Sound Field Estimation With Moving Microphones Using Kernel Ridge Regression

Jesper Brunnström, Martin Bo Møller, *Member, IEEE*, Jan Østergaard, *Senior Member, IEEE*, Shoichi Koyama, *Senior Member, IEEE*, Toon van Waterschoot, *Member, IEEE*, and Marc Moonen, *Fellow, IEEE*

***Abstract*—Sound field estimation with moving microphones can increase flexibility, decrease measurement time, and reduce equipment constraints compared to using stationary microphones. In this paper a kernel ridge regression (KRR)-based method is presented for sound field estimation with moving microphones. The proposed KRR method is constructed to estimate room impulse response (RIR) as a function of receiver position, using a discrete time continuous space sound field model based on the discrete Fourier transform and the Herglotz wave function. The proposed method allows for the inclusion of prior knowledge as a regularization penalty, similar to KRR methods for sound field estimation with stationary microphones, which is then novel for moving microphones. Using a directional weighting for the proposed method, the estimates are improved, which is demonstrated on both simulated and real data. Due to the high computational cost of sound field estimation with moving microphones, an approximate KRR method is also proposed, using random Fourier features (RFF) to approximate the kernel. The RFF method is shown to decrease computational cost while obtaining less accurate estimates compared to the KRR method, providing a trade-off between cost and performance.**

***Index Terms*—Kernel ridge regression, moving microphones, random fourier features, reproducing kernel hilbert space, sound field estimation.**

I. Introduction

THE development of effective sound field estimation methods is crucial for many sound field control applications [1], [2]. Sound field estimation is difficult due to the large number of spatial samples required to perfectly reconstruct the sound field [3]. Therefore in practice, the sound field is spatially undersampled, and estimation is performed by solving an ill-posed inverse problem using some type of prior knowledge as regularization.

One ubiquitous assumption for sound field estimation is that the microphones are stationary during the measurement process, which simplifies the derivations of the methods. However, estimating sound fields over large regions requires many spatial samples, which either takes a long time to collect or requires expensive large microphone arrays. Using moving microphones allows for more flexible measurements, which can shorten the measurement time or relax equipment requirements. In addition, sound field estimation could be applied in situations where the microphones cannot be kept stationary.

A special case of particular interest is a sound field generated by a source using an impulse as source signal. Such a sound field is referred to as a RIR function, viewed as a function of receiver position. The RIR function associated with a specific source is important for many applications including sound field control [4], [5], acoustic signal enhancement [6], [7], room acoustics modelling [8], [9], and room inference [10]. Estimating RIR functions is therefore the focus of this paper. It is impractical to use an impulse as a source signal, so instead swept-sine signals are commonly used [11], [12], [13]. Using stationary microphones, the recorded signal for each microphone is first deconvolved such that a RIR is obtained for each microphone position, followed by the estimation of RIRs at arbitrary positions through sound field interpolation. Using moving microphones, the deconvolution and sound field interpolation must be jointly performed, as the sound at each position is not sufficiently observed to deconvolve the signal directly. This complicates the derivation of the estimation methods, and considerably increases their computational cost.

While no specific constraints on the trajectories of the moving microphones are imposed in this paper, the trajectory needs to be known. This corresponds to the requirement that the microphone positions must be known when stationary microphones are used. In a similar way to how the positions of the stationary microphones affect the sound field estimation performance [3], [14], [15], the choice of trajectory also affects the estimation performance [16], [17], [18]. Under some constraints on the source signal, the trajectory can be obtained from the recorded audio signal itself, which means that a separate tracking system

Received 23 February 2026; revised 6 July 2026; accepted 29 July 2026. Date of publication 14 August 2026; date of current version 28 August 2026. This work was supported in part by the ESAT Laboratory of KU Leuven, in the frame of KU Leuven Research Council under Project C14-21-0075 "A holistic approach to the design of integrated and distributed digital signal processing algorithms for audio and speech communication devices", and in part by European Union's Horizon 2020 research and innovation programme through Marie Skłodowska-Curie under Grant Agreement 956369: 'Service-Oriented Ubiquitous Network-Driven Sound — SOUNDS'. The associate editor coordinating the review of this article and approving it for publication was Dr. Prasanga Samarasinghe. *(Corresponding author: Jesper Brunnström.)*

Jesper Brunnström, Toon van Waterschoot, and Marc Moonen are with the Department of Electrical Engineering (ESAT), STADIUS, KU Leuven, 3000 Leuven, Belgium (e-mail: pvjesper@gmail.com).

Martin Bo Møller was with Bang & Olufsen, Acoustics R&D, 7600 Struer, Denmark.

Jan Østergaard is with the Department of Electronic Systems, Aalborg University, 9220 Aalborg, Denmark.

Shoichi Koyama is with the National Institute of Informatics, Tokyo 101-8430, Japan

Digital Object Identifier 10.1109/TASLPRO.2026.3724402

2998-4173 © 2026 IEEE. All rights reserved, including rights for text and data mining, and training of artificial intelligence and similar technologies. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

Authorized licensed use limited to: KU Leuven University Library. Downloaded on September 29,2026 at 08:19:02 UTC from IEEE Xplore. Restrictions apply.

is not necessarily required [19]. If the obtained trajectory is not accurate, the estimates are degraded due to incorrect spatial information [20] as well as due to the Doppler shift [21].

A number of methods exist for sound field estimation with moving microphones. The methods in [22], [23], [24], [25], [26], [27] are suited for estimation on the one-dimensional trajectory along which the microphones move, some of which are relying on the specific properties of the perfect periodic sequence [13], [28] used as source signal. The problem has also been approached as a sampling problem [29], as well as by using compressive sensing to reduce measurement time [30]. One drawback of these methods is that they do not fully take advantage of the fact that the unknown quantity is a sound field, which has known properties.

Including a sound field model incorporating known properties of sound wave propagation is commonly done when stationary microphones are used. This can be done with a basis expansion, where the sound field is expressed as a superposition of position-dependent basis functions weighted by position-independent coefficients. The most commonly used basis functions are plane wave functions and spherical wave functions, which are analytical solutions to the homogenous Helmholtz equation [31]. Basis expansions have been used with spherical wave functions in [32], [33] and plane wave functions in [34] for the moving microphone case.

A basis expansion in general provides an approximate sound field model, because only a finite number of basis functions is used in order to maintain computational tractability. However, it is possible to use an infinite number of basis functions to represent the sound field, as was done with spherical harmonics in [33] for moving microphones and in [35], [36] for stationary microphones. These infinite-dimensional basis expansions actually use an implicit basis expansion, as the basis coefficients never have to be computed explicitly. Experiments have shown that such representations are more robust and produces better estimates [20], [33] compared to truncated finite-dimensional basis expansions [32].

The infinite-dimensional basis expansion method for stationary microphones in [35] is equivalent to KRR [37] for omnidirectional microphones. However, the KRR method has been further developed to incorporate prior knowledge of the sound field [38], [39], [40], as well as to take other properties of the sound field into account [41], [42]. Including such prior knowledge has been shown to improve the estimation performance significantly. The KRR method has only been developed in the frequency domain for stationary microphones, which means that the performance improvement from exploiting prior knowledge can not currently be obtained with moving microphones.

Recently, the KRR method for sound field estimation has been extended to operate in the discrete time-domain [43]. This allows for time-domain characteristics of the sound field to be incorporated as prior knowledge, but also extends the class of problems that can be solved using the method. One such problem is sound field estimation with moving microphones, which indeed requires a time-domain model in order to describe the obtained measurements. The proposed method in this paper builds upon the model described in [43], and extends it for the moving microphones case. The proposed KRR-based method is in special cases equivalent to the Bayesian estimation method proposed in [33]. The proposed method allows for additional prior knowledge to be included in the optimization problem, while the Bayesian method in [33] allows for the use of directional microphones.

It is well known that the computational cost of KRR scales poorly with the number of data points, which for moving microphones is the number of time-domain samples collected. A method for reducing the computational cost of KRR is to use RFF, where a finite basis expansion is used to approximate the kernel [44], [45]. In this paper an approximate KRR method for sound field estimation based on RFF is developed for both stationary and moving microphones. The RFF in this case correspond to a finite basis expansion using plane wave functions, as in [34], which demonstrates that finite plane wave basis expansion methods can be viewed as approximations of the proposed KRR method. This is similar to the relationship between the truncated spherical wave function expansion of [32] and the non-truncated expansion of [33]. However, one benefit of RFF is that flexible regularization penalties can be derived in the reproducing kernel Hilbert space (RKHS) domain, which then translates to RFF as well.

To summarize, in this paper a KRR approach to sound field estimation with moving microphones is proposed. The proposed method, like [33], expresses the sound field in an infinite-dimensional basis, which leads to improved estimation performance and robustness compared to existing methods that rely on finite-dimensional basis expansions. In addition, the proposed method allows for prior knowledge to be included in estimation problem in the same way as in the stationary microphone case [38], [39], [40], [43], [46], which further improves estimation performance. The paper builds on the KRR-based time-domain sound field estimation method developed in [43], extending the approach to moving microphones. The extension requires the formulation of KRR in the context of a new measurement model that allows for joint deconvolution and sound field interpolation. The proposed method is evaluated on both simulated data and real data from [20], demonstrating improved estimation performance compared to existing methods.

A. Notation

The vector space of N-dimensional real vectors and of N-dimensional complex vectors are denoted by $\mathbb{R}^N$ and $\mathbb{C}^N$ respectively. The unit sphere in three dimensions is denoted by $\mathbb{S}_2$, defined as $\{\boldsymbol{x} \mid \|\boldsymbol{x}\| = 1, \boldsymbol{x} \in \mathbb{R}^3\}$. The complex conjugate of a is $\overline{a}$, the transpose of matrix $\boldsymbol{A}$ is $\boldsymbol{A}^\top$, and the Hermitian transpose of $\boldsymbol{A}$ is $\boldsymbol{A}^\mathsf{H}$. The imaginary unit is $j^2 = -1$. The same notation is used for linear operators as for matrix multiplication, where the operator is applied to the variable to the right of it, e.g. in the expression Ta, the variable a is the argument of the linear operator T. The adjoint of a linear operator $T : \mathcal{H} \to \mathcal{H}'$ between Hilbert spaces $\mathcal{H}$ and $\mathcal{H}'$, is denoted by $T^* : \mathcal{H}' \to \mathcal{H}$. The selection of an element with index l from $\boldsymbol{a}$ is denoted by $(\boldsymbol{a})_l$, which is assumed to begin at zero. Variables referring to

Authorized licensed use limited to: KU Leuven University Library. Downloaded on September 29,2026 at 08:19:02 UTC from IEEE Xplore. Restrictions apply.

time-domain and frequency-domain quantities are marked with and without a tilde respectively, like $\tilde{\boldsymbol{a}}$ and $\boldsymbol{a}$.

II. Problem Statement

The task considered in this paper is to reconstruct a RIR for all positions within a region of interest, where a discrete set of time-domain samples is collected by continuously moving microphones. The region of interest $\Omega \subset \mathbb{R}^3$ is a simply connected source-free region. Outside the region Ω is a sound source of interest which emits a known signal ϕ, e.g. a loudspeaker playing a predetermined training signal. If multiple sources of interest are present, they can be viewed as a single source, following the strategies in [26] [33, Sec.IV.E].

The acoustic response of the room is assumed to be time-invariant over the duration of the measurement process. The acoustic response associated with the position $\boldsymbol{r} \in \Omega$ sampled with sampling rate $f_s \in \mathbb{R}_{>0}$ can be modelled in discrete time as a constant L-length finite impulse response (FIR) filter, the impulse response of which is referred to as a RIR. The RIRs associated with all positions $\boldsymbol{r} \in \Omega$ can then be represented by a RIR function $\tilde{\boldsymbol{u}} : \Omega \to \mathcal{V}_\mathfrak{t}$, where $\mathcal{V}_\mathfrak{t}$ is the Hilbert space of L-length sequences with inner product

$$\langle \tilde{\boldsymbol{a}}, \tilde{\boldsymbol{a}}' \rangle_{\mathcal{V}_\mathfrak{t}} = \tilde{\boldsymbol{a}}'^{\top} \tilde{\boldsymbol{a}} \tag{1}$$

where $\tilde{\boldsymbol{a}}, \tilde{\boldsymbol{a}}' \in \mathcal{V}_\mathfrak{t}$. The space $\mathcal{V}_\mathfrak{t}$ is isomorphic to the standard Euclidean vector space $\mathbb{R}^L$.

Now consider an omnidirectional microphone moving continuously inside Ω. The microphone collects samples, obtaining the time-domain signal $p(n) \in \mathbb{R}$, where $n \in \mathbb{Z}$ is the discrete-time time index. The position of the microphone is synchronously recorded, giving a sequence of positions $\boldsymbol{r}_0, \ldots, \boldsymbol{r}_{N-1}$ representing the trajectory along which the microphone moves. The source signal is similarly sampled, defining the known signal $\phi(n) \in \mathbb{R}$.

