ICASSP 2025 - 2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) | 979-8-3503-6874-1/25/$31.00 ©2025 IEEE | DOI: 10.1109/ICASSP49660.2025.10888108

A Zero-Shot Physics-Informed Dictionary Learning Approach for Sound Field Reconstruction

Stefano Damiano*†, Federico Miotello*‡, Mirco Pezzoli‡, Alberto Bernardini‡,
Fabio Antonacci‡, Augusto Sarti‡, Toon van Waterschoot†
*These authors contributed equally
†Dept. of Electrical Engineering (ESAT-STADIUS), KU Leuven, Leuven, Belgium
{stefano.damiano, toon.vanwaterschoot}@esat.kuleuven.be
‡Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano, Milan, Italy
{federico.miotello, mirco.pezzoli, alberto.bernardini, fabio.antonacci, augusto.sarti}@polimi.it

***Abstract*—Sound field reconstruction aims to estimate pressure fields in areas lacking direct measurements. Existing techniques often rely on strong assumptions or face challenges related to data availability or the explicit modeling of physical properties. To bridge these gaps, this study introduces a zero-shot, physics-informed dictionary learning approach to perform sound field reconstruction. Our method relies only on a few sparse measurements to learn a dictionary, without the need for additional training data. Moreover, by enforcing the Helmholtz equation during the optimization process, the proposed approach ensures that the reconstructed sound field is represented as a linear combination of a few physically meaningful atoms. Evaluations on real-world data show that our approach achieves comparable performance to state-of-the-art dictionary learning techniques, with the advantage of requiring only a few observations of the sound field and no training on a dataset.**

***Index Terms*—sound field reconstruction, dictionary learning, Helmholtz equation.**

I. Introduction

Sound field reconstruction is a fundamental task in acoustic engineering that focuses on estimating a pressure field in areas where direct measurements are unavailable. Directly measuring the acoustic field across a large volume and the entire audible frequency range requires a significant experimental effort, which is often impractical [1], [2]. Instead, sound field reconstruction techniques provide a more feasible solution by using a limited set of observations to interpolate and extrapolate the acoustic field [3]–[11]. This approach enables a precise acoustic representation of the environment, which is crucial for various applications, such as source localization [12], sound field control [13], auralization [14] and active noise cancellation [15].

Various techniques for sound field reconstruction are documented in the literature. Most approaches are based either on a parametric description of the acoustic scene [11], [16] or on models derived from the solutions to the wave equation [9], [17], such as plane waves [18], spherical waves [19], [20], or equivalent sources [3], [5], [6]. While offering robust solutions, these techniques often depend on strong assumptions, such as specific source propagation models, and frequently fall short in accurately modeling real-world scenarios, when the underlying assumptions are no longer valid.

Data-driven approaches try to overcome these limitations by leveraging on training processes carried out on data. One first class of such methods is based on modeling the acoustic field using neural networks. Deep learning models, such as convolutional neural networks [21], generative adversarial networks [22], and diffusion models [7], have been successful in learning the acoustic priors from training data and have been exploited to reconstruct pressure fields or address other acoustics-related tasks. Another class of data-driven methods is constituted by dictionary-learning techniques [23]. These approaches assume that a signal can be represented as a sparse combination of functions, called atoms, learned from observed data, and have demonstrated good performance and generalization capabilities in sound field reconstruction [4]. However, the amount of data required to train both deep learning and dictionary learning methods is substantial, and constitutes a bottleneck for their performance.

As a solution, some methods based on neural networks that do not require training on large datasets have also been proposed. For example, the deep prior technique [8], [24], [25] has shown effectiveness in modeling the acoustic prior of sound fields without needing training data. However, these methods are often unable to outperform data-driven techniques and may struggle to explicitly model the physical properties of sound fields. Similarly, physics-informed neural networks have been recently introduced to combine the explicit modeling capabilities of traditional model-based techniques with the flexibility of neural networks, while minimizing the amount of data required. These models have been effective in various acoustic problems by incorporating the physical equations governing sound propagation, such as the wave equation, to improve the accuracy [26]–[31]. While they produce physically meaningful solutions, they often lack interpretability.

To improve interpretability and to further integrate model-based techniques with data-driven methods, we propose a zero-shot physics-informed dictionary learning approach for sound field reconstruction. As a zero-shot technique, our solution only relies on few sparse measurements of a pressure field to learn a dictionary of atoms, with no additional data used for model training. Similarly to [32], [33], the physics-informed nature of our approach guides the dictionary learning process by encouraging the atoms to satisfy the underlying governing partial differential equation, i.e., the Helmholtz equation. By enforcing this acoustic prior information, the atoms do not need to be learned from additional training data other than the few observed measurements. This way, we are able to represent the sound field as a linear combination of a few physically meaningful atoms. Extensive evaluation on real-world data demonstrates the effectiveness of our method compared to a baseline based on a dictionary of Bessel functions and the state-of-the-art dictionary learning technique [4], that leverages on locally learned and data-driven functions.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 956962, from KU Leuven internal funds C3/23/056, and from FWO Research Project G0A0424N. This paper reflects only the authors' views and the Union is not liable for any use that may be made of the contained information.

