KU LEUVEN

STADIUS
Center for Dynamical Systems, Signal Processing and Data Analytics

Citation/Reference	Authors: Jesper Brunnström, Martin Bo Møller, Jan Østergaard, Toon van Waterschoot, Marc Moonen, Filip Elvander Title: Spatial covariance estimation for sound field reproduction using kernel ridge regression Journal/Conference/Book: European Signal Processing Conference (EUSIPCO) Vol/Book Chapter: Pages: ISBN:
Archived version	☐ Submitted version ■ Accepted (Author manuscript: the content is identical to the content of the published paper, but without the final typesetting by the publisher) ☐ Published ☐ Published (Open Access)
Acknowledgements	In this manuscript, the following funding is acknowledged: ■ EU H2020 ☐ EU Horizon Europe ☐ FWO ■ KULeuven Internal Funds ☐ VLAIO ☐ AI ☐ Other
Published version	
Copyright statement	
Author contact	jesper.brunnstroem@kuleuven.be

Spatial covariance estimation for sound field reproduction using kernel ridge regression

Jesper Brunnström* , Martin Bo Møller‡ , Jan Østergaard§ , Toon van Waterschoot* , Marc Moonen* and Filip Elvander†

*Dept. of Electrical Engineering (ESAT), STADIUS, KU Leuven, Leuven, Belgium

†Department of Information and Communications Engineering, Aalto University, Espoo, Finland

‡Bang & Olufsen, Acoustics R&D, Struer, Denmark

§Dept. of Electronic Systems, Aalborg University, Aalborg, Denmark

***Abstract*—Many sound field reproduction methods require knowledge of the spatial covariance, a quantity derived from room impulse responses in a region. The spatial covariance is commonly computed as the sample covariance, even though this often leads to poor estimates in practice due to spatial undersampling. This paper proposes a method for estimating sound fields generated by multiple sound sources, with the goal of improving spatial covariance estimation. The method includes prior knowledge, which allows for previous covariance estimates to be combined with new sound field measurements. A sound zone control problem is considered where the estimated spatial covariance is used to generate sound in a simulated room. The experiment shows that the proposed method provides an improvement in sound field reproduction performance when both sound field data and spatial covariance data can be combined.**

***Index Terms*—sound field estimation, spatial covariance, kernel ridge regression, sound field reproduction, spatial audio**

I. Introduction

Sound field reproduction tasks such as sound zone control [1] or spatial audio reproduction [2] require knowledge of the acoustic environment, in particular room impulse responses (RIRs) from each source to one or more regions in the room. Reproducing sound with incorrectly estimated RIRs can significantly affect the reproduction performance [3]–[5].

The spatial covariance, a quantity derived from the RIRs, appears naturally in many sound field reproduction methods [6]–[8]. Some methods do not use the RIRs at all, but rely only on the spatial covariance [9]–[11]. The spatial covariance used in sound field *reproduction* is a relationship between the emitted sound by several sound sources, averaged over a region in the room. This is in contrast to the spatial covariance in sound field *estimation*, which is a relationship between received sound at different positions [12]–[14].

The spatial covariance is commonly computed as the sample covariance of the RIRs at a set of measurement points. Such an approach requires a dense and evenly spaced set of measurements, which is time-consuming to collect. In addition, the variation of the acoustic environment over time means that the time-consuming process has to be repeated periodically. Sound field interpolation can be applied to mitigate the problem of too few measurements [15]. However, essential features are missing in existing methods, such as the combination of sound field measurements with prior knowledge of the spatial covariance.

For sound field estimation, it has been demonstrated that enforcing physical constraints leads to better estimates [12]. Because the spatial covariance is derived from sound fields, there is a set of physical constraints inherited from the sound fields which the spatial covariance should satisfy. Therefore, it could be advantageous to enforce these inherited constraints in the spatial covariance estimation process. In addition, the continuous region where the spatial covariance is defined should be taken into account, not only the discrete measurement points.

In this paper a spatial covariance estimation method using kernel ridge regression is proposed. The approach is a generalization of [12], estimating sound fields associated with multiple sources. In contrast to existing methods, the proposed methods allows for the inclusion of a prior spatial covariance in addition to the sound field data. The proposed method is evaluated on sound zone control in a simulated room, demonstrating improved sound zone control performance. Code associated with the proposed method is released under a permissive license at https://github.com/sounds-research/aspcol.