Given the presence of measurement noise $s(n) \in \mathbb{R}$, the measured time-domain signal can then be expressed as

$$p(n) = \langle \tilde{\boldsymbol{u}}(\boldsymbol{r}_n), \boldsymbol{\phi}(n) \rangle_{\mathcal{V}_\mathfrak{t}} + s(n) \tag{2}$$

for time indices $n = 0, \ldots, N-1$, and $\boldsymbol{\phi}(n) = (\phi(n), \ldots, \phi(n-L+1)) \in \mathcal{V}_\mathfrak{t}$ containing the most recent L samples of the source signal. The task is then to estimate the RIR function $\tilde{\boldsymbol{u}}$ from the available data.

For clarity, only a single moving microphone is considered. The results can be used without modification for M moving microphones when the source signal ϕ is periodic, in which case the trajectories and signals from an integer number of periods from the M microphones can be concatenated, appearing like a single microphone with a discontinuous trajectory.

III. Reproducing Kernel Hilbert Space

The RIR function will be estimated using KRR. This requires a definition of a RKHS containing permissible RIR functions, which are modelled as discrete-time sound fields following the approach of [43]. First, the discrete Fourier transform (DFT) is described, to relate the time- and frequency-domain representations of a sound field. Then, the RKHSs for all frequencies are aggregated using the DFT to obtain the desired function space, which is a RKHS representing time-domain sound fields. This section is a minimal restatement of definitions and results from [43], where more details can be found.

A. Frequency-Domain Representation

The DFT is a transform between the space of time-domain sequences $\mathcal{V}_\mathfrak{t} = \mathbb{R}^L$, and the space of frequency-domain sequences $\mathcal{V}_\mathfrak{f} = \mathbb{C}^{L_f - 2} \times \mathbb{R}^2$ if L is even and $\mathcal{V}_\mathfrak{f} = \mathbb{C}^{L_f - 1} \times \mathbb{R}$ if L is odd, where $L_f = \lfloor \frac{L}{2} \rfloor + 1$. The forward transform $\mathcal{F} : \mathcal{V}_\mathfrak{t} \to \mathcal{V}_\mathfrak{f}$ and the inverse transform $\mathcal{F}^{-1} : \mathcal{V}_\mathfrak{f} \to \mathcal{V}_\mathfrak{t}$ between $\tilde{\boldsymbol{v}} \in \mathcal{V}_\mathfrak{t}$ and $\boldsymbol{v} \in \mathcal{V}_\mathfrak{f}$ are defined here as

$$\begin{aligned} (\mathcal{F}\tilde{\boldsymbol{a}})_l &= \sum_{n=0}^{L-1} e^{2\pi j n l / L} (\tilde{\boldsymbol{a}})_n \quad 0 \le l < L_f \\ (\mathcal{F}^{-1}\boldsymbol{a})_n &= \mathfrak{Re}\left[\sum_{l=0}^{L_f - 1} c_l e^{-2\pi j n l / L} (\boldsymbol{a})_l \right] \quad 0 \le n < L, \end{aligned} \tag{3}$$

For sound field estimation, the frequency at $l = 0$ and the Nyquist frequency $l = \frac{L}{2}$ are generally not of interest, as antialiasing filters and source characteristics lead to measurements being close to zero at these frequencies. These two special cases are still presented to maintain internal consistency of the model.

The real-part operator is defined as $\mathfrak{Re}[v] = \frac{1}{2}(a + \overline{a})$, and

$$c_l = \frac{2}{L} \qquad 0 < l < \frac{L}{2} \tag{4}$$

with $c_l = \frac{1}{L}$ only for the special cases $l = 0$ or $l = \frac{L}{2}$. It is convenient that the transform is unitary, like the standard DFT is with appropriate scaling. This is achieved by choosing the inner product of $\mathcal{V}_\mathfrak{f}$ as

$$\langle \boldsymbol{a}, \boldsymbol{a}' \rangle_{\mathcal{V}_\mathfrak{f}} = \mathfrak{Re}[\boldsymbol{a}'^{\mathsf{H}} \boldsymbol{C} \boldsymbol{a}]. \tag{5}$$

where $\boldsymbol{C} \in \mathbb{R}^{L_f \times L_f}$ is a diagonal matrix defined by $(\boldsymbol{C})_{ll} = c_l$.

This transform is sometimes referred to as the *real* DFT, as it assumes the time-domain signal to be real, and removes the redundancy in the form of conjugate symmetry in the frequency-domain signal. The time-convention used in (3) is consistent with acoustics literature [47], but is opposite to some popular implementations of the fast Fourier transform (FFT) [48], [49].

B. Plane Wave Model

The sound field model is based on the Herglotz wave function for each of the sampled frequencies $\omega_l = \frac{2\pi f_s l}{L}$rad / s in $\mathcal{V}_\mathfrak{f}$, where again $0 \le l < L_f$ as in (3). Any time-domain sound field $\tilde{\boldsymbol{u}} : \Omega \to \mathcal{V}_\mathfrak{t}$ can be approximated arbitrarily well by [43, Section III.B][38, Appendix A]

$$\tilde{\boldsymbol{u}}(\boldsymbol{r}) = \mathcal{F}^{-1} \int_{\mathbb{S}_2} \boldsymbol{E}(\boldsymbol{r}, \boldsymbol{d}) \mathring{\boldsymbol{u}}(\boldsymbol{d}) \, ds(\boldsymbol{d}). \tag{6}$$

The directionality function $\mathring{\boldsymbol{u}} : \mathbb{S}_2 \to \mathcal{V}_\mathfrak{f}$ represents the complex plane wave coefficients at the sampled frequencies. The set of permissible directionality functions is $L_2(\mathbb{S}_2 \to \mathcal{V}_\mathfrak{f})$, which denotes the Hilbert space of square integrable functions from the unit sphere $\mathbb{S}_2$ to $\mathcal{V}_\mathfrak{f}$, where functions are considered equal if they

Authorized licensed use limited to: KU Leuven University Library. Downloaded on September 29,2026 at 08:19:02 UTC from IEEE Xplore. Restrictions apply.

only differ on a set with measure zero. The diagonal operator $\boldsymbol{E}(\boldsymbol{r}, \boldsymbol{d}) \in \mathcal{B}(\mathcal{V}_\mathfrak{f})$, for a plane wave incoming from the direction $\boldsymbol{d}$, can be represented by a matrix with diagonal elements

$$(\boldsymbol{E}(\boldsymbol{r}, \boldsymbol{d}))_{ll} = e^{-j\frac{\omega_l}{c}\boldsymbol{r}^\top \boldsymbol{d}} \qquad 0 \leq l < \frac{L}{2}, \tag{7}$$

where linear bounded operators from $\mathcal{V}_\mathfrak{f}$ to $\mathcal{V}_\mathfrak{f}$ are denoted by $\mathcal{B}(\mathcal{V}_\mathfrak{f})$. Elements of $\mathcal{V}_\mathfrak{f}$ must be real-valued at the Nyquist frequency, which means that $(\boldsymbol{E}(\boldsymbol{r}, \boldsymbol{d}))_{\frac{L}{2}\frac{L}{2}} = \cos(\frac{\omega_l}{c}\boldsymbol{r}^\top \boldsymbol{d})$. The speed of sound is denoted by c, which is assumed to be constant in Ω.

C. Reproducing Kernel Hilbert Space

The RKHS of sound fields $\tilde{\mathcal{H}}$ can be defined in terms of the operator $\mathcal{A} : L_2(\mathbb{S}_2 \to \mathcal{V}_\mathfrak{f}) \to \tilde{\mathcal{H}}$, which takes a directionality function $\mathring{\boldsymbol{u}}$ and transforms it into a sound field $\tilde{\boldsymbol{u}}$ according to (6). The RKHS can then be defined as

$$\tilde{\mathcal{H}} = \{\mathcal{A}\mathring{\boldsymbol{u}} \mid \mathring{\boldsymbol{u}} \in L_2(\mathbb{S}_2 \to \mathcal{V}_\mathfrak{f})\}, \tag{8}$$

with the inner product

$$\langle \tilde{\boldsymbol{u}}, \tilde{\boldsymbol{v}} \rangle_{\tilde{\mathcal{H}}} = \int_{\mathbb{S}_2} \langle \mathring{\boldsymbol{u}}(\boldsymbol{d}), \mathring{\boldsymbol{v}}(\boldsymbol{d}) \rangle_{\mathcal{V}_\mathfrak{f}} \, ds(\boldsymbol{d}). \tag{9}$$

A RKHS is characterized by its kernel function. The kernel function $\tilde{\Gamma} : \Omega \times \Omega \to \mathcal{B}(\mathcal{V}_\mathfrak{t})$ of $\tilde{\mathcal{H}}$, where $\mathcal{B}(\mathcal{V}_\mathfrak{t})$ denotes bounded linear operators from $\mathcal{V}_\mathfrak{t}$ to $\mathcal{V}_\mathfrak{t}$, is

$$\tilde{\Gamma}(\boldsymbol{r}, \boldsymbol{r}') = \mathcal{F}^{-1}\Gamma(\boldsymbol{r}, \boldsymbol{r}')\mathcal{F}, \tag{10}$$

where $\Gamma : \Omega \times \Omega \to \mathcal{B}(\mathcal{V}_\mathfrak{f})$ is diagonal and can be represented by a matrix with diagonal elements

$$(\Gamma(\boldsymbol{r}, \boldsymbol{r}'))_{ll} = \kappa_{\omega_l}(\boldsymbol{r}, \boldsymbol{r}') \qquad 0 \leq l < \frac{L}{2}. \tag{11}$$

At the Nyquist frequency the value is instead $(\Gamma(\boldsymbol{r}, \boldsymbol{r}'))_{\frac{L}{2}\frac{L}{2}} = \frac{1}{2}(\kappa_{\omega_l}(\boldsymbol{r}, \boldsymbol{r}') + \kappa_{\omega_l}(\boldsymbol{r}, -\boldsymbol{r}'))$.

Finally, the single-frequency kernel $\kappa_\omega : \Omega \times \Omega \to \mathbb{C}$ is

$$\kappa_\omega(\boldsymbol{r}, \boldsymbol{r}') = j_0\left(\frac{\omega}{c}\|\boldsymbol{r} - \boldsymbol{r}'\|_2\right), \tag{12}$$

where j_0 is the zeroth-order spherical Bessel function of the first kind.

IV. Sound Field Estimation

A. Kernel Ridge Regression

To estimate the RIR function $\tilde{\boldsymbol{u}}$, an optimization problem in the form of KRR will be formulated. The relationship between the function $\tilde{\boldsymbol{u}}$ and the measured data at time n can be expressed as a linear measurement operator $\mathcal{M}_n : \tilde{\mathcal{H}} \to \mathbb{R}$, defined as

$$\mathcal{M}_n\tilde{\boldsymbol{u}} = \langle \tilde{\boldsymbol{u}}(\boldsymbol{r}_n), \boldsymbol{\phi}(n) \rangle_{\mathcal{V}_\mathfrak{t}}, \tag{13}$$

which means the data model in (2) can be expressed as $p(n) = \mathcal{M}_n\tilde{\boldsymbol{u}} + s(n)$.

The goal is to minimize the difference between the measured data $p(n)$ and what the measurement model (13) predicts, while also applying appropriate regularization. Such a KRR problem can be stated as

$$\underset{\tilde{\boldsymbol{u}} \in \tilde{\mathcal{H}}}{\text{minimize}} \sum_{n=0}^{N-1} (p(n) - \mathcal{M}_n\tilde{\boldsymbol{u}})^2 + \lambda\|\tilde{\boldsymbol{u}}\|^2_{\tilde{\mathcal{H}}}, \tag{14}$$

where $\lambda \in \mathbb{R}_{\geq 0}$ is a parameter controlling the strength of the regularization.

According to the representer theorem [50], the optimal solution to a KRR problem can be stated in terms of the adjoints of the measurement operators, meaning that the optimal solution to (14) is

$$\tilde{\boldsymbol{u}} = \sum_{n=0}^{N-1} \mathcal{M}_n^* a_n \tag{15}$$

for a priori unknown parameters $a_n \in \mathbb{R}$, and where $\mathcal{M}_n^*$ is the adjoint of the measurement operator $\mathcal{M}_n$ defined in (13). When the measurement operator is a point-evaluation operator, as is the case with stationary microphones in [43], the adjoint coincides with the kernel function of the RKHS. In this case with moving microphones, the measurement operators $\mathcal{M}_n$ are not point-evaluation operators, and so their adjoints are distinct from the kernel function.

The adjoint $\mathcal{M}_n^* : \mathbb{R} \to \tilde{\mathcal{H}}$ is derived in Appendix A, and is given as

$$\mathcal{M}_n^* a = \tilde{\Gamma}(\cdot, \boldsymbol{r}_n)\boldsymbol{\phi}(n)a, \tag{16}$$

which can be inserted into (15). It can be noted that using the adjoint of the measurement operator in (13) rather than a point-evaluation operator leads to representers (15) that take the source signal $\boldsymbol{\phi}$ into account.