This work was supported by the Italian Ministry of University and Research (MUR) under the National Recovery and Resilience Plan (NRRP), and by the European Union (EU) under the NextGenerationEU project.

Authorized licensed use limited to: KU Leuven Libraries. Downloaded on April 09,2025 at 12:19:59 UTC from IEEE Xplore. Restrictions apply.

II. Problem Formulation

A. Problem statement

Let us consider a source-free, two-dimensional region of space Ω lying on a horizontal plane in a reverberant room. The complex-valued frequency-domain acoustic pressure at position $\mathbf{r} \in \Omega$ and angular frequency $\omega = 2\pi f$, where f denotes the temporal frequency, is expressed by $p(\mathbf{r}, \omega)$. We sample Ω at N^2 points forming a square $N \times N$ grid at positions $\mathbf{r}_n$, with equidistant points along both directions and inter-point distance h. The pressure values $\mathbf{p} \in \mathbb{C}^{N^2}$ at the N^2 grid points can then be expressed as $\mathbf{p} = [p(\mathbf{r}_1, \omega), p(\mathbf{r}_2, \omega), \ldots, p(\mathbf{r}_{N^2}, \omega)]^\top$. We assume that only a limited number $M < N^2$ of measurements $\tilde{\mathbf{p}} \in \mathbb{C}^M$ is available, with $\tilde{\mathbf{p}}$ being a subset of $\mathbf{p}$. Our goal is to recover the complete set of pressure values $\mathbf{p}$ from the observations $\tilde{\mathbf{p}}$, using a zero-shot physics-informed dictionary learning algorithm.

B. Background

In dictionary-based methods for sound field reconstruction, the acoustic pressure $\tilde{\mathbf{p}}$ at a sparse set of measurement positions can be modeled as [4]

$$\tilde{\mathbf{p}} = \tilde{\mathbf{D}}\mathbf{x} + \tilde{\mathbf{e}}\,, \tag{1}$$

where $\tilde{\mathbf{D}} \in \mathbb{C}^{M \times L}$ is a dictionary of L position-dependent basis functions evaluated at M arbitrary positions on the grid $\tilde{\mathbf{r}}_m, 1 \leq m \leq M$, $\mathbf{x} \in \mathbb{C}^L$ is the coefficients vector containing the unknown weights of the atoms, and $\tilde{\mathbf{e}}$ denotes the measurement noise. The atoms in the dictionary $\tilde{\mathbf{D}}$ typically contain functions that represent solutions to the Helmholtz equation, e.g. plane waves [3], Green's functions [6], or, alternatively, can be learned from data [4]. In a first analysis stage, and given a known $\tilde{\mathbf{D}}$ dictionary, the weights $\mathbf{x}$ can be retrieved by solving the regularized inverse problem [4]

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \|\tilde{\mathbf{p}} - \tilde{\mathbf{D}}\mathbf{x}\|_2^2 + \alpha \|\mathbf{x}\|_p^p\,, \tag{2}$$

where the ℓ_p norm is defined as

$$\|\mathbf{x}\|_p^p = \sum_{l=1}^{L} |x_l|^p\,. \tag{3}$$

Well-known approaches use the ℓ_0 or ℓ_1 norms to promote sparsity in the solution [3], [5], [6], [34], assuming $\tilde{\mathbf{D}}$ to be overcomplete. In a second synthesis stage, $\mathbf{p}$ can then be reconstructed at all N^2 positions through

$$\mathbf{p} = \mathbf{D}\hat{\mathbf{x}} \tag{4}$$

where $\mathbf{D} \in \mathbb{C}^{N^2 \times L}$ contains the position-dependent basis functions evaluated at $\mathbf{r}_n, 1 \leq n \leq N^2$.