II. Problem statement

A. Data model

Consider a sound field reproduction system with S sources in a room, intended to reproduce sound in the source-free region $\Omega \subset \mathbb{R}^3$. Each source has an associated room impulse response to each point $\boldsymbol{r} \in \Omega$ for a given time period, which can be modelled as a finite impulse response filter that is assumed constant over the period of the estimation process. The sound fields associated with all S sources can be expressed as a vector-valued function $\boldsymbol{u} : \Omega \to \mathbb{C}^S$, representing a single frequency-domain component of the RIR as a function of $\boldsymbol{r} \in \Omega$. The quantities will henceforth be dependent on frequency, but the frequency will be omitted from the notation for brevity.

This research work was carried out at the ESAT Laboratory of KU Leuven, in the frame of Research Council KU Leuven C14-21-0075 "A holistic approach to the design of integrated and distributed digital signal processing algorithms for audio and speech communication devices". This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 956369: 'Service-Oriented Ubiquitous Network-Driven Sound — SOUNDS'. The scientific responsibility is assumed by its authors.

The sound field is measured at the M points $\boldsymbol{r}_1, \boldsymbol{r}_2, \ldots, \boldsymbol{r}_M \in \Omega$. All sources are not necessarily measured at each point, which is relevant when sources are added to the system at different times, such as for ad-hoc arrays. For point $\boldsymbol{r}_m$ the sound fields for S_m sources are measured, where $1 \leq S_m \leq S$. It can be represented by a selection matrix $\boldsymbol{P}_m \in \mathbb{R}^{S_m \times S}$, which is defined as

$$\boldsymbol{P}_m^\top = \begin{bmatrix} \boldsymbol{e}_{\mathcal{I}_1} & \boldsymbol{e}_{\mathcal{I}_2} & \ldots & \boldsymbol{e}_{\mathcal{I}_{S_m}} \end{bmatrix}, \tag{1}$$

where $\mathcal{I}$ is an index set of the sources measured at $\boldsymbol{r}_m$, and $\boldsymbol{e}_j \in \mathbb{R}^S$ for $j \in \mathbb{N}$ is the standard basis vector whose jth element is 1. The available data is

$$\boldsymbol{h}_m = \boldsymbol{P}_m \boldsymbol{u}(\boldsymbol{r}_m) + \boldsymbol{\epsilon}_m, \qquad m = 1, \ldots, M, \tag{2}$$

where $\boldsymbol{\epsilon}_m \in \mathbb{C}^{S_m}$ represents additive noise.

B. Spatial covariance

The considered spatial covariance $\boldsymbol{R} \in \mathcal{S}_+(S)$ is

$$\boldsymbol{R} = \frac{1}{|\Omega|} \int_\Omega \boldsymbol{u}(\boldsymbol{r}) \boldsymbol{u}(\boldsymbol{r})^\mathsf{H} \, d\boldsymbol{r}, \tag{3}$$

where $^\mathsf{H}$ denotes the Hermitian transpose, $|\Omega|$ denotes the volume of the region Ω, and $\mathcal{S}_+(S)$ refers to the set of positive definite matrices in $\mathbb{C}^{S \times S}$. Note that in the sound zone control literature, $\boldsymbol{R}$ is often defined as the complex conjugate of (3) [15, Eq. 13]. The task considered in this paper is to estimate the covariance (3) using data of the form (2).

III. Spatial covariance estimation

A. Sound field model

The sound field can be estimated by solving optimization problems on an appropriate reproducing kernel Hilbert space (RKHS) $\mathcal{H}$, which will be formulated in this section. A more detailed account of the model can be found in [12], [16]. A sound field $\boldsymbol{u} \in \mathcal{H}$ associated with all sources in a source-free region can be expressed using the Herglotz wave function as

$$\boldsymbol{u}(\boldsymbol{r}) = \int_{\mathbb{S}^2} e^{-ik\boldsymbol{r}^\top \hat{\boldsymbol{d}}} \bar{\boldsymbol{u}}(\hat{\boldsymbol{d}}) \, d\hat{\boldsymbol{d}}, \tag{4}$$

where i is the imaginary unit, and $\bar{\boldsymbol{u}} : \mathbb{S}^2 \to \mathbb{C}^S$ is a square-integrable function on the unit sphere $\mathbb{S}^2$ representing the complex weight of each plane wave. The direction $\hat{\boldsymbol{d}} \in \mathbb{S}^2$ is the incoming plane wave direction. The wavenumber is $k = \frac{\omega}{c}$ where ω is the angular frequency and c is the speed of sound.

