Acoustic field estimation with differentiable physics

Samuel A. Verburg
Technical University of Denmark (DTU)
saveri@dtu.dk

Efren Fernandez-Grande
Technical University of Madrid (UPM)

Peter Gerstoft
Technical University of Denmark (DTU)

Abstract

This work introduces a differentiable physics approach to estimate acoustic fields from a limited number of spatially distributed observations. The initial conditions of the wave equation are approximated with a neural network, and the differential operator is computed with a differentiable numerical solver. We introduce an additional sparsity-promoting constraint to achieve meaningful solutions even under severe undersampling conditions. Numerical experiments demonstrate that the proposed approach can reconstruct sound fields under extreme data scarcity.

1 Introduction

The inverse problem of estimating a sound field over time and space from a limited number of spatially distributed observations is common in acoustic technologies [1, 2, 3]. A main challenge is the large number of observations required, which increases with the domain size and the frequency. Thus, efficient models to estimate sound fields with minimal data are of interest [4, 5, 6, 7, 8].

This study introduces a differentiable physics (DP) [9] approach for sound field estimation, where the initial condition is modeled with a neural network, and a differentiable finite-difference solver is used to solve the wave equation. We demonstrate that even if the network is trained for a given discretization, the sound field can be reconstructed at higher resolutions, as the network can be queried at any point in the domain. Furthermore, we propose a sparsity-promoting constraint to the initial condition. In a series of experiments we show that the proposed DP approach is robust, presents good learning convergence, while achieving small errors. The code is available at `https://github.com/samuel-verburg/differentiable-soundfield-reconstruction.git` and for details see the full paper [10].

2 Sound field estimation using differentiable physics

Let us consider the acoustic pressure field $p(\mathbf{r}, t)$ in the spatio-temporal domain $\Omega \times [0, T]$, where $\Omega \subset \mathbb{R}^2$, $\mathbf{r} \in \Omega$, and $t \in [0, T]$. The pressure field is the solution of the wave equation

$$\nabla^2 p(\mathbf{r}, t) - \frac{1}{c^2} \frac{\partial^2 p(r, t)}{\partial t^2} = 0, \tag{1}$$

with initial conditions, $p(\mathbf{r}, 0) = g(\mathbf{r})$ and $\frac{\partial p}{\partial t}(\mathbf{r}, 0) = 0$. In (1) $c \in \mathbb{R}$ is the medium wave speed, assumed to be a known constant. The domain is considered unbounded, with no reflected waves arriving from outside. To express this, a first-order absorptive boundary condition [11] is considered

$$\nabla p(\mathbf{r}, t) \cdot \mathbf{n} + \frac{1}{c} \frac{\partial p}{\partial t} = 0 \quad \text{at } \mathbf{r} \in \partial\Omega, \tag{2}$$

where $\mathbf{n}$ is the unit vector normal to the boundary $\partial\Omega$.

1st Workshop on Differentiable Systems and Scientific Machine Learning @ EurIPS 2025.

The goal of sound field reconstruction is to estimate the entire pressure field from noisy observations,

$$\hat{p}_{mn} = p(\mathbf{r}_m, t_n) + e_{mn} \quad \text{for } m = 0, \dots, M_{\text{ob}} - 1 \text{ and } n = 0, \dots, N-1, \tag{3}$$

where $\mathbf{r}_0, \dots, \mathbf{r}_{M_{\text{ob}}-1}$ are the sensor locations, $t_0, \dots, t_{N-1}$ are the time samples, and e_{mn} is additive noise. Measurements are typically performed using microphone arrays or distributed sensors. Therefore, the pressure is finely sampled over time, but only a few positions are sampled over space.