B. Solving Kernel Ridge Regression

Inserting the form of the optimal estimate (15) into the cost function (14) leads to a finite-dimensional convex optimization problem in terms of the parameters a_n. With the parameters and data combined into vectors as $\boldsymbol{a} = (a_0, \ldots, a_{N-1}) \in \mathbb{R}^N$ and $\boldsymbol{p} = (p(0), \ldots, p(N-1)) \in \mathbb{R}^N$, the optimization problem is

$$\underset{\boldsymbol{a} \in \mathbb{R}^N}{\text{minimize}} \ \|\boldsymbol{p} - \boldsymbol{K}\boldsymbol{a}\|^2_{\mathbb{R}^N} + \lambda\langle \boldsymbol{K}\boldsymbol{a}, \boldsymbol{a} \rangle_{\mathbb{R}^N}. \tag{17}$$

The kernel matrix $\boldsymbol{K} \in \mathbb{R}^{N \times N}$ is a positive semi-definite matrix defined as

$$\begin{aligned}(\boldsymbol{K})_{nm} &= \mathcal{M}_n\mathcal{M}_m^* \\ &= \langle \tilde{\Gamma}(\boldsymbol{r}_n, \boldsymbol{r}_m)\boldsymbol{\phi}(m), \boldsymbol{\phi}(n) \rangle_{\mathcal{V}_\mathfrak{t}}.\end{aligned} \tag{18}$$

The optimization problem in (17) is a regularized least squares problem, and hence has a unique optimum, which is

$$\boldsymbol{a} = (\boldsymbol{K} + \lambda\boldsymbol{I})^{-1}\boldsymbol{p}. \tag{19}$$

Finally, the RIR function estimate at an arbitrary point $\boldsymbol{r} \in \Omega$ is obtained by using the optimal $\boldsymbol{a}$ in (15), leading to the expression

$$\tilde{\boldsymbol{u}}(\boldsymbol{r}) = \sum_{n=0}^{N-1} \tilde{\Gamma}(\boldsymbol{r}, \boldsymbol{r}_n)\boldsymbol{\phi}(n)a_n. \tag{20}$$

Authorized licensed use limited to: KU Leuven University Library. Downloaded on September 29,2026 at 08:19:02 UTC from IEEE Xplore. Restrictions apply.

The estimate using this method is equivalent to the estimate obtained from the Bayesian approach taken in [33] for omnidirectional microphones.

C. Kernel Ridge Regression Including Prior Knowledge

In many cases there is more prior knowledge available than only the physical properties of sound propagation. In a KRR context, such prior knowledge can be included by constructing a regularization function penalizing unlikely sound fields, as has been demonstrated in [38], [39], [40], [43], [51] particularly for prior knowledge regarding the sound propagation direction. The KRR problem for moving microphones considered in this paper admits a similar method to construct regularization penalties as in [43].

Generally, the KRR problem stated in (14) is modified to

$$\underset{\tilde{\boldsymbol{u}}\in\tilde{\mathcal{H}}}{\text{minimize}} \sum_{n=0}^{N-1} \left(p(n) - \mathcal{M}_n\tilde{\boldsymbol{u}}\right)^2 + \lambda\|\mathcal{R}\tilde{\boldsymbol{u}}\|_{\mathcal{Z}}^2, \tag{21}$$

where $\mathcal{R} : \tilde{\mathcal{H}} \to \mathcal{Z}$ is an invertible bounded linear operator, and $\mathcal{Z}$ is a Hilbert space. The operator $\mathcal{R}$ should be constructed such that RIR functions that are likely to occur according to the available prior knowledge are given a small norm.

Solving (21) follows the derivations and results from Section IV-A and IV-B, except that the kernel function $\tilde{\Gamma}$ is replaced by the modified kernel function $\tilde{\Gamma}_r : \Omega \times \Omega \to \mathcal{B}(\mathcal{V}_\mathfrak{t})$, which is defined as

$$\tilde{\Gamma}_r(\boldsymbol{r}, \boldsymbol{r}') = \left((\mathcal{R}^*\,\mathcal{R})^{-1}\tilde{\Gamma}(\cdot, \boldsymbol{r}')\right)(\boldsymbol{r}). \tag{22}$$

Note that the kernel $\tilde{\Gamma}$ with one argument fixed is a member of the RKHS, i.e. $\tilde{\Gamma}(\cdot, \boldsymbol{r}') \in \tilde{\mathcal{H}}$. The notation in (22) should therefore be read as evaluating $(\mathcal{R}^*\,\mathcal{R})^{-1}\tilde{\Gamma}(\cdot, \boldsymbol{r}') \in \tilde{\mathcal{H}}$ in the point $\boldsymbol{r} \in \Omega$.

If there is prior knowledge available on the directionality of the RIR function, a directional weighting can be used. Following the definitions of [43, Section V.A], the regularization operator $\mathcal{R}$ is

$$(\mathcal{R}\,\tilde{\boldsymbol{u}})(\boldsymbol{r}) = \mathcal{F}^{-1}\int_{\mathbb{S}_2} \boldsymbol{E}(\boldsymbol{r}, \boldsymbol{d})\boldsymbol{W}(\boldsymbol{d})\mathring{\boldsymbol{u}}(\boldsymbol{d})ds(\boldsymbol{d}), \tag{23}$$

which is characterized by the weighting $\boldsymbol{W} : \mathbb{S}_2 \to \mathcal{B}(\mathcal{V}_\mathfrak{f})$. The modified kernel function is then given as

$$\tilde{\Gamma}_r(\boldsymbol{r}, \boldsymbol{r}') = \mathcal{F}^{-1}\Gamma_r(\boldsymbol{r}, \boldsymbol{r}')\mathcal{F} \tag{24}$$

with

$$\Gamma_r(\boldsymbol{r}, \boldsymbol{r}') = \int_{\mathbb{S}_2} \boldsymbol{E}(\boldsymbol{r}, \boldsymbol{d})(\boldsymbol{W}(\boldsymbol{d})^*\boldsymbol{W}(\boldsymbol{d}))^{-1}\boldsymbol{E}(-\boldsymbol{r}', \boldsymbol{d})\,ds(\boldsymbol{d}), \tag{25}$$

where $\boldsymbol{E}(\boldsymbol{r}', \boldsymbol{d})^* = \boldsymbol{E}(-\boldsymbol{r}', \boldsymbol{d})$ is used in (25).

A practical special case is when $\boldsymbol{W}$ is diagonal for all $\boldsymbol{d} \in \mathbb{S}_2$, referred to as a frequency-wise directional weighting. In that case, the weighting can be defined by a scalar-valued positive function $\gamma_l : \mathbb{S}_2 \to \mathbb{R}_{>0}$ per frequency, as

$$\left((\boldsymbol{W}(\boldsymbol{d})^*\boldsymbol{W}(\boldsymbol{d}))^{-1}\right)_{ll} = \gamma_l(\boldsymbol{d}). \tag{26}$$

With such a weighting, the frequency-domain kernel Γ_r is still diagonal, with diagonal elements $(\Gamma_r(\boldsymbol{r}, \boldsymbol{r}'))_{ll} = \kappa^\gamma_{\omega_l}(\boldsymbol{r}, \boldsymbol{r}')$ which for all frequencies except for $\omega_{\frac{L}{2}}$ is given as

$$\kappa^\gamma_\omega(\boldsymbol{r}, \boldsymbol{r}') = \int_{\mathbb{S}_2} e^{-j\frac{\omega}{c}(\boldsymbol{r}-\boldsymbol{r}')^\top\boldsymbol{d}}\gamma(\boldsymbol{d})\,ds(\boldsymbol{d}). \tag{27}$$

Such a weighting is employed in the frequency domain KRR methods of [39], [40], [41], [51], hence the methods therein for finding appropriate weighting functions γ_l can be directly applied.

The kernel is particularly simple when the weighting is $\gamma_l(\boldsymbol{d}) = e^{-\beta_l\boldsymbol{\eta}_l^\top\boldsymbol{d}}$, where $\boldsymbol{\eta}_l \in \mathbb{S}_2$ is a given direction and $\beta_l \in \mathbb{R}_{>0}$ is the strength of the weighting [43, Section V.A]. This weighting is proportional to the von Mises-Fisher distribution [52, Section 9.3.2]. Using this weighting in the KRR problem leads to a preference for estimates where the sound propagates in the $\boldsymbol{\eta}_l$ direction. The kernel's diagonal elements are then given as

$$\begin{aligned}(\Gamma_r(\boldsymbol{r}, \boldsymbol{r}'))_{ll} &= j_0(\sqrt{\boldsymbol{\xi}_l^\top\boldsymbol{\xi}_l}) \\ \boldsymbol{\xi}_l &= \frac{\omega_l}{c}(\boldsymbol{r} - \boldsymbol{r}') + j\beta_l\boldsymbol{\eta}_l\end{aligned} \tag{28}$$

for $0 \le l < \frac{L}{2}$, where it is assumed that $\beta_{\frac{L}{2}} = 0$ to ensure a closed-form solution. The non-weighted kernel (11) is recovered if $\beta = 0$.

D. Computation of the Kernel Matrix

The most straightforward implementation of the expression in (18) to compute $\boldsymbol{K}$ would require N^2 matrix-vector products, each with a computational cost of $\mathcal{O}(L^2)$. This would quickly lead to an infeasibly large computational cost. However, the structure of $\boldsymbol{K}$ in (18) reveals a cheaper way of computing the same matrix if a kernel obtained from a frequency-wise weighting as described in (26) is used, which includes the non-weighted kernel (11) as a special case.

Combining (10) and (18) gives

$$(\boldsymbol{K})_{nm} = \langle\Gamma(\boldsymbol{r}_n, \boldsymbol{r}_m)\boldsymbol{\phi}_f(m), \boldsymbol{\phi}_f(n)\rangle_{\mathcal{V}_\mathfrak{f}}, \tag{29}$$

where $\boldsymbol{\phi}_f(n) = \mathcal{F}\boldsymbol{\phi}(n)$. Using the definition of the inner product in (5), and exploiting the fact that Γ is diagonal, leads to

$$(\boldsymbol{K})_{nm} = \sum_{l=0}^{L_f-1} c_l\mathfrak{Re}[\kappa^\gamma_{\omega_l}(\boldsymbol{r}_n, \boldsymbol{r}_m)\overline{(\boldsymbol{\phi}_f(n))_l}(\boldsymbol{\phi}_f(m))_l]. \tag{30}$$

The expression in (30) requires $L_f N^2$ evaluations of the scalar-valued kernel function to construct $\boldsymbol{K}$. Although the computational cost can still be high for large N, the cost is significantly reduced compared to the straightforward implementation.

After constructing $\boldsymbol{K}$, the dominant computational cost is in solving the linear system in (19) for potentially large N. Therefore, it can be beneficial to use large scale matrix-free solvers [53], as was demonstrated in [34].

Authorized licensed use limited to: KU Leuven University Library. Downloaded on September 29,2026 at 08:19:02 UTC from IEEE Xplore. Restrictions apply.

V. Fast Sound Field Estimation With Random Fourier Features

The construction and inversion of matrix $\boldsymbol{K}$ in (17) with N^2 elements has a computational cost of $\mathcal{O}(N^3)$ for most algorithms. For KRR-based sound field estimation with stationary microphones in the frequency-domain [37], N is equal to the number of microphones, leading to a small matrix and low computational cost. Using moving microphones, N is equal to the number of collected time-domain samples, leading to very large computational cost for high sampling rates or long measurement times. In this section, a structured approach to lower the computational cost is proposed, by using RFF as introduced in [44]. Initially restricted to shift-invariant kernels, the method has been extended to more classes of kernels, with additional methods for reducing computational cost further [45]. In this section, specifically the approach in [44] is considered.

A. Random Fourier Features for Stationary Microphones

The RFF approach will first be described for stationary microphones, using the single-frequency KRR problem [37]. Although computational cost reduction is rarely needed with this method, it is a step towards the more useful moving microphone case.

The KRR optimization problem for stationary microphones is

$$\underset{u_\omega \in \mathcal{H}_\omega}{\text{minimize}} \sum_{m=0}^{M-1} |p_\omega(m) - u_\omega(\boldsymbol{r}_m)|^2 + \lambda \|u_\omega\|_{\mathcal{H}_\omega}^2, \tag{31}$$

where the data $p_\omega(m) \in \mathbb{C}$ is the component at frequency ω of the RIR measured by microphone m at position $\boldsymbol{r}_m \in \Omega$. The RKHS $\mathcal{H}_\omega$ contains functions $u_\omega : \Omega \to \mathbb{C}$ that represent the complex sound pressure at frequency ω, for which the kernel function is

$$\kappa_\omega^\gamma(\boldsymbol{r}, \boldsymbol{r}') = \int_{\mathbb{S}_2} e^{-j\frac{\omega}{c}(\boldsymbol{r}-\boldsymbol{r}')^\top \boldsymbol{d}} \gamma(\boldsymbol{d})\, ds(\boldsymbol{d}), \tag{32}$$

where γ is a scalar weighting function, corresponding to a frequency-wise weighting as defined in (26).

The considered RFF approach requires a shift-invariant kernel $\kappa : \mathbb{R}^3 \times \mathbb{R}^3 \to \mathbb{C}$, which means it satisfies $\kappa(\boldsymbol{r}, \boldsymbol{r}') = \underline{\kappa}(\boldsymbol{r} - \boldsymbol{r}')$ for $\underline{\kappa} : \mathbb{R}^3 \to \mathbb{C}$. Central to the approach is Bochner's theorem, which states that any shift-invariant kernel is the Fourier transform of a probability measure $\mu : \mathbb{R}^3 \to \mathbb{R}$. In this case,

$$\kappa_\omega^\gamma(\boldsymbol{r}, \boldsymbol{r}') = \underline{\kappa}_\omega^\gamma(\boldsymbol{r} - \boldsymbol{r}') = \int_{\mathbb{R}^3} e^{-j(\boldsymbol{r}-\boldsymbol{r}')^\top \boldsymbol{\xi}}\, \mu_\omega(\boldsymbol{\xi})\, d\boldsymbol{\xi}, \tag{33}$$

under the assumption that κ_ω^γ is scaled such that the total measure of μ_ω is one.