The RKHS $\mathcal{H}$ can be constructed from the set of functions of the form (4) along with the inner product

$$\langle \boldsymbol{u}, \boldsymbol{v} \rangle_\mathcal{H} = \int_{\mathbb{S}^2} \bar{\boldsymbol{v}}(\hat{\boldsymbol{d}})^\mathsf{H} \boldsymbol{W}(\hat{\boldsymbol{d}})^{-1} \bar{\boldsymbol{u}}(\hat{\boldsymbol{d}}) \, d\hat{\boldsymbol{d}}, \tag{5}$$

where $\boldsymbol{W} : \mathbb{S}^2 \to \mathcal{S}_+(S)$ is a directional weighting function. This directional weighting is a generalization of the weighting considered in [17], [18], and equivalent when $\boldsymbol{W}$ is chosen to be diagonal. Exploring choices of $\boldsymbol{W}$ is not the focus of this paper, so a diagonal weighting will be assumed.

The RKHS $\mathcal{H}$ is characterized by its kernel function $\Gamma : \Omega \times \Omega \to \mathcal{S}_+(S)$, which satisfies the reproducing property $\langle \boldsymbol{u}(\boldsymbol{r}), \boldsymbol{c} \rangle_{\mathbb{C}^S} = \langle \boldsymbol{u}, \Gamma(\cdot, \boldsymbol{r})\boldsymbol{c} \rangle_\mathcal{H}$ for all $\boldsymbol{c} \in \mathbb{C}^S$, where $\langle \boldsymbol{a}, \boldsymbol{c} \rangle_{\mathbb{C}^S} = \boldsymbol{c}^\mathsf{H} \boldsymbol{a}$ is the standard Euclidean inner product. The kernel function is

$$\Gamma(\boldsymbol{r}, \boldsymbol{r}') = \int_{\mathbb{S}^2} \boldsymbol{W}(\hat{\boldsymbol{d}}) e^{-ik(\boldsymbol{r}-\boldsymbol{r}')^\top \hat{\boldsymbol{d}}} \, d\hat{\boldsymbol{d}}, \tag{6}$$

which can in general be solved by numerical integration. When no weighting is used, the kernel function reduces to $\Gamma(\boldsymbol{r}, \boldsymbol{r}') = \boldsymbol{I}_S \, j_0(k\|\boldsymbol{r} - \boldsymbol{r}'\|_2)$, where $\boldsymbol{I}_S \in \mathbb{R}^{S \times S}$ is the identity matrix and j_0 is the zeroth order spherical Bessel function of the first kind.

B. Kernel ridge regression

The sound field $\boldsymbol{u} \in \mathcal{H}$ can be estimated by solving the optimization problem

$$\min_{\boldsymbol{u} \in \mathcal{H}} \mathcal{C} = \sum_{m=1}^M \|\boldsymbol{h}_m - \boldsymbol{P}_m \boldsymbol{u}(\boldsymbol{r}_m)\|_2^2 + \lambda \|\boldsymbol{u}\|_\mathcal{H}^2. \tag{7}$$

Consider the linear evaluation operator $\mathcal{M}_{\boldsymbol{r}_m} : \mathcal{H} \to \mathbb{C}^{S_m}$ defined as $\mathcal{M}_{\boldsymbol{r}_m} \boldsymbol{u} = \boldsymbol{P}_m \boldsymbol{u}(\boldsymbol{r}_m)$. According to the representer theorem [19], a minimizer of (7) has the form

$$\hat{\boldsymbol{u}}(\boldsymbol{r}) = \sum_{m=1}^M \mathcal{M}_{\boldsymbol{r}_m}^* \boldsymbol{a}_m = \sum_{m=1}^M \Gamma(\boldsymbol{r}, \boldsymbol{r}_m) \boldsymbol{P}_m^\top \boldsymbol{a}_m, \tag{8}$$

where $\boldsymbol{a}_m \in \mathbb{C}^{S_m}$ is a set of unknown parameters, and $\mathcal{M}_{\boldsymbol{r}_m}^* : \mathbb{C}^{S_m} \to \mathcal{H}$ denotes the Hilbert space adjoint of $\mathcal{M}_{\boldsymbol{r}_m}$.

Inserting (8) into the optimization problem (7) results in a finite-dimensional optimization problem in terms of the parameters $\boldsymbol{a} = (\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_M) \in \mathbb{C}^{M_\text{eff}}$, where $M_\text{eff} = \sum_{m=1}^M S_m$. The optimal solution is

$$\boldsymbol{a} = (\boldsymbol{\Gamma} + \lambda \boldsymbol{I})^{-1} \boldsymbol{h}. \tag{9}$$

The vector $\boldsymbol{h} = (\boldsymbol{h}_1, \boldsymbol{h}_2, \ldots, \boldsymbol{h}_M) \in \mathbb{C}^{M_\text{eff}}$ represents all available data. The Gram matrix $\boldsymbol{\Gamma} \in \mathbb{C}^{M_\text{eff} \times M_\text{eff}}$ is defined by its m, m'th block $\boldsymbol{\Gamma}_{mm'} = \boldsymbol{P}_m \Gamma(\boldsymbol{r}_m, \boldsymbol{r}_{m'}) \boldsymbol{P}_{m'}^\top \in \mathbb{C}^{S_m \times S_{m'}}$.