In the proposed DP approach a neural network $g_{\text{dp}}(\mathbf{r}; \boldsymbol{\theta})$ models the unknown initial pressure. The physical constraints are imposed by applying a numerical PDE solver to the network output. Training the DP neural network amounts to solving the optimization problem

$$\min_{\boldsymbol{\theta}} \left\{ \lambda_{\text{data}} \mathcal{L}_{\text{data}}(\mathbf{p}^0(\boldsymbol{\theta}), \mathbf{p}^1 \dots, \mathbf{p}^{N-1}) + \lambda_{\text{sp}} \mathcal{L}_{\text{sp}}(\mathbf{p}^0(\boldsymbol{\theta})) \right\}, \tag{4}$$

where $\mathbf{p}^1, \dots, \mathbf{p}^{N-1}$ is the numerical solution of the PDE computed in a M_{grid}-dimensional discretization grid, and $\mathbf{p}^0(\boldsymbol{\theta})$ is obtained by sampling the neural network $g_{\text{dp}}(\mathbf{r}; \boldsymbol{\theta})$ at the grid positions. Therefore, $\mathbf{p}^n$ is a vector in $\mathbb{R}^{M_{\text{grid}}}$ instead of being a continuous function. For simplicity, it is assumed that for each observation position, $\mathbf{r}_0, \dots, \mathbf{r}_{M_{\text{ob}}-1}$, there is a point in the discretization grid. A data fitting function, $\mathcal{L}_{\text{data}}$, that operates on the discrete pressure can be expressed as

$$\mathcal{L}_{\text{data}}\left(\mathbf{p}^0(\boldsymbol{\theta}), \dots, \mathbf{p}^{N-1}\right) = \frac{1}{M_{\text{ob}} N} \sum_{n=0}^{N-1} \left\| \mathbf{M}\mathbf{p}^n - \hat{\mathbf{p}}^n \right\|^2, \tag{5}$$

where $\mathbf{M}$ is a $M_{\text{ob}} \times M_{\text{grid}}$ binary matrix that extracts the pressure values at the observation positions, and $\hat{\mathbf{p}}^n \in \mathbb{R}^{M_{\text{ob}}}$ denotes the observations in Eq. (3) arranged as a vector.

The finite difference method [12] solves the PDE numerically. The solution is obtained by applying the explicit time integration scheme

$$\mathbf{p}^1 = \mathbf{p}^0 + 0.5\mathbf{L}\mathbf{p}^0, \text{ and } \mathbf{p}^{n+1} = 2\mathbf{p}^n - \mathbf{p}^{n-1} + \mathbf{L}\mathbf{p}^n \text{ for } n = 1, \dots, N-1, \tag{6}$$

where $\mathbf{L} = (c\Delta t/\Delta r)^2 \mathbf{L}_\Delta$ and $\mathbf{L}_\Delta \in \mathbb{R}^{M_{\text{grid}} \times M_{\text{grid}}}$ is the central difference approximation of the Laplace operator $\nabla^2[\cdot]$. The scalars Δt and Δr are the sampling period and grid spacing, respectively. The optimization objective in (4) excludes derivative terms, as both initial conditions $p(\mathbf{r}, 0) = g(\mathbf{r})$ and $\frac{\partial p}{\partial t}(\mathbf{r}, 0) = 0$ are implicit in the first step of the solver, Eq. (6). To handle the unbounded domain, the boundary condition of (2) is incorporated into the numerical solver. The finite difference approximation of the absorptive boundary computes the values of $\mathbf{p}^{n+1}$ at the boundary based on $\mathbf{p}^n$ at the boundary and adjacent points, as well as $\mathbf{p}^{n+1}$ at the adjacent points.

A sparsity-promoting constraint [13] for the initial condition is considered,

$$\mathcal{L}_{\text{sp}}(\mathbf{p}^0) = \frac{1}{M_{\text{grid}}} \|\mathbf{p}^0\|_1, \tag{7}$$

which expresses the assumption that the sound field is generated by a few sound sources, and thus $g(\mathbf{r})$ is zero in most of the domain . To balance the loss terms $\mathcal{L}_{\text{data}}$ and $\mathcal{L}_{\text{sp}}$ an annealing algorithm [14] is used to select the weights λ_{data} and λ_{sp}.