The implication of (33) is that $e^{-j(\boldsymbol{r}-\boldsymbol{r}')^\top \boldsymbol{\xi}}$ is an unbiased estimate of $\kappa_\omega^\gamma(\boldsymbol{r}, \boldsymbol{r}')$ if $\boldsymbol{\xi}$ is obtained by sampling $\mathbb{R}^3$ according to μ_ω. An estimate with less variance can be obtained by using more than one sample, leading to the approximation

$$\kappa_\omega^\gamma(\boldsymbol{r}, \boldsymbol{r}') \approx \frac{1}{D} \sum_{d=1}^{D} e^{-j(\boldsymbol{r}-\boldsymbol{r}')^\top \boldsymbol{\xi}_d} = \langle \boldsymbol{z}_\omega(\boldsymbol{r}), \boldsymbol{z}_\omega(\boldsymbol{r}') \rangle_{\mathbb{C}^D}, \tag{34}$$

where $\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_D$ is a list of the D samples of $\mathbb{R}^3$, and $\boldsymbol{z}_\omega(\boldsymbol{r}) = \frac{1}{\sqrt{D}}(e^{-j\boldsymbol{r}^\top \boldsymbol{\xi}_1}, \ldots, e^{-j\boldsymbol{r}^\top \boldsymbol{\xi}_D}) \in \mathbb{C}^D$ is the random basis vector.

The probability measure μ_ω that satisfies (33) for κ_ω^γ is

$$\mu_\omega(\boldsymbol{\xi}) = \delta\left(\|\boldsymbol{\xi}\|_{\mathbb{R}^3} - \frac{\omega}{c}\right) \gamma(\boldsymbol{\xi}) C_\gamma, \tag{35}$$

where δ is the continuous Dirac delta function, and C_γ is a normalization constant ensuring that the total measure is 1, which will depend on the choice of γ. The derivation of (35) is given in Appendix B. The RFF basis can be identified as a basis of plane waves sampled proportionally to the weighting function γ.

Applying the representer theorem to (31) results in

$$\underset{\boldsymbol{a} \in \mathbb{C}^M}{\text{minimize}}\ \|\boldsymbol{p}_\omega - \boldsymbol{K}_\omega \boldsymbol{a}_\omega\|_{\mathbb{C}^M}^2 + \lambda \langle \boldsymbol{K}_\omega \boldsymbol{a}_\omega, \boldsymbol{a}_\omega \rangle_{\mathbb{C}^M}, \tag{36}$$

where $(\boldsymbol{K}_\omega)_{mm'} = \kappa_\omega^\gamma(\boldsymbol{r}_m, \boldsymbol{r}_{m'})$, the parameter vector is $\boldsymbol{a}_\omega \in \mathbb{C}^M$, and $\boldsymbol{p}_\omega = (p_\omega(0), \ldots, p_\omega(M-1)) \in \mathbb{C}^M$. Using the kernel approximation in (34) means that the kernel matrix can be approximated as $\boldsymbol{K}_\omega \approx \boldsymbol{Z}_\omega \boldsymbol{Z}_\omega^{\mathsf{H}}$, where $\boldsymbol{Z}_\omega \in \mathbb{C}^{M \times D}$ is

$$\boldsymbol{Z}_\omega = \begin{bmatrix} \boldsymbol{z}_\omega(\boldsymbol{r}_1)^\top \\ \vdots \\ \boldsymbol{z}_\omega(\boldsymbol{r}_M)^\top \end{bmatrix}. \tag{37}$$

Optimization problem (36) can be restated in terms of the random basis parameter vector $\boldsymbol{b}_\omega \in \mathbb{C}^D$, which is defined as $\boldsymbol{b}_\omega = \boldsymbol{Z}_\omega^{\mathsf{H}} \boldsymbol{a}_\omega$, by inserting (34) into (36). The optimization problem is then

$$\underset{\boldsymbol{b}_\omega \in \mathbb{C}^D}{\text{minimize}}\ \|\boldsymbol{p}_\omega - \boldsymbol{Z}_\omega \boldsymbol{b}_\omega\|_{\mathbb{C}^M}^2 + \lambda \|\boldsymbol{b}_\omega\|_{\mathbb{C}^M}^2, \tag{38}$$

for which the optimal solution is

$$\boldsymbol{b}_\omega = (\boldsymbol{Z}_\omega^{\mathsf{H}} \boldsymbol{Z}_\omega + \lambda \boldsymbol{I})^{-1} \boldsymbol{Z}_\omega^{\mathsf{H}} \boldsymbol{p}_\omega. \tag{39}$$

Finally, the RIR function estimate is given as

$$u_\omega(\boldsymbol{r}) = \boldsymbol{z}_\omega(\boldsymbol{r})^\top \boldsymbol{b}_\omega. \tag{40}$$

B. Random Fourier Features for Moving Microphones

In this section, the RFF approach will be applied to the moving microphone case. The following derivations hold for all kernels obtained from a frequency-wise directional weighting as described in (26), which includes the non-weighted kernel (11) as a special case. Similar to (34), these kernels can be approximated as

$$\tilde{\Gamma}(\boldsymbol{r}, \boldsymbol{r}') \approx \mathcal{F}^{-1} \frac{1}{D} \sum_{d=1}^{D} \boldsymbol{E}(\boldsymbol{r}, \boldsymbol{\xi}_d) \boldsymbol{E}(-\boldsymbol{r}', \boldsymbol{\xi}_d) \mathcal{F}, \tag{41}$$

where the vectors $\boldsymbol{\xi}_d$ are obtained according to the probability measure $\mu(\boldsymbol{\xi}) = \delta(\|\boldsymbol{\xi}\|_{\mathbb{R}^3} - 1)\gamma(\boldsymbol{\xi}) C_\gamma$. Due to $\boldsymbol{E}$ being diagonal, it is possible to use different values of D for different frequencies in the approximation (41). However, for clarity, D is assumed to be the same for all frequencies.

Using (41) to calculate $\boldsymbol{K}$ as defined in (18) gives

$$(\boldsymbol{K})_{nm} \approx \frac{1}{D} \sum_{d=1}^{D} \langle \mathcal{F}^{-1} \boldsymbol{E}(\boldsymbol{r}_n, \boldsymbol{\xi}_d) \boldsymbol{E}(-\boldsymbol{r}_m, \boldsymbol{\xi}_d) \mathcal{F} \boldsymbol{\phi}(m), \boldsymbol{\phi}(n) \rangle_{\mathcal{V}_t}$$

Authorized licensed use limited to: KU Leuven University Library. Downloaded on September 29,2026 at 08:19:02 UTC from IEEE Xplore. Restrictions apply.

$$= \frac{1}{D}\sum_{d=1}^{D}\langle \boldsymbol{E}(-\boldsymbol{r}_m, \boldsymbol{\xi}_d)\boldsymbol{\phi}_f(m), \boldsymbol{E}(-\boldsymbol{r}_n, \boldsymbol{\xi}_d)\boldsymbol{\phi}_f(n)\rangle_{\mathcal{V}_\mathfrak{f}}. \tag{42}$$

The random basis vectors $\boldsymbol{v}_d(n) \in \mathcal{V}_\mathfrak{f}$ can be defined as $\boldsymbol{v}_d(n) = \frac{1}{\sqrt{D}}\boldsymbol{E}(-\boldsymbol{r}_n, \boldsymbol{\xi}_d)\boldsymbol{\phi}_f(n)$ which implies

$$(\boldsymbol{K})_{nm} \approx \sum_{d=1}^{D}\langle \boldsymbol{v}_d(n), \boldsymbol{v}_d(m)\rangle_{\mathcal{V}_\mathfrak{f}}. \tag{43}$$

Expression (43) can be simplified by making use of the isomorphism between $\mathcal{V}_\mathfrak{f}$ and $\mathbb{R}^L$. The invertible transform $\mathcal{S} : \mathcal{V}_\mathfrak{f} \to \mathbb{R}^L$ is used to separate the real and imaginary parts, i.e. as

$$\mathcal{S}\boldsymbol{b} = \Big(\sqrt{c_0}\mathfrak{Re}[b_0], \ldots, \sqrt{c_{\frac{L}{2}}}\mathfrak{Re}[b_{\frac{L}{2}}], \sqrt{c_1}\mathfrak{Im}[b_1], \ldots, \sqrt{c_{\frac{L}{2}-1}}\mathfrak{Im}[b_{\frac{L}{2}-1}]\Big), \tag{44}$$

assuming even L, with an analogous definition for odd L. The transform has the useful property that $\langle \boldsymbol{a}, \boldsymbol{b}\rangle_{\mathcal{V}_\mathfrak{f}} = \langle \mathcal{S}\boldsymbol{a}, \mathcal{S}\boldsymbol{b}\rangle_{\mathbb{R}^L}$. Defining $\tilde{\boldsymbol{v}}_d(n) = \mathcal{S}\boldsymbol{v}_d(n)$, (43) can be expressed as

$$(\boldsymbol{K})_{nm} \approx \sum_{d=1}^{D}\langle \tilde{\boldsymbol{v}}_d(n), \tilde{\boldsymbol{v}}_d(m)\rangle_{\mathbb{R}^L}. \tag{45}$$

Collecting the real-valued random basis vectors into a matrix $\boldsymbol{V} \in \mathbb{R}^{N\times LD}$ as

$$\boldsymbol{V} = \begin{bmatrix} \tilde{\boldsymbol{v}}_1(0)^\top & \ldots & \tilde{\boldsymbol{v}}_D(0)^\top \\ \vdots & \ddots & \vdots \\ \tilde{\boldsymbol{v}}_1(N-1)^\top & \ldots & \tilde{\boldsymbol{v}}_D(N-1)^\top \end{bmatrix} \tag{46}$$

leads to the concise approximation

$$\boldsymbol{K} \approx \boldsymbol{V}\boldsymbol{V}^\top. \tag{47}$$

A new set of parameters $\boldsymbol{b}_d \in \mathbb{R}^L$ can be defined as the parameter vector $\boldsymbol{a}$ transformed by the new basis as $\boldsymbol{b}_d = \sum_{n=0}^{N-1}\tilde{\boldsymbol{v}}_d(n)a_n$. More concisely, it can be expressed as $\boldsymbol{b} = \boldsymbol{V}^\top\boldsymbol{a}$ for $\boldsymbol{b} = (\boldsymbol{b}_1, \ldots, \boldsymbol{b}_D) \in \mathbb{R}^{LD}$. Then, inserting (47) into optimization problem (17) leads to

$$\underset{\boldsymbol{b}\in\mathbb{R}^{LD}}{\text{minimize}} \ \|\boldsymbol{p} - \boldsymbol{V}\boldsymbol{b}\|^2_{\mathbb{R}^N} + \lambda\|\boldsymbol{b}\|^2_{\mathbb{R}^{LD}}, \tag{48}$$

for which the optimal solution is

$$\boldsymbol{b} = (\boldsymbol{V}^\top\boldsymbol{V} + \lambda\boldsymbol{I})^{-1}\boldsymbol{V}^\top\boldsymbol{p}. \tag{49}$$

Finally, the RIR function estimate at an arbitrary point $\boldsymbol{r} \in \Omega$ is obtained as in (20), but adapted to the new basis. Inserting the kernel approximation (41) into (20) leads to

$$\tilde{\boldsymbol{u}}(\boldsymbol{r}) \approx \frac{1}{D}\sum_{d=1}^{D}\mathcal{F}^{-1}\boldsymbol{E}(\boldsymbol{r}, \boldsymbol{\xi}_d)\sum_{n=0}^{N-1}\boldsymbol{v}_d(n)a_n = \mathcal{F}^{-1}\frac{1}{\sqrt{D}}\sum_{d=1}^{D}\boldsymbol{E}(\boldsymbol{r}, \boldsymbol{\xi}_d)\mathcal{S}^{-1}\boldsymbol{b}_d. \tag{50}$$

The expression in (50) shows that each element in $\boldsymbol{b}$ has an interpretation as a plane wave coefficient. It becomes clearer if the reconstructed frequency-domain RIR function $\boldsymbol{u}(\boldsymbol{r}) = \mathcal{F}\tilde{\boldsymbol{u}}(\boldsymbol{r})$ is inspected, which is

$$(\boldsymbol{u}(\boldsymbol{r}))_l = \frac{1}{\sqrt{D}}\sum_{d=1}^{D}e^{-j\frac{\omega_l}{c}\boldsymbol{r}^\top\boldsymbol{\xi}_d}(\mathcal{S}^{-1}\boldsymbol{b}_d)_l \tag{51}$$

for $0 \le l < \frac{L}{2}$. It is clear that $(\mathcal{S}^{-1}\boldsymbol{b}_d)_l$ is the plane wave coefficient for the plane wave with direction $\boldsymbol{\xi}_d$ and frequency ω_l.