The optimal parameter estimate (9) can be inserted into (8) to obtain a sound field estimate for any $\boldsymbol{r} \in \Omega$. When the directional weighting and hence the kernel function is diagonal, (9) is equivalent to applying the interpolation method in [12], [16] for each source individually.

C. Spatial covariance estimate

Assuming a diagonal kernel, the spatial covariance can be expressed in terms of the parameters as

$$\boldsymbol{R} = \boldsymbol{A} \boldsymbol{K} \boldsymbol{A}^\mathsf{H} \;\text{ with }\; \boldsymbol{K} = \frac{1}{|\Omega|} \int_\Omega \Gamma(\boldsymbol{r}) \Gamma(\boldsymbol{r})^\mathsf{H} \, d\boldsymbol{r}. \tag{10}$$

The kernel values are represented by $\Gamma(\boldsymbol{r}) \in \mathbb{C}^{M_\text{eff} \times S} = \begin{bmatrix} \Gamma(\boldsymbol{r}, \boldsymbol{r}_1)^\top \boldsymbol{P}_1^\top & \ldots & \Gamma(\boldsymbol{r}, \boldsymbol{r}_M)^\top \boldsymbol{P}_M^\top \end{bmatrix}^\top$, and the parameter matrix is $\boldsymbol{A} = \begin{bmatrix} \boldsymbol{P}_1^\top \operatorname{diag}\{\boldsymbol{a}_1\} & \ldots & \boldsymbol{P}_M^\top \operatorname{diag}\{\boldsymbol{a}_M\} \end{bmatrix} \in \mathbb{C}^{S \times M_\text{eff}}$. The kernel weighting matrix $\boldsymbol{K}$ can be calculated efficiently when $\boldsymbol{W}(\hat{\boldsymbol{d}}) = \boldsymbol{I}$ and Ω is a ball [20, Section 3.2], or with numerical integration otherwise.

IV. Spatial covariance estimation with prior knowledge

Given that the acoustic environment can change over time, there could be a previous estimate of the spatial covariance $\boldsymbol{R}_0$ available. Alternatively, there could be prior knowledge obtained from theoretical or data-driven methods. It would be desirable to use such prior information to obtain a better estimate with fewer measurements, reducing the effort in updating the estimates according to changes in the acoustic environment. Such an optimization problem can be written as

$$\min_{\boldsymbol{u}\in\mathcal{H}} \mathcal{C} + \gamma\mathcal{J}\Big(\frac{1}{|\Omega|}\int_\Omega \boldsymbol{u}(\boldsymbol{r})\boldsymbol{u}(\boldsymbol{r})^{\mathsf{H}}\,d\boldsymbol{r}, \boldsymbol{R}_0\Big), \tag{11}$$

where $\mathcal{C}$ is the cost function of (7), $\gamma \in \mathbb{R}_{\geq 0}$ is a weighting parameter, $\mathcal{J} : \mathcal{S}_+(S) \times \mathcal{S}_+(S) \to \mathbb{R}_{\geq 0}$ is a covariance fitting function, and $\boldsymbol{R}_0 \in \mathcal{S}_+(S)$ is the prior covariance.

As demonstrated in (8), the optimal solution of a kernel ridge regression problem is given by linear combinations of the kernel function evaluated at the same points as the function $\boldsymbol{u}$ is evaluated in the cost function. Due to the integral, (11) is a cost function with an infinite number of evaluations at all points in Ω, which will not lead to a practical solution following the same approach.

The proposed solution is to apply the representer theorem [19] to an approximation of the original optimization problem, defined as

$$\min_{\boldsymbol{u}\in\mathcal{H}} \quad \mathcal{C} + \gamma\mathcal{J}\Big(\frac{1}{V}\sum_{v=1}^{V}\boldsymbol{u}(\boldsymbol{r}_v)\boldsymbol{u}(\boldsymbol{r}_v)^{\mathsf{H}}, \boldsymbol{R}_0\Big), \tag{12}$$

for some set of points $\boldsymbol{r}_v \in \Omega$ for $v = 1, \ldots, V$. The resulting form of the function estimate instead of (8) is then

$$\hat{\boldsymbol{u}}(\boldsymbol{r}) = \sum_{m=1}^{M+V} \Gamma(\boldsymbol{r}, \boldsymbol{r}_m)\boldsymbol{P}_m^\top \boldsymbol{a}_m, \tag{13}$$