Dictionary learning relies on the assumption that a signal can be expressed as a sparse combination of functions, called atoms, learned from observed data [23]. As opposed to defining a closed-form dictionary $\mathbf{D}$ of basis functions based on some model of the sound field, the dictionary learning framework has been adopted in [4] to learn the dictionary atoms from an extensive set of training data. In the training phase, the overcomplete dictionary $\mathbf{D}$ and the corresponding training weights $\mathbf{x}_{\text{tr}}$ are jointly obtained from a set of training signals $\mathbf{p}_{\text{tr}}$ via the optimization problem

$$\hat{\mathbf{D}}, \hat{\mathbf{x}}_{\text{tr}} = \arg\min_{\mathbf{D}, \mathbf{x}_{\text{tr}}} \|\mathbf{p}_{\text{tr}} - \mathbf{D}\mathbf{x}_{\text{tr}}\|_2^2 + \alpha \|\mathbf{x}_{\text{tr}}\|_1\,, \tag{5}$$

solved by means of alternate optimization algorithms [35], [36]. The learned dictionary is then used to reconstruct the sound field in the evaluation setting using the available observations $\tilde{\mathbf{p}}$, by solving the inverse problem (2) with a sparsity-promoting norm to retrieve the weights $\hat{\mathbf{x}}$, finally plugged in (4) to obtain $\mathbf{p}$.

Inspired by image-processing techniques [37], in [4] patch-based processing is adopted to analyze the pressure field. The global domain Ω is divided in overlapping subdomains, where the pressure field is analyzed locally; the global sound field is then reassembled from the local reconstructions. Dictionary learning is adopted to jointly estimate a dictionary and the weights for each analyzed local patch. The model is pre-trained using real-world data recorded in a training room, and the dictionary is then used to effectively reconstruct the sound field in an unseen test room, where only a sparse set of measurements is available, by solving the ℓ_1-regularized inverse problem (2). Note that the pre-training requires data recorded at all N^2 grid points in the training room in order to obtain the dictionary that will be used to reconstruct the sound field in the test room. Despite its effectiveness, this approach requires a pre-training stage and, thus, training data. Moreover, since the dictionary atoms are entirely learned from data, they might lack a physical meaning as they do not necessarily follow the governing Helmholtz equation.

III. Proposed Method

To promote the retrieval of a physically meaningful dictionary, we introduce a zero-shot physics-informed dictionary learning algorithm that, differently from [4], operates at a global level (i.e., without considering local patches) and does not require pre-training. Inspired by [33], we encourage structure in the dictionary by enforcing each atom to satisfy the Helmholtz equation for a certain angular frequency ω. The Helmholtz equation can be expressed for each dictionary atom (i.e., column) $\mathbf{d}_l = \mathbf{D}_{(:,l)} \in \mathbb{C}^{N^2}$, with $l = 1, \ldots, L$, as [38]

$$(\nabla^2 + k^2)\mathbf{d}_l = \mathbf{0}\,, \tag{6}$$

where $k = \omega/c$ denotes the wave number and $c = 343\,\text{m/s}$ the speed of sound in air. If the distance h between adjacent grid points is sufficiently small, the Helmholtz equation can be discretized using the finite difference method (FDM) [33]. To this end, we first define the second-order difference matrix $\mathbf{H}(k) \in \mathbb{R}^{N^2 \times N^2}$, a symmetric, 5-diagonal Toeplitz matrix whose upper row $\mathbf{H}_{(1,:)}(k)$ is defined as

$$\mathbf{H}_{(1,:)}(k) = [\overbrace{-4 + k^2h^2 \quad 1 \quad 0 \quad \ldots \quad 1}^{N} \quad 0 \quad \ldots \quad 0]\,. \tag{7}$$

Note that the main diagonal elements of $\mathbf{H}$ depend on frequency through the wave number k. Then, (6) can be expressed as

$$\mathbf{H}(k)\mathbf{d}_l = \mathbf{0}\,. \tag{8}$$

With this definition, and considering a set of L angular frequencies ω_l, we modify (5) to introduce a regularization term that enforces the atoms of $\mathbf{D}$ to be solutions of the Helmholtz equation as

$$\begin{aligned}\hat{\mathbf{D}}, \hat{\mathbf{x}} = \arg\min_{\mathbf{D}, \mathbf{x}} &\|\tilde{\mathbf{p}} - \tilde{\mathbf{D}}\mathbf{x}\|_2^2 + \\ &+ \alpha\|\mathbf{x}\|_1 + \beta \sum_{l=1}^{L} \|\mathbf{H}(\frac{\omega_l}{c})\mathbf{d}_l\|_2^2\,.\end{aligned} \tag{9}$$

Note that $\tilde{\mathbf{D}}$ is a subset of $\mathbf{D}$, containing the M rows corresponding to the positions where measurements are available. Without the Helmholtz regularization term, only these rows would be updated and, thus, the learned dictionary would be meaningless and lead to a poor reconstruction. The Helmholtz regularization term encourages each atom of the *full* dictionary $\mathbf{D}$ to represent a solution of the Helmholtz equation at a specific frequency ω_l: this way, the entire dictionary gets updated, capturing a wideband representation of the

Authorized licensed use limited to: KU Leuven Libraries. Downloaded on April 09,2025 at 12:19:59 UTC from IEEE Xplore. Restrictions apply.