Central to the proposed DP approach is the neural network, $g_{\text{dp}}(\mathbf{r}; \boldsymbol{\theta})$, to model the initial condition as a continuous, smooth function that maps any input coordinate within the domain to a corresponding output value. Therefore, even if the PDE is solved on a fixed discrete grid during training, the resolution can be increased by sampling the neural network on a finer grid, and then solve the PDE with a higher resolution numerical solver.

3 Numerical experiments

A 2+1D domain is defined, where the spatial domain is a square of side length $L = 1$, the temporal domain has a duration of $T = 0.343$, and the speed of sound is $c = 1$. The temporal domain is divided into $n = 50$ samples, giving a sampling period $\Delta t = 7.0 \times 10^{-3}$. For the finite difference solver a regular discretization grid of $M_{\text{grid}} = 100^2$ is defined.

Single pulse: Synthetic data for multiple sound fields is generated. For the first one, the initial condition is a single Gaussian pulse of unit amplitude and scale $\sigma = 0.02$ placed at the center of the domain,

$$g(\mathbf{r}) = \exp\left(-0.5\|\mathbf{r} - \mathbf{r}_0\|^2/\sigma^2\right), \tag{8}$$

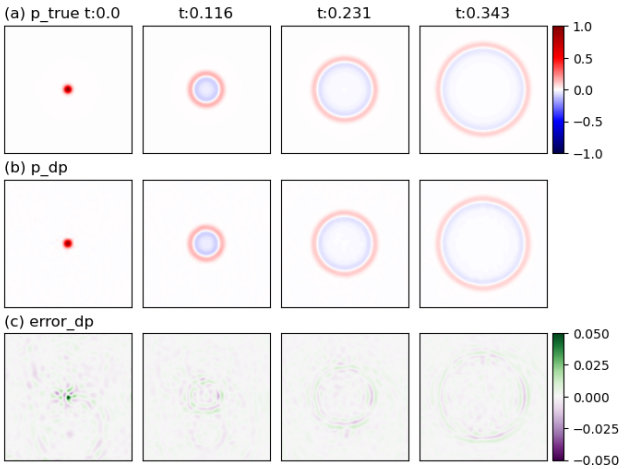

Figure 1: Sound field consisting on a single pulse at the domain center. Each column corresponds to a time frame. Row (a): reference solution. Row (b): DP model estimation. Row (c): difference between reference and DP.

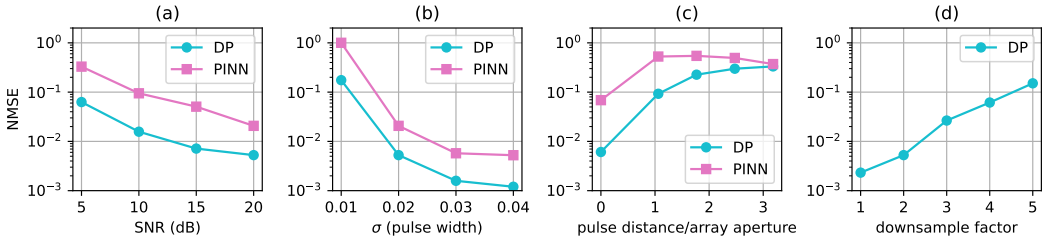

Figure 2: Normalized Mean squared error vs (a) the SNR, (b) the source width, (c) the pulse distance to the array center, and (d) the downsample factor between the evaluation and training grids.

where $\mathbf{r}_0 = (L/2, L/2)$. The observations used for the reconstruction conform a pseudo-random array where the sensor locations are sampled from the discretization grid within $[0.1L, 0.9L] \times [0.1L, 0.9L]$ and with a minimum distance of 0.05 between sensors. The number of time samples is $N = 50$ and the number of sensors is $M_{\text{obs}} = 20$. Additive Gaussian noise is added such that the SNR is 20 dB. The observed data and reference values of the sound field are obtained from the analytical solution of the acoustic wave equation in free field with a Gaussian pulse as initial condition [15]. The reference sound field is computed on a grid of twice the spatial and temporal resolutions of the DP finite difference grid.