where $\boldsymbol{a}_m \in \mathbb{C}^S$ and $\boldsymbol{P}_m = \boldsymbol{I}_S$ for all $m = M+1, \ldots, M+V$. The points $\boldsymbol{r}_m$ for $1 \leq m \leq M$ are points in Ω where sound field measurements were taken, while the points for $M+1 \leq m \leq M+V$ are virtual points that can be chosen within Ω. The expression in (13) can be inserted into (11) to obtain a finite-dimensional optimization problem in terms of $\boldsymbol{a} \in \mathbb{C}^{M_{\text{eff}}+SV}$. The spatial covariance can then be computed as in (10), with $\Gamma(\boldsymbol{r}) \in \mathbb{C}^{(M_{\text{eff}}+SV)\times S}$ and $\boldsymbol{A} \in \mathbb{C}^{S\times(M_{\text{eff}}+SV)}$ defined analogously.

Two covariance fitting functions are considered, defined as

$$\begin{aligned}\mathcal{J}_{\text{Frob}}(\boldsymbol{R}, \boldsymbol{R}') &= \|\boldsymbol{R} - \boldsymbol{R}'\|_F^2 \\ \mathcal{J}_{\text{Wass}}(\boldsymbol{R}, \boldsymbol{R}') &= \text{tr}(\boldsymbol{R} + \boldsymbol{R}' - 2(\boldsymbol{R}'^{\frac{1}{2}}\boldsymbol{R}\boldsymbol{R}'^{\frac{1}{2}})^{\frac{1}{2}}).\end{aligned} \tag{14}$$

The squared Frobenius norm $\mathcal{J}_{\text{Frob}}$ is fast to compute, but is not a natural distance for positive definite matrices. The Wasserstein distance $\mathcal{J}_{\text{Wass}}$ is more complex to compute, but has a closer connection to the geometry of positive definite matrices [21].

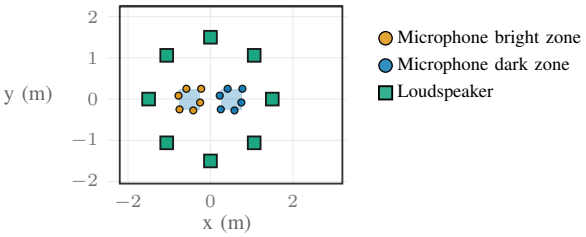


Fig. 1. Positions in the plane of the sources, microphones and the sound zones.

V. Evaluation

A. Simulation

Consider a scenario where careful sound field measurements are first made in a room, followed by a change to the acoustic environment, which requires the spatial covariance to be updated with new data. A few sound field measurements of the form (2) are collected in the new acoustic environment, which provide insufficient spatial sampling on their own. By combining the covariance data from the old environment and the sound field data from the new environment, the new covariance estimate should be improved. A simulation based experiment is performed, representing this scenario.

The first considered spatial covariance estimation method is referred to as *sound field data*, and computes the sample covariance from M sound field measurements of type (2) from the new environment. The second method, referred to as *kernel ridge regression (KRR)*, uses the same M sound field measurements from the new environment, but estimates the sound field as (9), followed by the spatial covariance estimate (10). *Covariance-informed kernel ridge regression (CIKRR)* is the proposed method. As a reference, the covariance data from the old environment can be used directly without any data from the new environment, which is referred to as *covariance data*.

Two cuboid regions of size $50 \times 50 \times 15\,\text{cm}$, referred to as the bright zone and dark zone, are placed in the interior of a circle of $S = 8$ sources. $M = 6$ microphones are placed on the boundary of each zone, the positions of which are shown in Fig 1. RIRs are generated using the image-source method, with a reverberation time of 0.36 s at a sampling rate of 1600 Hz [22], [23]. The data $\boldsymbol{h}_m$ is transformed from time-domain RIRs using a 800-point discrete Fourier transform before adding white Gaussian noise at a signal-to-noise ratio (SNR) of 40 dB.

The results are evaluated on an equally spaced grid E of 432 evaluation points with a spacing of 4 cm within each region. The sound field estimation error is measured by

$$\text{NMSE}_{\text{sf}} = \frac{\sum_{e\in E}\|\hat{\boldsymbol{u}}(\boldsymbol{r}_e) - \boldsymbol{u}(\boldsymbol{r}_e)\|_2^2}{\sum_{e\in E}\|\boldsymbol{u}(\boldsymbol{r}_e)\|_2^2} \tag{15}$$

and the covariance estimation error is measured by $\text{NMSE}_{\text{cov}} = \|\hat{\boldsymbol{R}} - \boldsymbol{R}\|_F^2/\|\boldsymbol{R}\|_F^2$. The true spatial covariance is computed as the sample covariance of the noise-free sound field at the evaluation points.