Algorithm 1: The Helmholtz-regularized dictionary learning algorithm

Initialize $\mathbf{D}$, $\mathbf{x}$
Set α, β, L, I
for $i = 0$ to I **do**
Sparse coding:
$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \|\tilde{\mathbf{p}} - \tilde{\mathbf{D}}\mathbf{x}\|_2^2 + \alpha\|\mathbf{x}\|_1$
Dictionary update:
$\hat{\mathbf{D}} = \arg\min_{\mathbf{D}} \|\tilde{\mathbf{p}} - \tilde{\mathbf{D}}\mathbf{x}\|_2^2 + \beta \sum_{l=1}^{L} \|\mathbf{H}(\frac{\omega_l}{c})\mathbf{d}_l\|_2^2$
end

sound field. If all frequencies ω_l fall within a given range, the learned dictionary can be effectively used to reconstruct the sound field at unseen frequencies within the same range, as will be discussed in Sec. IV. Minimizing (9) ensures that the solution remains sparse in terms of the coefficients $\mathbf{x}$. Since both $\mathbf{D}$ and $\mathbf{x}$ are unknown, the solution to the nonconvex problem in (9) is found using an alternating optimization approach [4]. This process consists of two separate update steps. First, in the sparse coding step, the optimal coefficients $\mathbf{x}$ are determined for a fixed dictionary $\mathbf{D}$ according to (2). Then, in the dictionary update step, each atom of the dictionary $\mathbf{D}$ is optimized by minimizing the discrepancy between the measured pressure field and the reconstructed field at positions $\tilde{\mathbf{r}}_m$. The two steps are repeated until a maximum number of iterations I is reached. The algorithm is summarized in Algorithm 1.

Unlike the dictionary learning method [4], our approach is *zero-shot*, as it does not involve any pre-training on datasets. In fact, the algorithm is provided only with the observed pressure field $\tilde{\mathbf{p}}$; both the dictionary and the weights are jointly learned using this information alone, along with the Helmholtz regularization, that influences the dictionary update step (while the sparse coding remains unchanged). Finally, once the dictionary $\mathbf{D}$ and weights $\mathbf{x}$ are learned, the sound field in the whole evaluation region is reconstructed via (4).

IV. Evaluation

We conduct an evaluation campaign based on experimental data to assess the performance of the proposed method. The experiments rely on the dataset introduced in [1]: the measurements are taken in a classroom with dimensions $[6.61, 9.45, 2.97]\mathrm{m}$ having non-flat walls, an acoustically absorbing ceiling and containing scattering objects (e.g., desks and chairs), with $T_{60} = 0.5\,\mathrm{s}$. A single source is located in a room corner and impulse responses are recorded on a square grid of 69×69 points with $h = 2.5\,\mathrm{cm}$ spacing, over a planar aperture with size $1.7\,\mathrm{m} \times 1.7\,\mathrm{m}$, centered at position $[3.78, 1.90, 1.86]\mathrm{m}$ (further details can be found in [4]).

To assess the performance of the proposed method, we compare it with a baseline (BL) and to the state-of-the-art dictionary learning sound field reconstruction method [4] (OLDL). The baseline adopts a fixed dictionary inspired by the *synthesized* dictionary used in [4]. The frequency-dependent BL dictionary $\mathbf{D}_{\text{bl}}$ is built using $L = 21$ atoms, drawn from (frequency-dependent) multivariate normal distributions with covariance matrix $\mathbf{\Sigma} \in \mathbb{R}^{N \times N}$, whose entry at position s, t is obtained by [4]

$$\Sigma_{s,t} = j_0(k\|\mathbf{r}_s - \mathbf{r}_t\|_2)\,, \tag{10}$$

where j_0 is the zero-order spherical Bessel function. Similarly to the proposed method, each atom $\mathbf{d}_l$ is generated using spherical Bessel functions at frequency ω_l (thus, each atom represents a different frequency).

The OLDL method is trained using the procedure described in [4], based on analyzing the sound field in local patches and pre-training the dictionary $\mathbf{D}_{\text{oldl}}$ in a training room[1].

We implement the proposed method in Python[2], and use the CVXPY [39] library to solve the optimization problem. For each experiment, we use a dictionary of $L = 21$ atoms and a range of frequencies of 200 Hz for the Helmholtz regularization: this results in atoms modeling frequencies spaced by 10 Hz. Preliminary analysis showed that increasing the number of atoms does not lead to an improved performance. We set the maximum number of iterations $I = 40$ for solving the optimization problem, and set the hyper-parameters $\alpha = 1$ and $\beta = 0.1$, obtained via a grid search. To emphasize the benefit of the physics-informed learning procedure, the dictionary $\mathbf{D}$ is initialized using the baseline dictionary $\mathbf{D}_{\text{bl}}$.

