of a spatial sound field in a reverberant environment," *IEEE J. Sel. Top. Signal Process.* **13**(1), 172–184 (2019).
- ¹³E. Fernandez-Grande, A. Xenaki, and P. Gerstoft, "A sparse equivalent source method for near-field acoustic holography," *J. Acoust. Soc. Am.* **141**(1), 532–542 (2017).
- ¹⁴E. Fernandez-Grande and L. Daudet, "Compressive acoustic holography with block-sparse regularization," *J. Acoust. Soc. Am.* **143**(6), 3737–3746 (2018).
- ¹⁵B. Rafaely, "Analysis and design of spherical microphone arrays," *IEEE Trans. Speech Audio Process.* **13**(1), 135–143 (2004).
- ¹⁶S. A. Verburg and E. Fernandez-Grande, "Reconstruction of the sound field in a room using compressive sensing," *J. Acoust. Soc. Am.* **143**(6), 3770–3779 (2018).
- ¹⁷S. Kaneko, T. Suenaga, H. Akiyama, Y. Miyake, S. Tominaga, F. Shirakihara, and H. Okumura, "Development of a 64-channel spherical microphone array and a 122-channel loudspeaker array system for 3d sound field capturing and reproduction technology research," in *Audio Engineering Society Convention 144*, Milan, Italy (May 23–26) (Audio Engineering Society, New York, 2018).
- ¹⁸P. Samarasinghe, T. Abhayapala, and M. Poletti, "Wavefield analysis over large areas using distributed higher order microphones," *IEEE/ACM Trans. Audio Speech Lang. Process.* **22**(3), 647–658 (2014).
- ¹⁹S. Kaneko and R. Duraiswami, "Multiple scattering ambisonics: Three-dimensional sound field estimation using interacting spheres," *JASA Express Lett.* **1**(8), 084801 (2021).
- ²⁰M. Pezzoli, M. Cobos, F. Antonacci, and A. Sarti, "Sparsity-based sound field separation in the spherical harmonics domain," in *ICASSP 2022–2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Singapore, Singapore (May 22–27) (IEEE, New York, 2022), pp. 1051–1055.
- ²¹E. G. Williams, *Fourier Acoustics: Sound Radiation and Nearfield Acoustical Holography* (Academic Press, New York, 1999).
- ²²Y. C. Pati, R. Rezaifar, and P. S. Krishnaprasad, "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition," in *Proceedings of 27th Asilomar Conference on Signals, Systems and Computers*, Pacific Grove, CA (November 1–3) (IEEE, New York, 1993), pp. 40–44.
- ²³J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Inf. Theory* **53**(12), 4655–4666 (2007).
- ²⁴I. Gorodnitsky and B. Rao, "Sparse signal reconstruction from limited data using focuss: A re-weighted minimum norm algorithm," *IEEE Trans. Signal Process.* **45**(3), 600–616 (1997).
- ²⁵A. Beck and M. Teboulle, "Gradient-based algorithms with applications to signal recovery," in *Convex Optimization in Signal Processing and Communications*, edited by D. P. Palomar and Y. C. Eldar (Cambridge University Press, Cambridge, UK, 2009), pp. 42–88.
- ²⁶M. Zibulevsky and M. Elad, "L1-L2 optimization in signal and image processing," *IEEE Signal Process. Mag.* **27**(3), 76–88 (2010).
- ²⁷S. Boyd, N. Parikh, E. Chu, B. Peleato, J. Eckstein *et al.*, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.* **3**(1), 1–122 (2011).
- ²⁸Z. Lin, R. Liu, and Z. Su, "Linearized alternating direction method with adaptive penalty for low-rank representation," *Adv. Neural Inform. Process. Syst.* **25**, 612–620 (2011).
- ²⁹Y. Takida, S. Koyama, N. Ueno, and H. Saruwatari, "Reciprocity gap functional in spherical harmonic domain for gridless sound field decomposition," *Signal Process.* **169**, 107383 (2020).
- ³⁰J. G. C. Ribeiro, S. Koyama, and H. Saruwatari, "Kernel interpolation of acoustic transfer functions with adaptive kernel for directed and residual reverberations," in *ICASSP 2023–2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Rhodes Island, Greece (June 4–10) (IEEE, New York, 2023), pp. 1–5.