The regularization parameter is set $\lambda = 10^{-5}$. The points $\boldsymbol{r}_m$ for $m = M+1, \ldots, M+V$ are chosen randomly from a uniform distribution within Ω. The optimization problem (11)

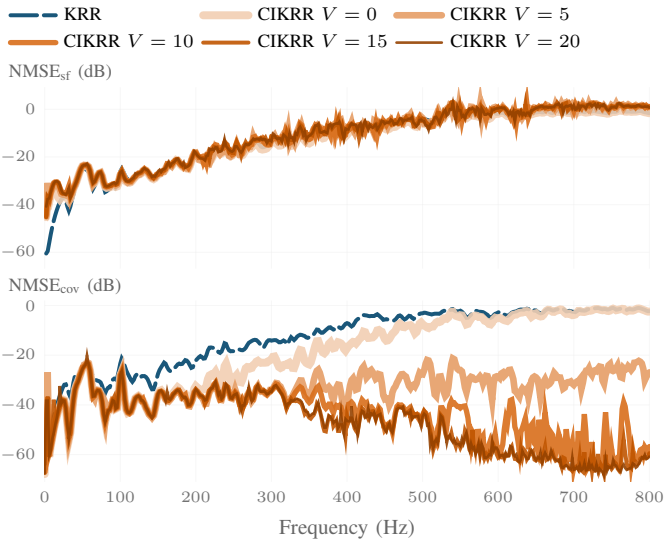


Fig. 2. Normalized mean square error of the sound field (top), and the spatial covariance (bottom) estimates for different values of the extra parameters V.

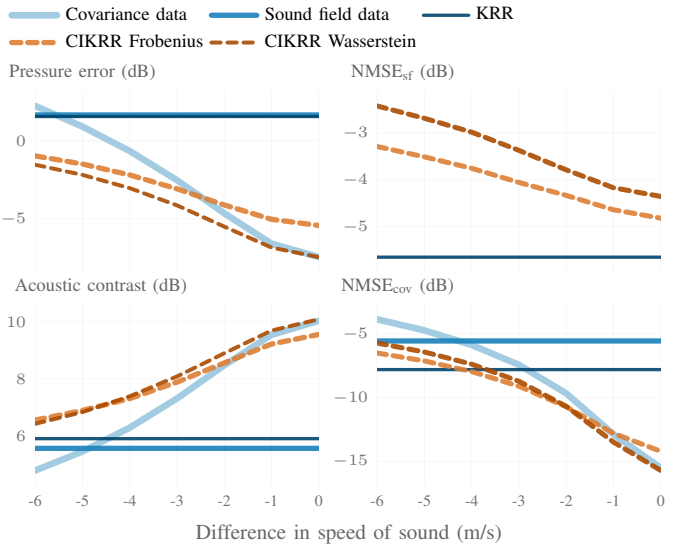


Fig. 3. Sound zone control performance (left) and estimation performance (right) as a function of the difference in the speed of sound for the environment where the covariance data is measured, compared to the speed of sound where the sound field data is measured and the methods are evaluated.

is solved using an ADAM optimizer with 10^5 steps and a step size of $2 \cdot 10^{-3}$ [24]. The matrix $\boldsymbol{K}$ is computed using Monte Carlo integration with 1000 samples.

B. Sound zone control method

The considered sound zone control method is given in [10], a method which only require spatial covariances, and is a generalization of pressure matching and acoustic contrast control. The source signals $\boldsymbol{y} \in \mathbb{C}^S$ are intended to be reproduced in the bright zone, but are generated without taking into account the leakage into the dark zone. The source signals are therefore modified using the filter $\boldsymbol{F} \in \mathbb{C}^{S \times S}$ as $\tilde{\boldsymbol{y}} = \boldsymbol{F}\boldsymbol{y}$ to be similar to $\boldsymbol{y}$ in the bright zone, while being close to 0 in the dark zone. The filter $\boldsymbol{F}$ that minimizes the pressure error is given by

$$\boldsymbol{F} = (\overline{\boldsymbol{R}_b} + \overline{\boldsymbol{R}_d} + \delta \boldsymbol{I})^{-1} \overline{\boldsymbol{R}_b}, \tag{16}$$

where $\overline{\ }$ denotes complex conjugate, and $\boldsymbol{R}_b, \boldsymbol{R}_d \in \mathcal{S}_+(S)$ the spatial covariances for the bright and dark zones.