Figure 1 shows the reference sound field and estimation for the single Gaussian pulse. The DP results show an accurate reconstruction throughout the spatio-temporal domain, with only noticeable differences at $t = 0$. The normalized mean square error (NMSE) computed over all the spatio-temporal points on the evaluation grid is 5.3×10^{-3}.

The reconstruction performance is analyzed by training the model in different scenarios. As benchmark, a conventional phyics-informed neural network (PINN) [16] is trained to solve the same estimation problem. Figure 2(a) shows the NMSE for various noise levels. The proposed DP model largely outperforms the PINN for all tested SNRs, presenting errors almost one order of magnitude smaller. Figure 2(b) shows the NMSE as a function of the pulse scale σ, which is directly related to the frequency content of the acoustic field. The PINN fails to reconstruct the sound field of highest frequency (smallest σ), presenting a NMSE close to 1, while the DP model consistently achieves lower errors. It is worth noting that the DP network has one layer less and half the number of units per layer that the PINN. Figure 2(c) shows the NMSE as a function of the distance between the pulse and the array center normalized by the array aperture. The experiment serves to evaluate the extrapolation capabilities of the models to areas where there is no observed data and the estimation

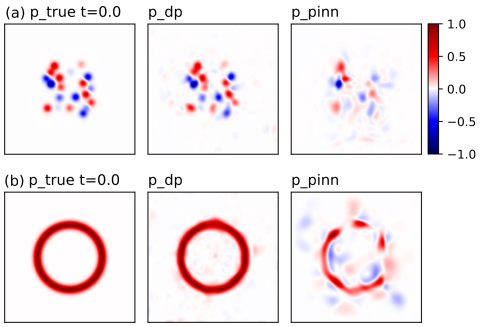

Figure 3: Initial conditions for different sound fields. The columns correspond to the reference, DP estimation, and PINN estimation, respectively. Row (a) twenty pulses. Row (b) continuous ring-like source.

relies only on the physics of wave propagation. The PINN model presents a large error as soon as the source is outside the array aperture as the physics are only included as a weak constraint. Conversely, the DP model output satisfies the underlying physics by design. Figure 2(d) shows the NMSE for different discretization grids used during training the DP model. The x-axis corresponds to the downsample factor of the training grid with respect to the reference grid. Since the initial conditions are approximated with a continuous function we can upscale the estimation to any desired resolution. As expected, the NMSE increases for lower training resolutions, which is caused by the accumulation of numerical errors and the fact that coarser grids are not able to represent high spatial frequencies present in the initial condition. Therefore, care must be taken to select an appropriate grid during training. Nonetheless it is possible to obtain estimations with errors below 10^{-1} even when the network is trained with coarse grids.

Complex source distributions Additional sound fields are synthesized to test the methods in other challenging sound field reconstruction problems, in particular when sparsity cannot be assumed. Figure 3(a) corresponds to the combination of twenty Gaussian pulses randomly distributed in $[0.3L, 0.7L]$ with random amplitudes in $[-1, 1]$ and $\sigma = 0.02$. The reference is shown in the first column of Figure 3, while the DP and PINN estimations are shown in the second and third columns, respectively. The PINN clearly fails to learn a meaningful solution, while the DP is able to recover the initial pressure in both cases. The DP model's ability to recover such a complex sound field with as few as 20 sensors is remarkable. Assuming that a characteristic wavelength of the sound field is 2σ and considering that $L/\sigma = 50$, a grid of 50^2 sensors would be required to recover the sound field according to classical sampling theory (without sparsity constraint). Figure 3(b) shows a ring-like initial condition. In this case, the observed and reference data are obtained from a high resolution finite difference simulation. The DP model can reconstruct the sound field even when the source is a continuous line, and the initial pressure is not as sparse.