The RFF approach leads to a plane wave expansion method similar to [34], meaning that both the RFF approach and [34] can be viewed as approximations of KRR through the discretization of the Herglotz wave function (6). Their advantage is that the computational cost is dependent on the number of basis functions LD rather than the number of samples N, which is preferable for long measurement times. However, in contrast to [34], the RFF approach naturally extends to new kernels, such as those obtained by a regularization operator $\mathcal{R}$ distinct from the identity operator, or those constructed for related sound field estimation problems [41], [54]. In addition, approaches to further improve the effectiveness and reduce computational cost can be obtained from the many proposed extensions to the RFF scheme [45].

VI. Evaluation

In this section the proposed KRR and RFF methods will be evaluated, and will be compared against existing sound field estimation methods with stationary microphones and moving microphones. Simulated data will be used to characterize the performance of the methods, followed by real data to validate the performance in practice.

The estimation performance will primarily be evaluated in terms of the normalized mean square error (NMSE), defined as

$$\text{NMSE} = \frac{\sum_{\boldsymbol{r}\in\mathcal{E}}\|\tilde{\boldsymbol{u}}(\boldsymbol{r}) - \tilde{\boldsymbol{u}}^{\text{true}}(\boldsymbol{r})\|^2_{\mathcal{V}_\mathfrak{t}}}{\sum_{\boldsymbol{r}\in\mathcal{E}}\|\tilde{\boldsymbol{u}}^{\text{true}}(\boldsymbol{r})\|^2_{\mathcal{V}_\mathfrak{t}}}, \tag{52}$$

where $\mathcal{E} = \{\boldsymbol{r}_1, \ldots, \boldsymbol{r}_E\}$ is the set of the E evaluation points used in the experiment. In some cases the NMSE as a function of frequency is considered, in which the normalization is made frequency-wise, such that NMSE for frequency ω_l is

$$\text{NMSE}_{\omega_l} = \frac{\sum_{\boldsymbol{r}\in\mathcal{E}}|(\boldsymbol{u}(\boldsymbol{r}))_l - (\boldsymbol{u}^{\text{true}}(\boldsymbol{r}))_l|^2}{\sum_{\boldsymbol{r}\in\mathcal{E}}|(\boldsymbol{u}^{\text{true}}(\boldsymbol{r}))_l|^2}, \tag{53}$$

where $\boldsymbol{u}(\boldsymbol{r}) = \mathcal{F}\tilde{\boldsymbol{u}}(\boldsymbol{r})$. Due to the randomness of the RFF method, all NMSE values for RFF are obtained as the mean NMSE obtained from the same experiment performed 10 times with different random seeds used for sampling the basis functions. One standard error of the sample mean is displayed above and below the mean as a coloured band.

The source signal ϕ used in all experiments is a perfect periodic sweep [13], [28], a suitable signal for acoustic measurements. The signal is constructed to sweep from 0Hz to a maximum frequency of $\frac{f_s}{2}$Hz, over a period of L samples, which is the appropriate signal length for a perfect periodic sweep [13].

Authorized licensed use limited to: KU Leuven University Library. Downloaded on September 29,2026 at 08:19:02 UTC from IEEE Xplore. Restrictions apply.

A. Methods

A number of sound field estimation methods will be compared, including the proposed KRR and RFF, using stationary or moving microphones. Following is a list of the methods in addition to the names used to refer to them.

KRR-S: Using stationary microphones, the KRR method using the unweighted kernel (12). The method is often referred to as kernel interpolation with a diffuse kernel [37].

KRR-SD: Using stationary microphones, the KRR method using a directional weighting as shown in (28). The direction $\boldsymbol{\eta}$ is chosen as the direction from the source to the center of the region of interest.

KRR-M: Using moving microphones, the method is the proposed method described in Section IV without any weighting, i.e., $\mathcal{R}$ is chosen as the identity operator. This method is equivalent to the Bayesian method in [33] with omnidirectional microphones.

KRR-MD: Using moving microphones, the method is the proposed method described in Section IV using a frequency-wise directional weighting. The same weighting is used as KRR-SD.

RFF-S: Using stationary microphones, the RFF method proposed in Section V-A, using uniformly sampled basis directions. This method is the RFF approximation of KRR-S.

RFF-SD: Using stationary microphones, the RFF method proposed in Section V-A, using directionally informed sampling of the basis directions. This method is the RFF approximation of KRR-SD. Because the probability measure μ_ω with the chosen weighting function is the von Mises-Fisher distribution, it can be easily sampled from [55].

RFF-M: Using moving microphones, the RFF method proposed in Section V-B using uniformly sampled basis directions. This method is the RFF approximation of KRR-M.

RFF-MD: Using moving microphones, the RFF method proposed in Section V-B, using directionally informed sampling of the basis directions. The same distribution is used as RFF-SD. This method is the RFF approximation of KRR-MD.

Spatial spectrum: Using moving microphones, the method is based on a basis expansion using spherical wave functions [32]. The method can be viewed as a truncated version of the Bayesian method in [33]. Alternatively, it can be viewed as a spherical wave function counterpart to the plane wave basis expansion methods of [34] and RFF. The method has a single additional parameter $r_{\max}$, which is set according to the maximum radius of the trajectory.

Nearest neighbour: Using stationary microphones, the estimated RIR for any position is chosen as the measured RIR from the closest microphone, without making any changes to it. This method serves as an indication of the difficulty of the chosen estimation problem, and is a baseline which all considered methods should improve upon.

B. Random Fourier Features for Stationary Microphones

To begin with, the validity of the RFF method for stationary microphones will be demonstrated. To demonstrate each method's representation power, the experiment is constructed to have very favourable conditions for sound field estimation.

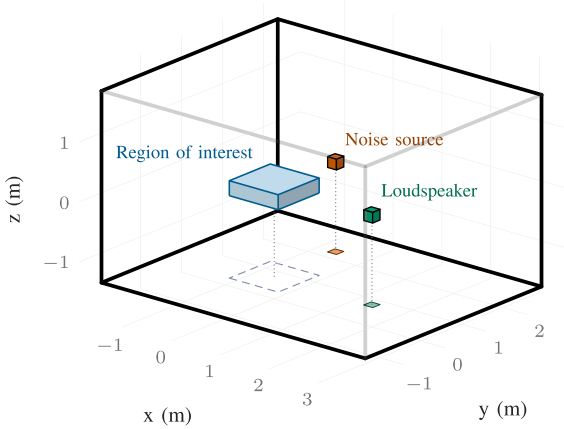


Fig. 1. The geometry of the simulated room, region of interest, loudspeaker and noise source. The loudspeaker and noise source are modelled as point sources in the experiments, but given non-zero volume here to make them visible in the illustration.

This also aligns with the moving microphone case where there are generally more data points compared to the number of basis functions. More realistic conditions will be considered for the moving microphone case, which is the focus of this paper.

The experiment is performed using RIRs simulated using the image-source method [56], [57] at a sampling rate of $f_s = 2000$ Hz. The RIRs are high-pass filtered at 20 Hz to simulate the low-frequency reduction in amplitude from a loudspeaker. The simulated cuboid room is of size $5.4 \times 4.3 \times 3.2\ \mathrm{m}^3$, within which the region of interest is a cuboid of size $1 \times 1 \times 0.25\ \mathrm{m}^3$. The geometry of the room and source is shown in Fig. 1.

Favourable conditions for sound field estimation are created by randomly placing 128 microphones with uniform distribution in the region of interest. The experiment is performed for 10 realizations of the microphone positions, one for each realization of the sampling of the basis functions. The reverberation time is $\mathrm{RT}_{60} = 0.12$ s. The RIRs at the microphone positions are used as data after transforming the RIRs into the frequency domain using a 1000-point DFT. The regularization parameter is set to $\lambda = 10^{-4}$.

The estimation performance for different numbers of basis functions can be seen in Fig. 2 as a function of frequency. The NMSE decreases as the number of basis functions increases, until it converges to the same error as the corresponding KRR estimate. KRR-S and KRR-SD consistently outperforms RFF-S and RFF-SD respectively for all frequencies, which is expected. The experiment validates the proposal of using randomly sampled plane waves as an approximation of the kernel function in (34).

A higher number of basis functions are necessary for higher frequencies. Given that a higher number of basis functions increases computational cost, this implies that it could be effective to choose D in a frequency-dependent manner. A comparison between the RFF-S and RFF-SD in Fig. 2 indicates that the directional weighting improves estimation performance when using a lower number of basis functions.

Authorized licensed use limited to: KU Leuven University Library. Downloaded on September 29,2026 at 08:19:02 UTC from IEEE Xplore. Restrictions apply.

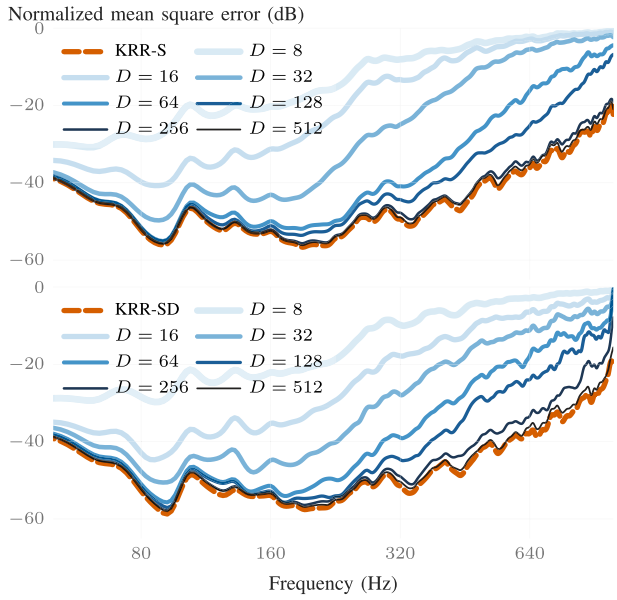


Fig. 2. Estimation performance for different numbers of random basis functions for stationary microphones. The figure shows that a higher number of basis functions are needed for higher frequencies. The non-weighted methods KRR-S and RFF-S are compared (top) and the directionally weighted methods KRR-SD and RFF-SD (bottom). RRF-S and RFF-SD with D basis functions are denoted by $D = 8, \ldots, D = 512$ in their respective plots.

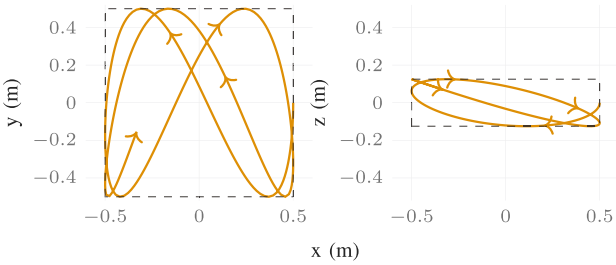


Fig. 3. The Lissajous trajectory for the moving microphone used in the simulations, shown projected onto a horizontal plane (left) and a vertical plane (right).

C. Random Fourier Features for Moving Microphones

The RFF method is of more practical use for the moving microphone case, where the number of data points and computational cost of KRR becomes high. In this section KRR-M, KRR-MD, RFF-M, and RFF-MD will be compared on simulated data in terms of estimation performance and computational cost. Except where mentioned explicitly, the same simulation parameters and geometry is used as in Section V-A. A sampling rate of $f_s = 1000$Hz is used, and the RIR length is $L = 500$ samples, corresponding to 0.5s. Following the approach of [32], [33], [34], the microphone moves along a Lissajous curve. The trajectory is normalized to a speed of 0.5m/s, which is shown in Fig. 3 for a measurement length of $N = 16000$. The regularization parameter is set to $\lambda = L \cdot 10^{-3}$.

The moving microphone data is simulated according to data model (2), with one simulated RIR per data point. While a more accurate simulation of a continuously moving microphone can be achieved by oversampling, it does not affect the presented results at the considered speed and maximum frequency. Further analysis of a sampled signal from a continuously moving microphone can be found in [21], and of the effect of position measurement errors in [20].

Estimates using RFF-M and RFF-MD are calculated for $D = 4, 8, 16$ and 32 and compared against KRR-M and KRR-MD respectively. The results are shown in Fig. 4, where the estimates are calculated using only the first $N = 1000, 2000, 4000, 8000,$ and 16000 samples of data of the trajectory shown in Fig. 3.

To demonstrate the relationship in computational cost between KRR and RFF, the mean clock time of calculating the estimates are shown in Fig. 4. The code was executed on the CPU of a laptop with an AMD Ryzen 7 PRO 4750 U processor. All methods are implemented in Python using JIT-compiled code in JAX [48], and are available at [58]. The computational cost of KRR-M and KRR-MD increases much faster compared to RFF-M and RFF-MD as the number of data points N increases. Especially KRR-MD is costly, and is therefore not calculated for $N = 16000$. In contrast, RFF-M and RFF-MD have essentially the same computational cost, that instead depends considerably on D.

While KRR-M and KRR-MD has higher computational cost than RFF-M and RFF-MD for large N, respectively, they provide more accurate estimates for any given N. In Fig. 4 the NMSE is shown, which demonstrates clearly that for a given amount of data, KRR-M and KRR-MD obtains lower NMSE than RFF-M and RFF-MD respectively. This means that there is a trade-off between estimation performance and computational cost that also depends on the amount of data available. Therefore, as a rule of thumb, RFF-M and RFF-MD can be considered advantageous for long acquisition times. The NMSE in Fig. 4 is calculated from the frequency bins between 20–480Hz, because the source has little energy for very low frequencies, and the estimates very close to the Nyquist frequency are not reliable, as Fig. 5 shows. Finally, the estimation performance as a function of frequency is shown in Fig. 5 for $N = 8000$. A similar pattern to RFF-S and RFF-D can be seen, where small values of D are sufficient to estimate low frequencies well, but not higher frequencies.