- ³¹J. G. C. Ribeiro, S. Koyama, R. Horiuchi, and H. Saruwatari, "Sound field estimation based on physics-constrained kernel interpolation adapted to environment," *IEEE/ACM Trans. Audio Speech Lang. Process.* **32**, 4369–4383 (2024).
- ³²B. Rafaely, "Plane-wave decomposition of the sound field on a sphere by spherical convolution," *J. Acoust. Soc. Am.* **116**(4), 2149–2157 (2004).
- ³³C. A. Anderson, P. D. Teal, and M. A. Poletti, "Spatial correlation of radial Gaussian and uniform spherical volume near-field source distributions," *IEEE/ACM Trans. Audio Speech Lang. Process.* **24**(1), 143–150 (2016).
- ³⁴R. Boloix-Tortosa, J. J. Murillo-Fuentes, F. J. Payán-Somet, and F. Pérez-Cruz, "Complex Gaussian processes for regression," *IEEE Trans. Neural Networks Learn. Syst.* **29**(11), 5499–5511 (2018).
- ³⁵C. M. Bishop and N. M. Nasrabadi, *Pattern Recognition and Machine Learning* (Springer, New York, 2006), Vol. 4.
- ³⁶J. Shawe-Taylor and N. Cristianini, *Kernel Methods for Pattern Analysis* (Cambridge University Press, Cambridge, UK, 2004).
- ³⁷N. Grbic, S. Nordholm, and A. Cantoni, "Optimal FIR subband beamforming for speech enhancement in multipath environments," *IEEE Signal Process. Lett.* **10**(11), 335–338 (2003).
- ³⁸J. G. C. Ribeiro, "Physics-grounded kernel interpolation of acoustics transfer functions," Ph.D. dissertation, The University of Tokyo, Tokyo, Japan, 2024.
- ³⁹R. Duraiswami, D. Zotkin, and N. Gumerov, "Interpolation and range extrapolation of HRTFs [head related transfer functions]," in *2004 IEEE International Conference on Acoustics, Speech, and Signal Processing*, Montreal, QC, Canada (May 17–21) (IEEE, New York, 2004), Vol. 4, pp. iv.
- ⁴⁰J. Fliege, "Integration nodes for the sphere," available at <https://www.personal.soton.ac.uk/jf1w07/nodes/nodes.html> (Last viewed July 4, 2024).
- ⁴¹M. Olivieri, A. Bastine, M. Pezzoli, F. Antonacci, T. Abhayapala, and A. Sarti, "Acoustic imaging with circular microphone array: A new approach for sound field analysis," *IEEE/ACM Trans. Audio Speech Lang. Process.* **32**, 1750–1761 (2024).

- ⁴²J.-W. Choi, F. Zotter, B. Jo, and J.-H. Yoo, “Multiarray eigenbeam-ESPRIT for 3D sound source localization with multiple spherical microphone arrays,” *IEEE/ACM Trans. Audio Speech Lang. Process.* **30**, 2310–2325 (2022).
- ⁴³J. Han, L. Shao, D. Xu, and J. Shotton, “Enhanced computer vision with Microsoft Kinect sensor: A review,” *IEEE Trans. Cybern.* **43**(5), 1318–1334 (2013).
- ⁴⁴G. H. Koopmann, L. Song, and J. B. Fahline, “A method for computing acoustic fields based on the principle of wave superposition,” *J. Acoust. Soc. Am.* **86**(6), 2433–2438 (1989).
- ⁴⁵M. Johnson, S. Elliott, K. Baek, and J. Garcia-Bonito, “An equivalent source technique for calculating the sound field inside an enclosure containing scattering objects,” *J. Acoust. Soc. Am.* **104**(3), 1221–1231 (1998).
- ⁴⁶Y. Haneda, K. Furuya, and H. Itou, “Design of multipole loudspeaker array based on spherical harmonic expansion,” in *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Prague, Czech Republic (May 22–27) (IEEE, New York, 2011), pp. 141–144.
- ⁴⁷A. H. Sayed, *Adaptive Filters* (Wiley, New York, 2011).