A. Metrics

We evaluate the performance of the proposed method based on two metrics: the normalized mean-squared error (NMSE) in dB scale

$$\mathrm{NMSE} = 10\log_{10}\left(\frac{\|\mathbf{p} - \hat{\mathbf{p}}\|_2^2}{\|\mathbf{p}\|_2^2}\right)\,, \tag{11}$$

where $\hat{\mathbf{p}} \in \mathbb{C}^{N^2}$ denotes the estimated pressure at the N^2 points of the evaluation grid, and the normalized cross-correlation (NCC)

$$\mathrm{NCC} = \frac{|\hat{\mathbf{p}}^{\mathrm{H}}\mathbf{p}|}{\|\hat{\mathbf{p}}\|_2^2\|\mathbf{p}\|_2^2}\,, \tag{12}$$

where $(\cdot)^{\mathrm{H}}$ indicates the Hermitian transpose. Specifically, NCC produces a value ranging from -1 to 1, where 1, 0, and -1 correspond to perfect similarity, no correlation, and perfect anti-correlation, respectively [31].

B. Results

To ensure a robust evaluation, we examine the behavior of the proposed, BL and OLDL [4] methods across varying frequencies and different numbers of available measurements M. All experiments are repeated 5 times, using randomly extracted configurations of measurement microphones M, and the average and standard deviation of the results over the 5 folds are reported for each experiment.

We first consider a fixed number of measurement microphones $M = 50$ and analyze the performance of the proposed method across frequencies, considering the range from 600 Hz to 1200 Hz, chosen to allow for a direct comparison with [4]. The entire frequency range is analyzed in the sub-bands $[500, 700]$Hz, $[700, 900]$Hz, $[900, 1100]$Hz, $[1100, 1300]$Hz. For each sub-band, a different dictionary of L atoms is generated for the BL, OLDL and proposed methods, and the reconstruction is performed at 81 frequency (i.e., every 2.5 Hz) within the same sub-band; we report the average of the results across each frequency band. Note that, since the atoms represent frequencies spaced by 10 Hz, the reconstruction is thus performed also at non-directly-modeled frequencies, where an interpolation between dictionary atoms is implicitly performed during the reconstruction.

In Fig. 1 we report the NMSE and the NCC across the considered frequency range. The performance of all methods degrades as frequency increases: for a fixed number of microphones M this is to be expected, as the number of measurements per wavelength decreases as frequency increases, making the problem more challenging. The

[1]Code available at https://github.com/manvhah/local_soundfield_reconstruction

[2]Code available at https://github.com/steDamiano/PIDL-sound-field-reconstruction

Authorized licensed use limited to: KU Leuven Libraries. Downloaded on April 09,2025 at 12:19:59 UTC from IEEE Xplore. Restrictions apply.

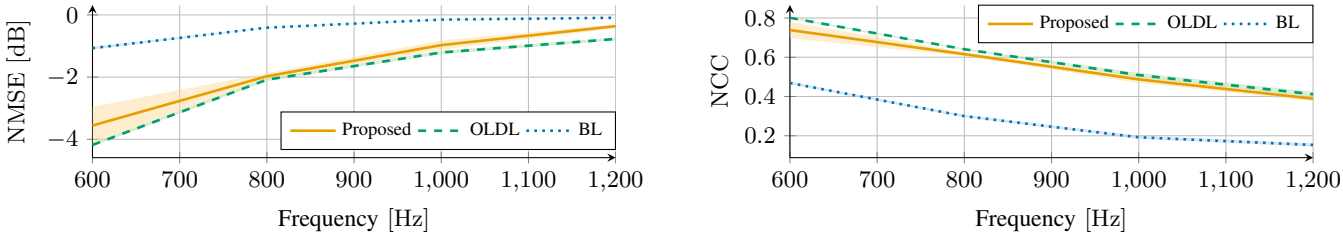


Fig. 1: Average (solid line) and standard deviation (shaded area) of the NMSE and NCC over the 5 folds, with respect to frequency. The number of available measurements is $M = 50$.