D. Estimation of Simulated Room Impulse Responses

Using a directional weighting for moving microphones is a new feature of the proposed method, so it will be considered more closely in this section. As the effectiveness of the chosen directional weighting depends on the directionality of the RIR function itself, the effect of the reverberation time will be investigated. In addition, the robustness of the different methods to the choice of regularization parameter will be shown. Except where explicitly mentioned, the same simulation parameters and geometry are used as in Section VI-C.

The first $N = 8000$ samples of the trajectory in Fig. 3 is used as data for the moving microphone, moving at the same speed of 0.5 m/s and is shown in Fig. 6. The $M = 16$ stationary microphones are placed in the center of each $L = 500$ length segment of the trajectory. This choice leads to a fair comparison with the stationary microphone methods, as the spatial sampling covers the same region. The estimates are evaluated on an equally

Authorized licensed use limited to: KU Leuven University Library. Downloaded on September 29,2026 at 08:19:02 UTC from IEEE Xplore. Restrictions apply.

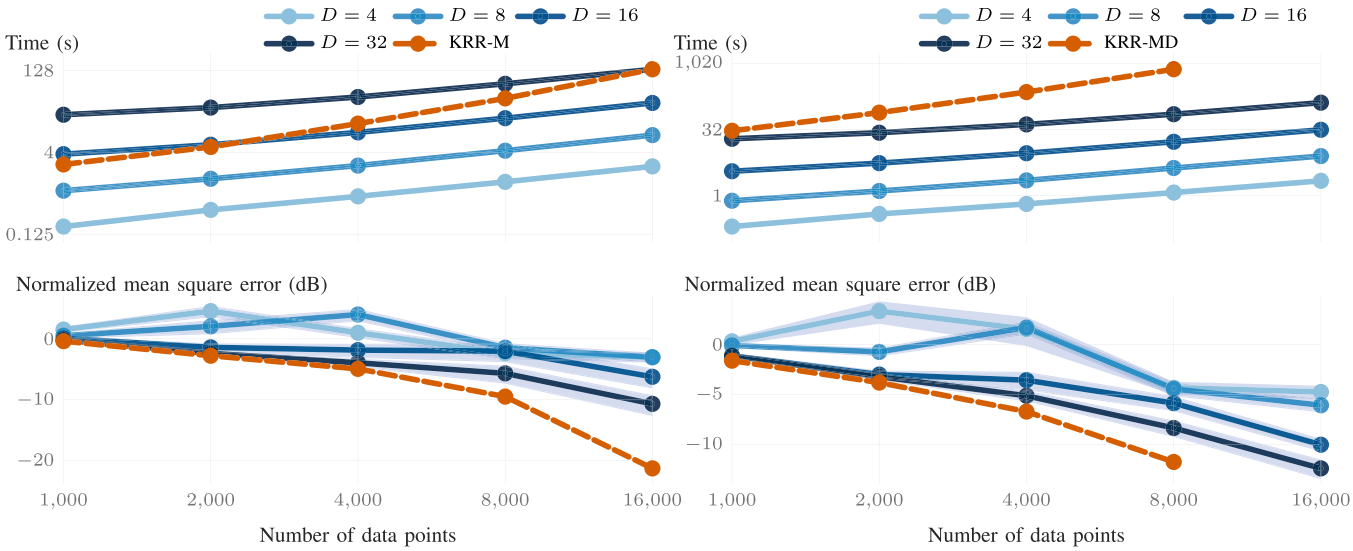


Fig. 4. The clock time (top) and the NMSE (bottom) for RFF-M and KRR-M (left) and RFF-MD and KRR-MD (right). RRF-M and RFF-MD with D basis functions are denoted by $D = 4, \ldots, D = 32$ in their respective plots.

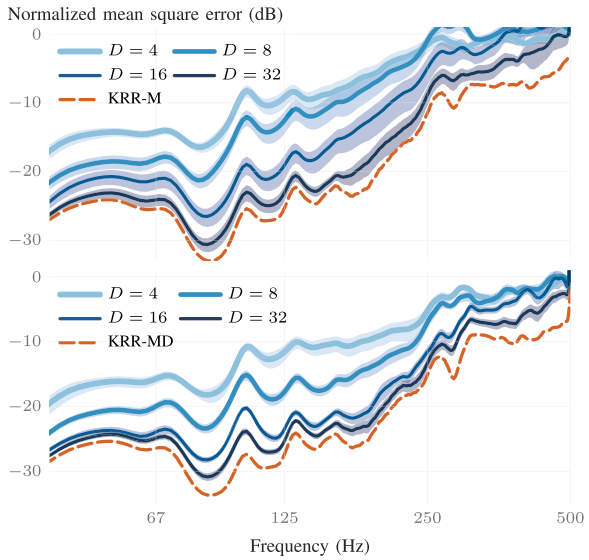


Fig. 5. Estimation performance for different numbers of random basis functions using moving microphones and $N = 8000$. The non-weighted methods KRR-M and RFF-M are compared (top) and the directionally weighted methods KRR-MD and RFF-MD (bottom). RRF-M and RFF-MD with D basis functions are denoted by $D = 4, \ldots, D = 32$ in their respective plots.

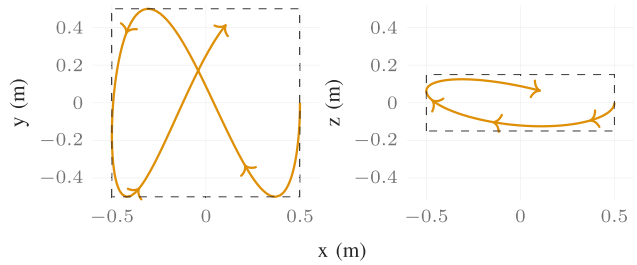


Fig. 6. The 3-dimensional trajectory used in the experiment, shown projected onto a horizontal plane (left) and a vertical plane (right).

spaced grid of $E = 2000$ evaluation points within the region of interest, where the distance between neighbouring points is 0.05 m.

Two noise models are considered. *White noise* refers to spatiotemporally white Gaussian noise added directly to the microphone signals. *Spatial noise* refers to the same noise propagated from a noise source that is placed in the room at $(0, 1.5, 0)$ m, shown in Fig. 1, leading to a spatially and temporally correlated noise. The noise for both noise models is scaled such that the microphone signals have a mean signal-to-noise ratio (SNR) of 30 dB

All considered methods except nearest neighbour require a scalar regularization parameter to be set. Estimates are calculated using the regularization parameter $\lambda = L\lambda_0$ for KRR-M, KRR-MD, RFF-M, and RFF-MD, and using the regularization parameter $\lambda = \lambda_0$ for the other methods, where $\lambda_0 = 10^t$ for $t = (-8, -7, \ldots, 0)$. A reverberation time of 0.2 s is used. The NMSE for all parameter values are shown in Fig. 8. The results indicate a difference in robustness, where KRR degrades considerably less compared to RFF and spatial spectrum when a suboptimal λ is chosen. All methods degrade significantly when λ is chosen too large. KRR-S and KRR-SD can be seen to be particularly robust to small λ when spatial noise is used. This is due to the microphones recording simultaneously, which means that all measurements contain the same noise up to a deterministic linear transform.

KRR-M is equivalent to the maximum a posteriori estimate of [33] if the regularization parameter is chosen as $\lambda = \frac{L\sigma_s^2}{\sigma_\alpha^2}$, where σ_s^2 is the power of the noise s, and σ_α^2 is the prior variance of the spherical wave function coefficients in [33]. Even though the value of σ_α^2 is not necessarily known, the expression suggests that a reasonable choice of the regularization parameter λ is proportional to the inverse of the SNR, which is particularly helpful in environments with time-varying noise power levels.

Authorized licensed use limited to: KU Leuven University Library. Downloaded on September 29,2026 at 08:19:02 UTC from IEEE Xplore. Restrictions apply.

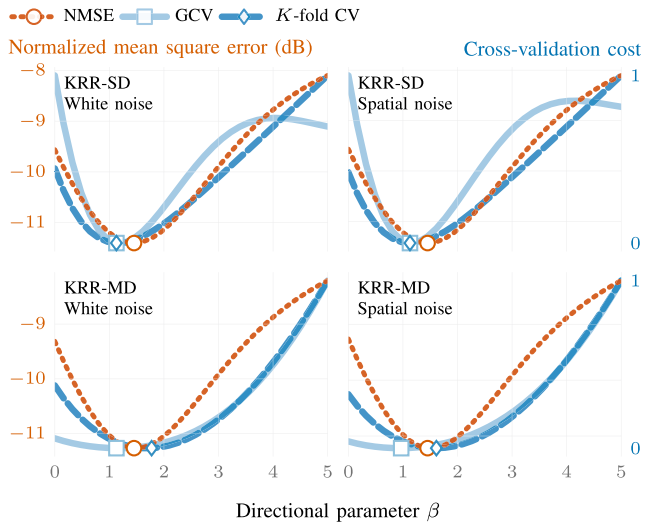


Fig. 7. Estimation performance compared to cross-validation cost for different values of β. The minimum NMSE and cross-validation cost is shown with a marker.

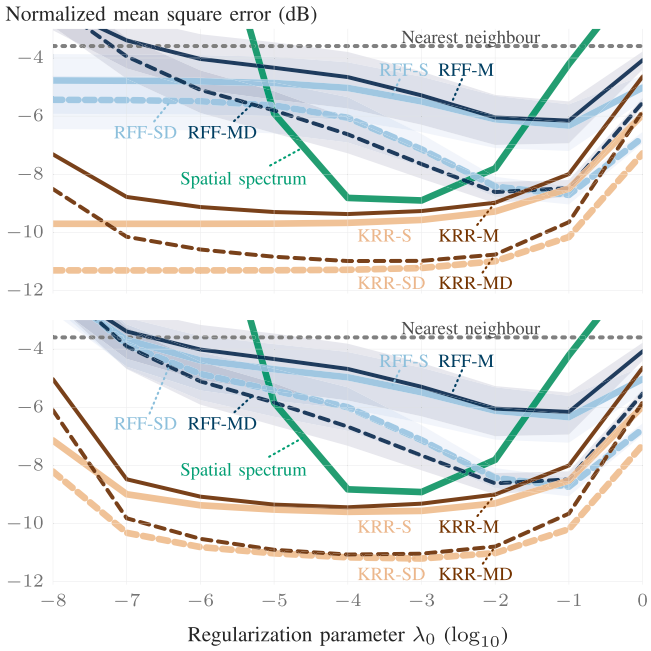


Fig. 8. Estimation performance as a function of the regularization parameter value for simulated data. The added noise is spatial noise (top) and white noise (bottom).

A similar connection exists between KRR-S and the Bayesian method in [35].

In practice, the parameters associated with the directional kernel, β and $\boldsymbol{\eta}$ must also be determined. Conventional techniques such as cross-validation can be applied to the proposed method. To demonstrate, using $\lambda_0 = 10^{-3}$ for all methods, β is obtained from the data using two common cross-validation methods, generalized cross-validation (GCV) [59] and K-fold cross-validation (K-fold CV) with $K = 4$ folds [60, Ch. 5]. Fig. 7 shows the cross-validation cost normalized between 0 and 1, together with the NMSE for KRR-SD and KRR-MD for both white noise and spatial noise. It is assumed that $\boldsymbol{\eta}$ is known, obtained from prior knowledge of the geometry, a source-localization method, or a principled learning approach such as [39].

To investigate the performance of the directional weighting, estimates are calculated for different reverberation times, for which the result is shown in Fig. 9. The reverberation times shown in Fig. 9 are the median RT_{60} between the source and the evaluation points as measured using the Schroeder integration method [61]. The directional parameter β is chosen the same for all frequencies, by calculating estimates using KRR-SD for 32 equally spaced values between 0 and 5, and using the β which has the least NMSE on the evaluation points. The used values of β are also shown in Fig. 9. The maximum value of 5 is arbitrarily chosen as a value larger than ideal for reverberant environments even with short reverberation times.

Fig. 9 shows that the chosen directional weighting is particularly effective at short reverberation times, regardless of noise type. When the RT_{60} is below 0.15s, all directionally weighted methods display better performance compared to the non-weighted methods. As the reverberation time increases, the improvement for KRR-MD and KRR-SD becomes less pronounced, although their estimates remain more accurate than KRR-M and KRR-S. This indicates the value of the proposed approach, as the estimation performance is robustly improved, even with a very simple directional weighting. To improve performance for long reverberation times, a more complicated weighting can be used which takes into account the directionality of the early reflections and the late reverberation [40].

Fig. 9 also shows that the accuracy of the estimates obtained from using stationary versus moving microphones are consistently very similar. While it is possible to find weighting functions using optimization methods such as [40] for the proposed moving microphone method, finding such a function can be computationally costly, a cost which is made considerably larger by the large cost of the original moving microphone estimation problems. Given the close connection between KRR for moving and stationary microphones, this opens up the possibility of finding suitable weighting functions using stationary microphones. Such a procedure could be done using only a few measurement microphones, or a surrogate simulation that is close enough to the real room. The weighting functions can then be applied to the moving microphone data, giving a good trade-off between speed, computational cost, and estimation performance.