4 Conclusion

We propose a differentiable physics approach for sound field reconstruction. Integration of a numerical solver in the training of a neural network enables the incorporation of hard physical constraints robustly. While the cost for training iteration of the DP approach is higher, the optimization is more stable than in conventional PINNs, and convergence of the learning process is achieved in a fraction of the optimizer steps (5×10^4 for DP and 5×10^5 for PINN). For the tested examples, the inference time of the trained DP and PINN approaches is similar. Formulating the solver in a differentiable way using AD makes the training process very simple since only the forward solver is required. The DP approach is generalizable beyond the training discretization, and the solutions obtained can be scaled to higher resolutions. Additionally, incorporating a sparsity-promoting constraint enables the reconstruction of sound fields with very little data. The experiments show that the DP model achieves accurate reconstructions and low errors even in challenging, highly undersampled problems.

References

[1] 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 Trans. Audio Speech. Lang. Process.*, vol. 25, no. 10, pp. 1929–1941, 2017.

[2] S. A. Verburg and E. Fernandez-Grande, "Reconstruction of the sound field in a room using compressive sensing," *J. Acoust. Soc. Am.*, vol. 143, no. 6, pp. 3770–3779, 2018.

[3] S. Koyama and L. Daudet, "Sparse representation of a spatial sound field in a reverberant environment," *IEEE J. Sel. Top. Signal Process.*, vol. 13, no. 1, pp. 172–184, 2019.

[4] D. Caviedes-Nozal, N. A. 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, 2021.

[5] E. Fernandez-Grande, X. Karakonstantis, D. Caviedes-Nozal, and P. Gerstoft, "Generative models for sound field reconstruction," *J. Acoust. Soc. Am.*, vol. 153, no. 2, pp. 1179–1190, 2023.

[6] J. G. Ribeiro, S. Koyama, R. Horiuchi, and H. Saruwatari, "Sound field estimation based on physics-constrained kernel interpolation adapted to environment," *IEEE Trans. Audio Speech Lang. Process.*, 2024.

[7] H. Bi and T. D. Abhayapala, "Point neuron learning: A new physics-informed neural network architecture," *EURASIP J. Audio Speech Music Process.*, vol. 2024, no. 1, p. 56, 2024.

[8] 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," *IEEE Signal Process. Mag.*, vol. 41, no. 6, pp. 60–71, 2025.

[9] N. Thuerey, B. Holzschuh, P. Holl, G. Kohl, M. Lino, Q. Liu, P. Schnell, and F. Trost, *Physics-based Deep Learning*. WWW, 2021. [Online]. Available: https://physicsbaseddeeplearning.org

[10] S. A. Verburg, E. Fernandez-Grande, and P. Gerstoft, "Differentiable physics for sound field reconstruction," *J. Acoust. Soc. Am. DOI 10.1121/10.0039862*, 2025.

[11] R. Clayton and B. Engquist, "Absorbing boundary conditions for acoustic and elastic wave equations," *B. seismological society of America*, vol. 67, no. 6, pp. 1529–1540, 1977.

[12] H. P. Langtangen and S. Linge, *Finite difference computing with PDEs: a modern software approach*. Springer Nature, 2017.

[13] E. J. Candès and M. B. Wakin, "An introduction to compressive sampling," *IEEE signal processing magazine*, vol. 25, no. 2, pp. 21–30, 2008.

[14] S. Wang, Y. Teng, and P. Perdikaris, "Understanding and mitigating gradient flow pathologies in physics-informed neural networks," *SIAM J. Sci. Comput.*, vol. 43, no. 5, pp. A3055–A3081, 2021.

[15] C. K. Tam and J. C. Webb, "Dispersion-relation-preserving finite difference schemes for computational acoustics," *J. Comput. Phys.*, vol. 107, no. 2, pp. 262–281, 1993.

[16] B. Moseley, A. Markham, and T. Nissen-Meyer, "Solving the wave equation with physics-informed deep learning," *arXiv preprint arXiv:2006.11894*, 2020.