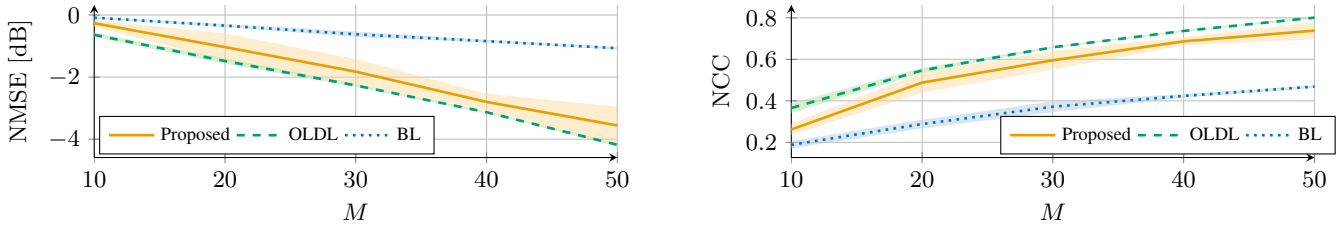


Fig. 2: Average (solid line) and standard deviation (shaded area) of the NMSE and NCC over the 5 folds, with respect to the number of available measurements M, in the frequency range $[500, 700]$Hz.

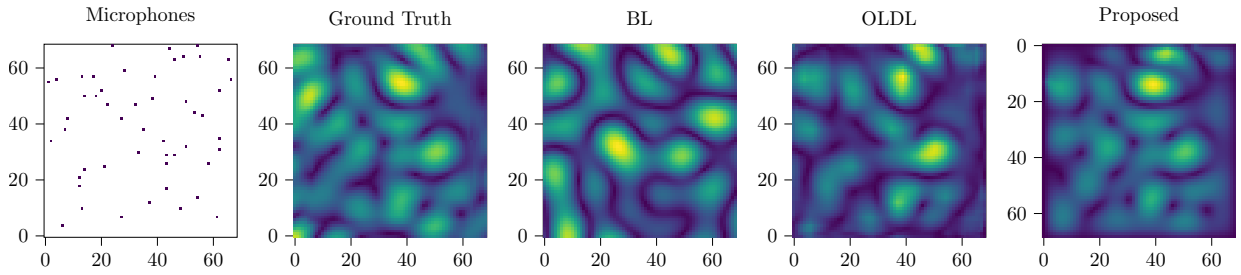


Fig. 3: Magnitude of the sound field at $f = 600\,\text{Hz}$ on the 69×69 point grid, reconstructed using the BL, OLDL and proposed method and compared with the measured ground truth. The reconstruction is performed using $M = 50$ microphones, depicted on the leftmost plot.

OLDL and the proposed method both outperform the BL in the entire frequency range, achieving a comparable performance. Nevertheless, while OLDL is a supervised learning method requiring pre-training on an extensive dataset, the proposed method works in a zero-shot setting, where the only data used for the reconstruction consists of the recording of the 50 measurement microphones in the evaluation room. We remark that, in the proposed method, $\mathbf{D}$ is intialized using the BL dictionary $\mathbf{D}_{\text{bl}}$: this ensures that the improvement of the proposed method over the BL directly results from the physically-informed learning strategy.

We then consider the frequency range $[500, 700]$Hz and analyze the performance when a decreasing number of microphones is available. Fig. 2 shows that, as expected, the performance of all methods degrades as the number of microphones decreases. As in Fig. 1, both OLDL and the proposed method outperform BL for all considered M, with OLDL achieving slightly better scores.

We finally show in Fig. 3 the reconstructed sound field at $f = 600\,\text{Hz}$ using $M = 50$ measurement microphones using the three methods, as compared to the ground truth measured field. This analysis confirms that the proposed method, similarly to OLDL, is able to achieve closer match to the measured field than BL.

The results underline the potential of the proposed dictionary learning approach for sound field reconstruction, and the effectiveness of imposing a physical constraint on the dictionary atoms as an alternative to pre-training the dictionary on an extensive training set. The proposed approach, in fact, guarantees results comparable with the state-of-the-art supervised dictionary learning method [4] without the need for any training data.

V. Conclusion

In this paper, we proposed a physics-informed dictionary learning approach for sound field reconstruction. By introducing a regularization term in the learning objective that encourages the dictionary atoms to satisfy the Helmholtz equation, the proposed approach is able to learn a wideband representation of the sound field in a room in a zero-shot setting, requiring neither any prior knowledge on the acoustic scene nor the need for pre-training the dictionary using extensive training data. Preliminary studies indicate that the proposed approach achieves a reconstruction performance comparable with state-of-the-art supervised dictionary learning methods, without needing any training data and having the advantage of retrieving a physically meaningful dictionary. Future work will focus on comparing the proposed method with other non-dictionary-based techniques and on investigating the introduction of prior knowledge on the analyzed acoustic scene in the algorithm.