The regularization parameter δ is chosen as $\delta = 10^{-2}\|\boldsymbol{R}_d\|_2$ for *sound field data* and *KRR* and $\delta = 10^{-3}\|\boldsymbol{R}_d\|_2$ for *covariance data* and *CIKRR*. The constants are of the form $10^{-\alpha}$ for integer α chosen to minimize the pressure error. The unmodified source signals $\boldsymbol{y}$ are generated as white Gaussian noise in the time-domain, independent for each source.

C. Experiment 1: Number of extra parameters V

The approximation of the integral in (13) raises the question of how to set the parameter V. It would be reasonable to believe that larger values leads to better performance, as the integral is more accurately approximated. For this experiment the proposed method using $\mathcal{J}_{\text{Frob}}$ as covariance fitting function is compared for different values of V, when the covariance data is the true covariance. The parameter γ is set to $\gamma = 1$.

In Fig. 2 the NMSE_{sf} and NMSE_{cov} are shown as a function of V for the bright zone. The extra degrees of freedom given by selecting a larger V allows the estimate to fit both sound field and covariance data accurately. The estimates quickly converge, suggesting that a small V provides most of the benefit, in turn leading to only a small increase in computational cost.

D. Experiment 2: Sound zone control

Changes to the speed of sound happen over time due to temperature or humidity changes, and degrades sound field reproduction performance. Given that the covariance data is degraded by the change to the speed of sound, the proposed method utilizes sound field measurements made after the change to maintain sound zone control performance. The approach can be contrasted with previous work where the speed of sound is explicitly compensated for [25].

The covariance data $\boldsymbol{R}_0$ is computed from 64 noise-free measurements taken at random points in the region, in a simulated room where the speed of sound is lower compared to the final value of 343 m/s. The sound field data is measured at the same speed of sound as the evaluation values, 343 m/s. Only the case where the speed of sound is increased is considered, as the results can be expected to be similar when the speed is decreased [26, Section 4.1]. The parameters V and γ are set to $V = 10$ and $\gamma = 10^{-3}$. The acoustic contrast and pressure error are computed from 5 s of signals.

The estimation and sound zone control performance are shown in Fig. 3, as a function of the speed of sound at which the covariance data was measured. The NMSE_{sf} is consistently higher for *CIKRR* compared to *KRR*. However, the spatial covariance estimates are consistently improved compared to the covariance data, which means that *CIKRR* provides the best covariance estimate when the covariance data is better or of similar quality to the sound field data. *CIKRR* with the

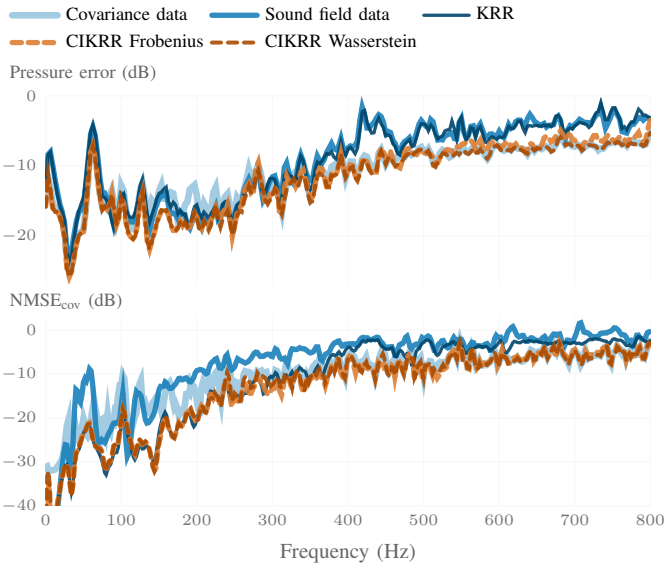


Fig. 4. Pressure error (top) and normalized mean square error for the spatial covariance estimate (bottom) when the covariance data is collected at −2 m/s compared to the evaluation environment.

Wasserstein metric achieves the lowest pressure error for all but the cases where the covariance data is of very high quality, in which case the results are similar to *covariance data*. As the quality of the covariance data degrades, *CIKRR* with both Wasserstein and Frobenius metric degrades slower compared to the data itself. Finally, the pressure error of *CIKRR* with the Wasserstein metric is equal or lower compared to the *covariance data*, and significantly improved over *sound field data* and *KRR*, as can be seen in Fig 4.

VI. Conclusion

A method for sound field estimation based on kernel ridge regression has been proposed, intended to be used for estimation of the spatial covariance required in sound field reproduction. In contrast to existing approaches, the proposed method allows for the combination of both sound field data and covariance data. The estimation and sound field reproduction performance has been evaluated using simulations, where it is shown that the proposed method provides improved results, especially when the covariance data is of similar or better quality compared to the sound field data.