E. Estimation of Real Room Impulse Responses

Finally, the performance of the proposed method will be demonstrated on real RIRs. The dataset from [20] is used, which has two microphones moving along circular trajectories of radii 0.45 m and 0.5 m. The dataset contains $E = 60$ evaluation points where RIRs are recorded using stationary microphones, which will be used as evaluation points. The trajectory, evaluation points, and source position are shown in Fig. 10, all of which are on the same horizontal plane in the room. Due to the similarity in performance between stationary and moving microphones

Authorized licensed use limited to: KU Leuven University Library. Downloaded on September 29,2026 at 08:19:02 UTC from IEEE Xplore. Restrictions apply.

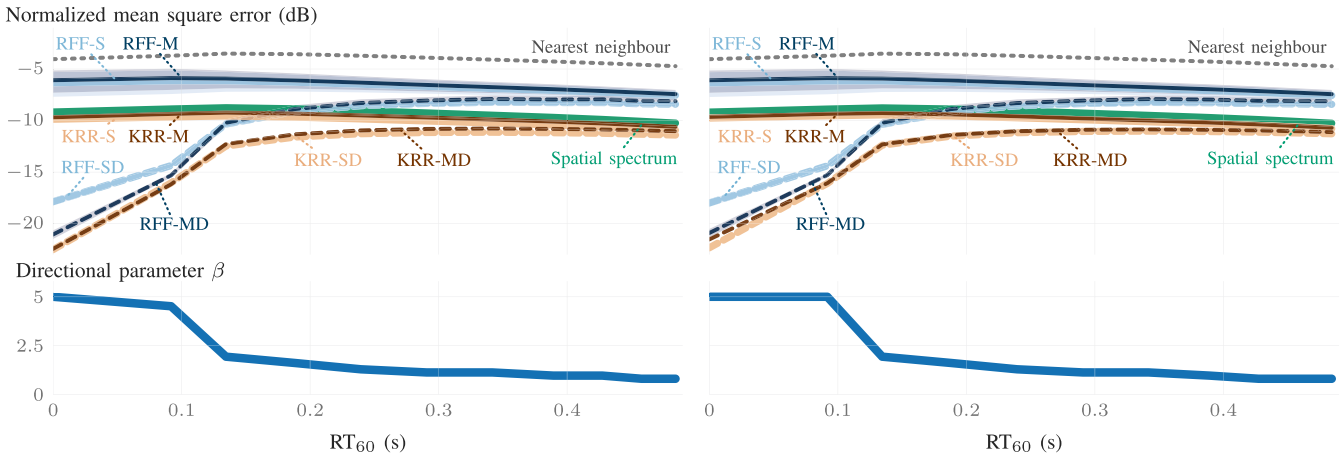


Fig. 9. Estimation performance for varying simulated reverberation times (top), and the used value of the parameter β (bottom).

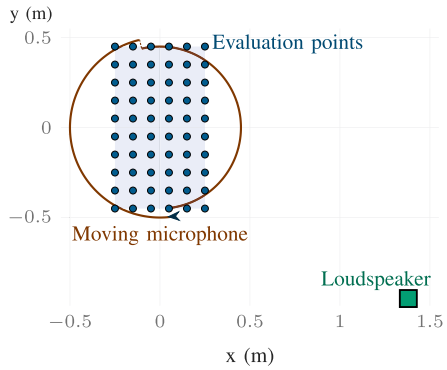


Fig. 10. Trajectory for the moving microphones and positions of the evaluation points used in the experiment with real data from [20].

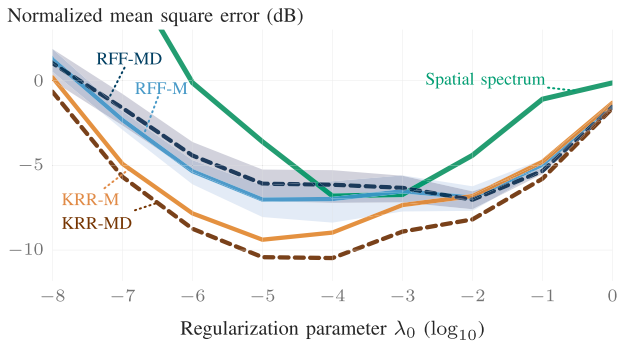


Fig. 11. Estimation performance as a function of the regularization parameter for the real data.

demonstrated in Section VI-D, and due to the data available in the dataset, only moving microphone methods are considered in this section. A sampling rate of $f_s = 1000$ Hz is used. The SNR for the used data is 21dB, which means that the data has more noise than the simulated data in Section VI-D.

The regularization parameter for the considered methods are once again investigated as described in Section VI-D, but for the real data. The estimation accuracy as a function of the regularization parameter is shown in Fig. 11. The regularization parameter value giving the lowest NMSE in Fig. 11 for each method is used for the remaining experiments.

TABLE I
NORMALIZED MEAN SQUARE ERROR IN DECIBELS FOR REAL DATA

KRR-M	KRR-MD	RFF-M	RFF-MD	Spatial spectrum
−9.15	**−10.23**	−6.32	−6.33	−6.76

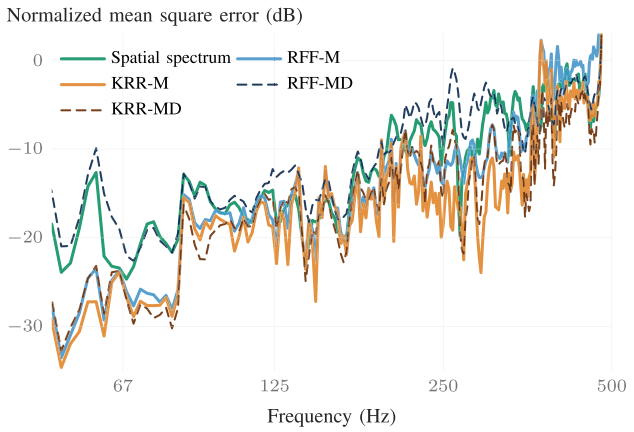


Fig. 12. Estimation performance as a function of frequency using real data. The standard error for RFF-M and RFF-MD is not displayed in this plot.

The NMSE of each method using the best regularization parameter is shown in Table I, and the NMSE as a function of frequency is shown in Fig. 12. The proposed KRR-MD has the lowest NMSE, indicating its applicability to real data. The KRR methods KRR-M and KRR-MD both perform better than the three finite basis expansion methods RFF-M, RFF-MD and spatial spectrum.

VII. CONCLUSION

A new method for sound field estimation using moving microphones has been proposed. The method is based on KRR, solving an optimization problem over a RKHS with functions representing time-domain sound fields, specifically RIRs as a

Authorized licensed use limited to: KU Leuven University Library. Downloaded on September 29,2026 at 08:19:02 UTC from IEEE Xplore. Restrictions apply.

function of receiver position. The KRR method is shown to be more robust and produce more accurate estimates compared to methods that project the data onto a finite set of basis functions. One additional benefit over existing methods is that the proposed KRR method allows for prior information to be incorporated in the form of a regularization penalty. Using a directional weighting in the regularization has been shown to be effective, obtaining the lowest error for both simulated and real data. An approximate KRR method has been proposed as well, using RFF to approximate the kernel. The RFF method has been shown to be less costly but less accurate compared to the KRR method, especially for large N.

Appendix A
Adjoint of Measurement Operator

The expression for the adjoint of the measurement operator $\mathcal{M}_n$ can be obtained from the definition of the adjoint $\langle \mathcal{M}_n \tilde{\boldsymbol{u}}, a\rangle_{\mathbb{R}} = \langle \tilde{\boldsymbol{u}}, \mathcal{M}_n^* a\rangle_{\tilde{\mathcal{H}}}$. Expanding the left-hand side gives

$$\begin{aligned}\langle \mathcal{M}_n \tilde{\boldsymbol{u}}, a\rangle_{\mathbb{R}} &= a\langle \tilde{\boldsymbol{u}}(\boldsymbol{r}_n), \boldsymbol{\phi}(n)\rangle_{\mathcal{V}_\mathfrak{t}} \\ &= a\int_{\mathbb{S}_2} \langle \mathcal{F}^{-1}\boldsymbol{E}(\boldsymbol{r}_n, \boldsymbol{d})\mathring{\boldsymbol{u}}(\boldsymbol{d}), \boldsymbol{\phi}(n)\rangle_{\mathcal{V}_\mathfrak{t}}\, ds(\boldsymbol{d}) \\ &= \int_{\mathbb{S}_2} \langle \mathring{\boldsymbol{u}}(\boldsymbol{d}), a\boldsymbol{E}(-\boldsymbol{r}_n, \boldsymbol{d})\mathcal{F}\boldsymbol{\phi}(n)\rangle_{\mathcal{V}_\mathfrak{f}}\, ds(\boldsymbol{d}) \\ &= \langle \tilde{\boldsymbol{u}}, \tilde{\boldsymbol{v}}_n\rangle_{\tilde{\mathcal{H}}}, \end{aligned} \tag{54}$$

where the directionality function of $\tilde{\boldsymbol{v}}_n = \mathcal{M}_n^* a$ is

$$\mathring{\boldsymbol{v}}_n = a\boldsymbol{E}(-\boldsymbol{r}_n, \boldsymbol{d})\mathcal{F}\boldsymbol{\phi}(n). \tag{55}$$

Inserting the directionality function $\mathring{\boldsymbol{v}}_n$ into the point evaluation expression in (6) gives

$$\begin{aligned}(\mathcal{M}_n^* a)(\boldsymbol{r}) &= a\mathcal{F}^{-1}\int_{\mathbb{S}_2} \boldsymbol{E}(\boldsymbol{r}, \boldsymbol{d})\boldsymbol{E}(-\boldsymbol{r}_n, \boldsymbol{d})\mathcal{F}\boldsymbol{\phi}(n)\, ds(\boldsymbol{d}) \\ &= a\tilde{\Gamma}(\boldsymbol{r}, \boldsymbol{r}_n)\boldsymbol{\phi}(n), \end{aligned} \tag{56}$$

which is the adjoint expression stated in (16).

Appendix B
Fourier Transform of Frequency-Domain Kernel

The probability measure is derived for kernels with a directional weighting function $\gamma : \mathbb{S}_2 \to \mathbb{R}$ [5], which is defined as

$$\kappa_\omega^\gamma(\boldsymbol{r}, \boldsymbol{r}') = \int_{\mathbb{S}_2} \gamma(\boldsymbol{d}) e^{-j\frac{\omega}{c}(\boldsymbol{r}-\boldsymbol{r}')^\top \boldsymbol{d}}\, ds(\boldsymbol{d}). \tag{57}$$

Applying Bochner's theorem (33) leads to

$$\int_{\mathbb{S}_2} \gamma(\boldsymbol{d}) e^{-j\frac{\omega}{c}(\boldsymbol{r}-\boldsymbol{r}')^\top \boldsymbol{d}}\, ds(\boldsymbol{d}) = \int_{\mathbb{R}^3} \mu_\omega(\boldsymbol{\xi}) e^{-j(\boldsymbol{r}-\boldsymbol{r}')^\top \boldsymbol{\xi}}\, d\boldsymbol{\xi}. \tag{58}$$

The probability measure μ_ω can be determined by inspection. The equality holds if the measure μ_ω is

$$\mu_\omega(\boldsymbol{\xi}) = \delta\left(\|\boldsymbol{\xi}\|_{\mathbb{R}^3} - \frac{\omega}{c}\right)\gamma(\boldsymbol{d})C_\gamma, \tag{59}$$

where C_γ is a normalization constant such that μ_ω is a probability measure. Because $\boldsymbol{d}$ is a unit vector scaled by $\frac{\omega}{c}$, $\boldsymbol{\xi}$ must always satisfy $\|\boldsymbol{\xi}\| = \frac{\omega}{c}$. The implication is that with a directionally weighted kernel, the plane wave basis should be sampled with directions according to the weighting.

When the weighting is set to $\gamma(\boldsymbol{d}) = 1$ for all $\boldsymbol{d} \in \mathbb{S}_2$, then the kernel κ_ω is obtained, due to the identity

$$\int_{\mathbb{S}_2} e^{-j\boldsymbol{r}^\top \boldsymbol{d}}\, ds(\boldsymbol{d}) = j_0(\sqrt{\boldsymbol{r}^\top \boldsymbol{r}}) \tag{60}$$

for any complex $\boldsymbol{r} \in \mathbb{C}^3$ [38, Eq. 65]. The associated probability measure is the uniform distribution on the sphere $\mathbb{S}_2$, defined explicitly as

$$\mu_\omega(\boldsymbol{\xi}) = \frac{1}{4\pi\frac{\omega^2}{c^2}}\delta\left(\|\boldsymbol{\xi}\|_{\mathbb{R}^3} - \frac{\omega}{c}\right). \tag{61}$$

Acknowledgment

The scientific responsibility is assumed by the authors.