Acknowledgment

We would like to express our gratitude to Manuel Hahmann for sharing code and data that helped in the development of this work.

Authorized licensed use limited to: KU Leuven Libraries. Downloaded on April 09,2025 at 12:19:59 UTC from IEEE Xplore. Restrictions apply.

REFERENCES

[1] M. Hahmann, S. A. Verburg Riezu, and E. Fernandez Grande, "Acoustic frequency responses in a conventional classroom," 2021. [Available]: https://findit.dtu.dk/en/catalog/2686334155.

[2] S. Koyama, T. Nishida, K. Kimura, T. Abe, N. Ueno, and J. Brunnström, "Meshrir: A dataset of room impulse responses on meshed grid points for evaluating sound field analysis and synthesis methods," in *Proc. 2021 IEEE Workshop Appl. Signal Process. Audio Acoust. (WASPAA)*, pp. 1–5, 2021.

[3] N. Antonello, E. De Sena, M. Moonen, P. A. Naylor, and T. van Waterschoot, "Room Impulse Response Interpolation Using a Sparse Spatio-Temporal Representation of the Sound Field," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 25, pp. 1929–1941, Oct. 2017.

[4] M. Hahmann, S. A. Verburg, and E. Fernandez-Grande, "Spatial reconstruction of sound fields using local and data-driven functions," *J. Acoust. Soc. Amer.*, vol. 150, pp. 4417–4428, Dec. 2021.

[5] S. Koyama and L. Daudet, "Sparse Representation of a Spatial Sound Field in a Reverberant Environment," *IEEE J. Selected Topics Signal Process.*, vol. 13, pp. 172–184, Mar. 2019.

[6] S. Damiano, F. Borra, A. Bernardini, F. Antonacci, and A. Sarti, "A Compressive Sensing Approach for the Reconstruction of the Soundfield Produced by Directive Sources in Reverberant Rooms," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 32, pp. 2667–2679, 2024.

[7] F. Miotello, L. Comanducci, M. Pezzoli, A. Bernardini, F. Antonacci, and A. Sarti, "Reconstruction of sound field through diffusion models," in *Proc. 2024 IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, pp. 1476–1480, 2024.

[8] M. Pezzoli, D. Perini, A. Bernardini, F. Borra, F. Antonacci, and A. Sarti, "Deep prior approach for room impulse response reconstruction," *Sensors*, vol. 22, no. 7, 2022.

[9] N. Ueno, S. Koyama, and H. Saruwatari, "Kernel ridge regression with constraint of helmholtz equation for sound field interpolation," in *Proc. 16th Int. Workshop Acoust. Signal Enhancement (IWAENC)*, 2018.

[10] D. Sundström, F. Elvander, and A. Jakobsson, "Impulse response interpolation using optimal transport," in *Proc. 57th Asilomar Conf. Signals, Systems Computers*, pp. 1350–1354, 2023.

[11] M. Pezzoli, F. Borra, F. Antonacci, S. Tubaro, and A. Sarti, "A parametric approach to virtual miking for sources of arbitrary directivity," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 28, pp. 2333–2348, 2020.

[12] N. Antonello, E. De Sena, M. Moonen, P. A. Naylor, and T. van Waterschoot, "Joint Source Localization and Dereverberation by Sound Field Interpolation Using Sparse Regularization," in *Proc. 2018 IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, (Calgary, Canada), pp. 6892–6896, Apr. 2018.

[13] T. Abe, S. Koyama, N. Ueno, and H. Saruwatari, "Amplitude matching for multizone sound field control," *IEEE/ACM Trans. Audio, Speech Language Process.*, vol. 31, pp. 656–669, 2022.

[14] M. Vorländer, *Auralization*. Springer, 2020.

[15] 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, Language Process.*, vol. 29, pp. 3052–3063, 2021.

[16] L. McCormack, A. Politis, R. Gonzalez, T. Lokki, and V. Pulkki, "Parametric ambisonic encoding of arbitrary microphone arrays," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 30, pp. 2062–2075, 2022.

[17] 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, 2018.

[18] W. Jin and W. B. Kleijn, "Theory and design of multizone soundfield reproduction using sparse methods," *IEEE/ACM Trans. Audio, Speech Language Process.*, vol. 23, no. 12, pp. 2343–2355, 2015.

[19] F. Borra, I. D. Gebru, and D. Markovic, "Soundfield Reconstruction in Reverberant Environments Using Higher-order Microphones and Impulse Response Measurements," in *Proc. 2019 IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, pp. 281–285, May 2019.

[20] M. Pezzoli, M. Cobos, F. Antonacci, and A. Sarti, "Sparsity-based sound field separation in the spherical harmonics domain," in *Proc. 2022 IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, pp. 1051–1055, 2022.