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] M. Olsen and M. B. Møller, "Sound zones: On the effect of ambient temperature variations in feed-forward systems," in *Preprints AES Conv.* May 2017, pp. 1009–1018.

[4] M. B. Møller, J. K. Nielsen, E. Fernandez-Grande, and S. K. Olesen, "On the influence of transfer function noise on sound zone control in a room," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 27, no. 9, pp. 1405–1418, Sept. 2019.

[5] J. Zhang, L. Shi, M. G. Christensen, W. Zhang, L. Zhang, and J. Chen, "Robust pressure matching with ATF perturbation constraints for sound field control," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, May 2022, pp. 8712–8716.

[6] P. A. Nelson, "Active control of acoustic fields and the reproduction of sound," *J. Sound Vib.*, vol. 177, no. 4, pp. 447–477, Nov. 1994.

[7] T. Lee, L. Shi, J. K. Nielsen, and M. G. Christensen, "Fast generation of sound zones using variable span trade-off filters in the DFT-domain," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 29, pp. 363–378, Dec. 2020.

[8] V. Molés-Cases, S. J. Elliott, J. Cheer, G. Piñero, and A. Gonzalez, "Weighted pressure matching with windowed targets for personal sound zones," *J. Acoust. Soc. Am.*, vol. 151, no. 1, pp. 334–345, Jan. 2022.

[9] J.-W. Choi and Y.-H. Kim, "Generation of an acoustically bright zone with an illuminated region using multiple sources," *J. Acoust. Soc. Am.*, vol. 111, no. 4, pp. 1695–1700, Apr. 2002.

[10] J. Brunnström, T. van Waterschoot, and M. Moonen, "Sound zone control for arbitrary sound field reproduction methods," in *Proc. European Signal Process. Conf. (EUSIPCO)*, Sept. 2023, pp. 341–345.

[11] J. Brunnström, T. van Waterschoot, and M. Moonen, "Signal-to-interference-plus-noise ratio based optimization for sound zone control," *IEEE Open J. Signal Process.*, vol. 4, pp. 257–266, Feb. 2023.

[12] 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 (IWAENC)*. Sept. 2018, pp. 436–440.

[13] D. Caviedes-Nozal, N. A. B. Riis, F. M. Heuchel, J. Brunskog, P. Gerstoft, and E. Fernandez-Grande, "Gaussian processes for sound field reconstruction," *J. Acoust. Soc. Am.*, vol. 149, no. 2, pp. 1107–1119, Feb. 2021.

[14] D. Sundström, J. Lindström, and A. Jakobsson, "Recursive spatial covariance estimation with sparse priors for sound field interpolation," in *Proc. IEEE Stat. Signal Process. Workshop (SPP)*, July 2023, pp. 517–521.

[15] 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. (ICASSP)*, May 2022, pp. 1071–1075.

[16] 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, Apr. 2021.

[17] 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.

[18] 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, Sept. 2024.

[19] S. Diwale and C. Jones, "A generalized representer theorem for Hilbert space-valued functions," Sept. 2018, arXiv preprint, arXiv:1809.07347.

[20] H. Ito, S. Koyama, N. Ueno, and H. Saruwatari, "Three-dimensional spatial active noise control based on kernel-induced sound field interpolation," in *Proc. Int. Congr. Acoust. (ICA)*, Sept. 2019, pp. 1101–1108.

[21] L. Malagò, L. Montrucchio, and G. Pistone, "Wasserstein Riemannian geometry of Gaussian densities," *Info. Geo.*, vol. 1, no. 2, pp. 137–179, Dec. 2018.

[22] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *J. Acoust. Soc. Am.*, vol. 65, no. 4, pp. 943–950, Apr. 1979.

[23] R. Scheibler, E. Bezzam, and I. Dokmanić, "Pyroomacoustics: A Python package for audio room simulations and array processing algorithms," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Apr. 2018, pp. 351–355.

[24] J. Bradbury, R. Frostig, P. Hawkins, M. J. Johnson, C. Leary, D. Maclaurin, G. Necula, A. Paszke, J. VanderPlas, S. Wanderman-Milne, and Q. Zhang, "JAX: Composable transformations of Python+NumPy programs," 2018, [Online]. Available: http://github.com/jax-ml/jax.

[25] S. S. Bhattacharjee, J. R. Jensen, and M. G. Christensen, "Sound zone control robust to sound speed change," Oct. 2024, arXiv preprint, arXiv:2410.07978.

[26] M. B. Møller, *Sound Zone Control inside Spatially Confined Regions in Acoustic Enclosures*, Ph.D. thesis, Aalborg University, Sept. 2019.