References

[1] T. Betlehem, W. Zhang, M. A. Poletti, and T. D. Abhayapala, "Personal sound zones: Delivering interface-free audio to multiple listeners," *IEEE Signal Process. Mag.*, vol. 32, no. 2, pp. 81–91, Mar. 2015.
[2] W. Zhang, P. N. Samarasinghe, H. Chen, and T. D. Abhayapala, "Surround by sound: A review of spatial audio recording and reproduction," *Appl. Sci.*, vol. 7, no. 5, May 2017, Art. no. 532.
[3] T. Ajdler, L. Sbaiz, and M. Vetterli, "The plenacoustic function and its sampling," *IEEE Trans. Signal Process.*, vol. 54, no. 10, pp. 3790–3804, Oct. 2006.
[4] S. Kuo and D. Morgan, "Active noise control: A tutorial review," *Proc. IEEE*, vol. 87, no. 6, pp. 943–975, Jun. 1999.
[5] 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. Process.*, vol. 29, pp. 3052–3063, Aug. 2021.
[6] P. A. Naylor and N. D. Gaubitch, *Speech Dereverberation*. Berlin, Germany: Springer Sci. Bus. Media, Jul. 2010.
[7] T. van Waterschoot and M. Moonen, "Fifty years of acoustic feedback control: State of the art and future challenges," *Proc. IEEE*, vol. 99, no. 2, pp. 288–327, Feb. 2011.
[8] L. Savioja and U. P. Svensson, "Overview of geometrical room acoustic modeling techniques," *J. Acoust. Soc. Amer.*, vol. 138, no. 2, pp. 708–730, Aug. 2015.
[9] M. Vorländer, *Auralization: Fundamentals of Acoustics, Modelling, Simulation, Algorithms and Acoustic Virtual Reality (RWTHedition Series)*. Cham, Switzerland: Springer, 2020.
[10] I. Dokmanić, R. Parhizkar, A. Walther, Y. M. Lu, and M. Vetterli, "Acoustic echoes reveal room shape," *Proc. Nat. Acad. Sci. USA*, vol. 110, no. 30, pp. 12186–12191, Jul. 2013.
[11] A. Farina, "Simultaneous measurement of impulse response and distortion with a swept-sine technique," in *Proc. AES Conv.*, 2000.
[12] S. Müller and P. Massarani, "Transfer-function measurement with sweeps," *J. Audio Eng. Soc.*, vol. 49, no. 6, pp. 443–471, Jun. 2001.
[13] C. Antweiler, A. Telle, and P. Vary, "NLMS-type system identification of MISO systems with shifted perfect sequences," in *Proc. Int. Workshop Acoust. Signal Enhancement*, Seattle, WA, USA, 2008.
[14] T. Nishida, N. Ueno, S. Koyama, and H. Saruwatari, "Region-restricted sensor placement based on gaussian process for sound field estimation," *IEEE Trans. Signal Process.*, vol. 70, pp. 1718–1733, 2022.
[15] S. A. Verburg, F. Elvander, T. van Waterschoot, and E. Fernandez-Grande, "Optimal sensor placement for the spatial reconstruction of sound fields," *EURASIP J. Audio Speech Music Process.*, vol. 2024, no. 1, Aug. 2024, Art. no. 41.
[16] J. Unnikrishnan and M. Vetterli, "Sampling and reconstruction of spatial fields using mobile sensors," *IEEE Trans. Signal Process.*, vol. 61, no. 9, pp. 2328–2340, May 2013.

Authorized licensed use limited to: KU Leuven University Library. Downloaded on September 29,2026 at 08:19:02 UTC from IEEE Xplore. Restrictions apply.

[17] K. Gröchenig, J. L. Romero, J. Unnikrishnan, and M. Vetterli, "On minimal trajectories for mobile sampling of bandlimited fields," *Appl. Comput. Harmon. Anal.*, vol. 39, no. 3, pp. 487–510, Nov. 2015.
[18] F. Katzberg, M. Maass, and A. Mertins, "Coherence based trajectory optimization for compressive sensing of sound fields," in *Proc. Eur. Signal Process. Conf.*, 2021, pp. 116–120.
[19] F. Katzberg, M. Maass, R. Pallenberg, and A. Mertins, "Positional tracking of a moving microphone in reverberant scenes by applying perfect sequences to distributed loudspeakers," in *Proc. Int. Workshop Acoust. Signal Enhancement*, 2022, pp. 1–5.
[20] J. Brunnström, M. B. Møller, T. van Waterschoot, M. Moonen, and J. Østergaard, "Experimental validation of sound field estimation methods using moving microphones," in *Proc. Forum Acusticum*, Malaga, Spain, 2025, p. 16.
[21] F. Katzberg, M. Maass, and A. Mertins, "Doppler frequency analysis for sound-field sampling with moving microphones," *Front. Signal Process.*, vol. 4, Apr. 2024, Art. no. 1304069.
[22] T. Ajdler, L. Sbaiz, and M. Vetterli, "Dynamic measurement of room impulse responses using a moving microphone," *J. Acoust. Soc. Amer.*, vol. 122, no. 3, pp. 1636–1645, Sep. 2007.
[23] C. Antweiler, A. Telle, P. Vary, and G. Enzner, "Perfect-sweep NLMS for time-variant acoustic system identification," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2012, pp. 517–520.
[24] N. Hahn and S. Spors, "Comparison of continuous measurement techniques for spatial room impulse responses," in *Proc. Eur. Signal Process. Conf.*, Budapest, Hungary, 2016, pp. 1638–1642.
[25] N. Hahn and S. Spors, "Continuous measurement of spatial room impulse responses using a non-uniformly moving microphone," in *Proc. IEEE Int. Workshop Appl. Signal Process. Audio Acoust.*, New Paltz, NY, USA, 2017, pp. 205–208.
[26] N. Hahn and S. Spors, "Simultaneous measurement of spatial room impulse responses from multiple sound sources using a continuously moving microphone," in *Proc. Eur. Signal Process. Conf.*, 2018, pp. 2180–2184.
[27] K. MacWilliam, T. Dietzen, R. Ali, and T. van Waterschoot, "State-space estimation of spatially dynamic room impulse responses using a room acoustic model-based prior," *Front. Signal Process.*, vol. 4, Sep. 2024, Art. no. 1426082.
[28] C. Antweiler and M. Dörbecker, "Perfect sequence excitation of the NLMS algorithm and its application to acoustic echo control," *Ann. Télécommun.*, vol. 49, no. 7, pp. 386–397, Jul. 1994.
[29] F. Katzberg, R. Mazur, M. Maass, P. Koch, and A. Mertins, "Sound-field measurement with moving microphones," *J. Acoust. Soc. Amer.*, vol. 141, no. 5, pp. 3220–3235, May 2017.
[30] F. Katzberg, R. Mazur, M. Maass, P. Koch, and A. Mertins, "A compressed sensing framework for dynamic sound-field measurements," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 26, no. 11, pp. 1962–1975, Nov. 2018.
[31] E. G. Williams and J. A. Mann, *Fourier Acoustics: Sound Radiation and Nearfield Acoustical Holography*, vol. 108. Cambridge, MA, USA: Academic Press, 1999.
[32] F. Katzberg, M. Maass, and A. Mertins, "Spherical harmonic representation for dynamic sound-field measurements," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2021, pp. 426–430.
[33] J. Brunnström, M. B. Møller, and M. Moonen, "Bayesian sound field estimation using moving microphones," *IEEE Open J. Signal Process.*, vol. 6, pp. 312–322, 2025.
[34] S. Verburg, Y. Pene, and E. Fernandez-Grande, "Dynamic room impulse response measurements with a robotic arm," in *Proc. Forum Acusticum*, Malaga, Spain, 2025.
[35] N. Ueno, S. Koyama, and H. Saruwatari, "Sound field recording using distributed microphones based on harmonic analysis of infinite order," *IEEE Signal Process. Lett.*, vol. 25, no. 1, pp. 135–139, Jan. 2018.
[36] J. Brunnström, M. B. Møller, J. Østergaard, and M. Moonen, "Bayesian sound field estimation using uncertain data," in *Proc. Int. Workshop Acoust. Signal Enhancement*, Aalborg, Denmark, 2024, pp. 329–333.
[37] N. Ueno, S. Koyama, and H. Saruwatari, "Kernel ridge regression with constraint of Helmholtz equation for sound field interpolation," in *Proc. Int. Workshop Acoust. Signal Enhancement*, Tokyo, Japan, 2018, pp. 436–440.
[38] 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.
[39] R. Horiuchi, S. Koyama, J. G. C. Ribeiro, N. Ueno, and H. Saruwatari, "Kernel learning for sound field estimation with l1 and l2 regularizations," in *Proc. IEEE Int. Workshop Appl. Signal Process. Audio Acoust.*, 2021, pp. 261–265.
[40] J. Ribeiro, S. Koyama, and H. Saruwatari, "Sound field estimation based on physics-constrained kernel interpolation adapted to environment," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 32, pp. 4369–4383, 2024.
[41] J. G. C. Ribeiro, S. Koyama, and H. Saruwatari, "Physics-constrained adaptive kernel interpolation for region-to-region acoustic transfer function: A bayesian approach," *EURASIP J. Audio Speech Music Process.*, vol. 2024, Sep. 2024, Art. no. 43.
[42] R. Matsuda, J. G. C. Ribeiro, H. Akiyama, and J. Trevino, "Kernel ridge regression based sound field estimation using a rigid spherical microphone array," in *Proc. IEEE Int. Workshop Appl. Signal Process. Audio Acoust.*, 2025, pp. 1–5.
[43] J. Brunnström, M. B. Møller, J. Østergaard, S. Koyama, T. van Waterschoot, and M. Moonen, "Time-domain sound field estimation using kernel ridge regression," *IEEE Trans. Audio Speech Lang. Process.*, vol. 34, pp. 1243–1258, 2026.
[44] A. Rahimi and B. Recht, "Random features for large-scale kernel machines," in *Proc. Adv. Neural Inf. Process. Syst.*, 2007, pp. 1177–1184.
[45] F. Liu, X. Huang, Y. Chen, and J. A. K. Suykens, "Random features for kernel approximation: A survey on algorithms, theory, and beyond," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 10, pp. 7128–7148, Oct. 2022.
[46] J. Brunnström, M. B. Møller, J. Østergaard, T. van Waterschoot, M. Moonen, and F. Elvander, "Spatial covariance estimation for sound field reproduction using kernel ridge regression," in *Proc. Eur. Signal Process. Conf.*, Palermo, Italy, 2025, pp. 86–90.
[47] P. A. Martin, *Multiple Scattering: Interaction of Time-Harmonic Waves With N Obstacles* (Encyclopedia of Mathematics and Its Applications Series), vol. 107. Cambridge, U.K.: Cambridge Univ. Press, 2006.
[48] J. Bradbury et al., "JAX: Composable transformations of Python NumPy programs," 2018. [Online]. Available: github.com/jax-ml/jax
[49] C. R. Harris et al., "Array programming with NumPy," *Nature*, vol. 585, no. 7825, pp. 357–362, Sep. 2020.
[50] S. Diwale and C. Jones, "A generalized representer theorem for Hilbert space-valued functions," Sep. 2018, *arXiv:1809.07347*.
[51] J. Brunnström, S. Koyama, and M. Moonen, "Variable span trade-off filter for sound zone control with kernel interpolation weighting," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2022, pp. 1071–1075.
[52] K. Mardia and P. Jupp, *Directional Statistics*. Hoboken, NJ, USA: John Wiley Sons, Inc., 2000.
[53] C. C. Paige and M. A. Saunders, "LSQR: An algorithm for sparse linear equations and sparse least squares," *ACM Trans. Math. Softw.*, vol. 8, no. 1, pp. 43–71, Mar. 1982.
[54] R. Matsuda and M. Otani, "Sound field estimation for source-included region based on Gaussian process using prior source information," *J. Acoust. Soc. Amer.*, vol. 157, no. 2, pp. 1403–1417, Feb. 2025.
[55] W. Jakob, "Numerically stable sampling of the von mises fisher distribution on S2 (and other tricks)" Interactive Geometry Lab, ETH, Zürich, Switzerland, Tech. Rep., Jun. 2015. [Online]. Available: https://www.mitsubarenderer.org/~wenzel/files/vmf.pdf
[56] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *J. Acoust. Soc. Amer.*, vol. 65, no. 4, pp. 943–950, Apr. 1979.
[57] R. Scheibler, E. Bezzam, and I. Dokmanić, "Pyroomacoustics: A Python package for audio room simulation and array processing algorithms," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2018, pp. 351–355.
[58] J. Brunnström, "Audio signal processing collection (aspcol)," 2025. [Online]. Available: github.com/sounds-research/aspcol
[59] G. H. Golub, M. Heath, and G. Wahba, "Generalized cross-validation as a method for choosing a good ridge parameter," *Technometrics*, vol. 21, no. 2, pp. 215–223, May 1979.
[60] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning* (Adaptive Computation and Machine Learning Series). Cambridge, MA, USA: MIT Press, 2006.
[61] M. R. Schroeder, "New method of measuring reverberation time," *J. Acoust. Soc. Amer.*, vol. 37, no. 3, pp. 409–412, Mar. 1965.

Authorized licensed use limited to: KU Leuven University Library. Downloaded on September 29,2026 at 08:19:02 UTC from IEEE Xplore. Restrictions apply.