[21] F. Lluis, P. Martinez-Nuevo, M. Bo Møller, and S. Ewan Shepstone, "Sound field reconstruction in rooms: Inpainting meets super-resolution," *J. Acoust. Society Amer.*, vol. 148, no. 2, pp. 649–659, 2020.

[22] E. Fernandez-Grande, X. Karakonstantis, D. Caviedes-Nozal, and P. Gerstoft, "Generative models for sound field reconstruction," *J. Acoust. Soc. Amer.*, vol. 153, pp. 1179–1190, 02 2023.

[23] I. Tošić and P. Frossard, "Dictionary learning," *IEEE Signal Process. Mag.*, vol. 28, no. 2, pp. 27–38, 2011.

[24] D. Ulyanov, A. Vedaldi, and V. Lempitsky, "Deep image prior," in *Proc. IEEE Conf. Computer Vision Pattern Recognition*, pp. 9446–9454, 2018.

[25] F. Miotello, M. Pezzoli, L. Comanducci, F. Antonacci, and A. Sarti, "Deep prior-based audio inpainting using multi-resolution harmonic convolutional neural networks," *IEEE/ACM Trans. Audio, Speech Language Process.*, 2023.

[26] S. Koyama, J. G. Ribeiro, T. Nakamura, N. Ueno, and M. Pezzoli, "Physics-informed machine learning for sound field estimation: Fundamentals, state of the art, and challenges [special issue on model-based and data-driven audio signal processing]," *IEEE Signal Processing Magazine*, vol. 41, no. 6, pp. 60–71, 2025.

[27] D. Sundström, S. Koyama, and A. Jakobsson, "Sound field estimation using deep kernel learning regularized by the wave equation," in *2024 18th International Workshop on Acoustic Signal Enhancement (IWAENC)*, pp. 319–323, IEEE, 2024.

[28] J. G. Ribeiro, S. Koyama, R. Horiuchi, and H. Saruwatari, "Sound field estimation based on physics-constrained kernel interpolation adapted to environment," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 2024.

[29] F. Miotello, F. Terminiello, M. Pezzoli, A. Bernardini, F. Antonacci, and A. Sarti, "A physics-informed neural network-based approach for the spatial upsampling of spherical microphone arrays," in *2024 18th International Workshop on Acoustic Signal Enhancement (IWAENC)*, pp. 215–219, IEEE, 2024.

[30] M. Olivieri, X. Karakonstantis, M. Pezzoli, F. Antonacci, A. Sarti, and E. Fernandez-Grande, "Physics-informed neural network for volumetric sound field reconstruction of speech signals," *EURASIP Journal on Audio, Speech, and Music Processing*, vol. 2024, no. 1, p. 42, 2024.

[31] E. Morena, R. Malvermi, M. Pezzoli1, F. Antonacci, and A. Sarti, "Reconstruction of directional sources using physics-informed neural networks," in *Proc. INTER-NOISE and NOISE-CON Congr. Conf.*, 2024.

[32] Y. Lu and Y. Wang, "A physics-constrained dictionary learning approach for compression of vibration signals," *Mechanical Syst. and Signal Process.*, vol. 153, p. 107434, 2021.

[33] H. V. Tetali, K. S. Alguri, and J. B. Harley, "Wave Physics Informed Dictionary Learning In One Dimension," in *Proc. 2019 IEEE 29th Int. Workshop Machine Learn. Signal Process. (MLSP)*, (Pittsburgh, PA, USA), pp. 1–6, Oct. 2019.

[34] S. A. Verburg and E. Fernandez-Grande, "Reconstruction of the sound field in a room using compressive sensing," *J. Acoust. Soc. Amer.*, vol. 143, pp. 3770–3779, June 2018.

[35] M. Aharon, M. Elad, and A. Bruckstein, "K-svd: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4311–4322, 2006.

[36] J. Mairal, F. Bach, J. Ponce, and G. Sapiro, "Online learning for matrix factorization and sparse coding," *J. Mach. Learn. Res.*, vol. 11, p. 19–60, mar 2010.

[37] A. M. Bruckstein, D. L. Donoho, and M. Elad, "From sparse solutions of systems of equations to sparse modeling of signals and images," *SIAM Review*, vol. 51, no. 1, pp. 34–81, 2009.

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

[39] S. Diamond and S. Boyd, "CVXPY: A Python-embedded modeling language for convex optimization," *J. Machine Learning Research*, vol. 17, no. 83, pp. 1–5, 2016.

Authorized licensed use limited to: KU Leuven Libraries. Downloaded on April 09,2025 at 12:19:59 UTC from IEEE Xplore. Restrictions apply.