forum acusticum 2026
8-12 September · Graz, Austria

DATA-DRIVEN EQUATION DISCOVERY IN WEAKLY NONLINEAR SOUND BEAMS

Tereza Filipská[1], Viktor Hruška[1], Aneta Furmanová[1]

[1] *Czech Technical University in Prague, Faculty of Electrical Engineering, Technická 2, 166 27 Prague, Czech Republic*

This paper is a revised version of the authors' submitted manuscript that was peer-reviewed by at least two anonymous reviewers.

Abstract

In this study, we build on our previous article, in which we presented the possibilities of data-driven equation discovery in the field of nonlinear acoustics. Now we focus on another benchmarking task: finding equations for weakly nonlinear sound beams. To this end, we rely on simulations of the compressible Navier-Stokes equations. In this case, we are dealing with wave propagation from a piston-oscillating boundary condition. After the appropriate transformation to retarded time and making use of the weak formulation, we again apply our algorithm and are able to re-create the textbook case to a high degree of accuracy: the well-known Khokhlov-Zabolotskaya-Kuznetsov (KZK) equation is discovered from data.

Keywords: *Equation discovery, KZK equation, nonlinear acoustics, paraxial approximation, sound beams*

1. Introduction

Acoustic waves of finite amplitude naturally occur in many forms: as intense traveling waves with characteristic steepening, or as standing waves in resonators [1, 2]. In the following paragraphs, we focus on one specific subset of scenarios: weakly nonlinear sound beams.

Their scope of application is vast, both in theory and in practice. From a theoretical standpoint there is the subtle interplay between nonlinearity and diffraction, leading to the well-known concepts of parametric arrays, self-demodulation and strong wave focusing. The applications of these results range from underwater acoustics [3] and medical ultrasonography [4] to some unusual experiments and devices from the audio engineering [5].

With the exception of highly simplified cases, such problems are difficult to solve analytically – especially when it comes to investigating the sound field in its spatial entirety. At the same time, this is a relatively well-explored area of research. Such situation is suitable for application of physics-informed machine learning. These methods are particularly promising because they may eventually enable the discovery of simplified models or solutions to partially specified problems that are not sufficiently well-defined for conventional numerical methods. A crucial first step, however, is benchmarking. That is, establishing that equation discovery methods can reliably recover solutions already known from the literature.

The field of sparse identification of nonlinear dynamics (SINDy) has rapidly developed in recent years [6–9]. While standard frameworks typically rely on point-wise numerical differentiation, the latest methods employ weak or integral formulations to handle noisy data and sharp gradients more robustly.

Our approach here builds directly on our previous article [10], in which we demonstrated the possibilities of data-driven discovery of wave equations for the weakly nonlinear limit. The result was a workflow that uncovered the Westervelt and Kuznetsov equations from data simulated using compressible Navier-Stokes equations, even in more than one spatial dimension. Our primary objective is to adapt the methodological workflow so that it is specifically tailored to sound beams. This shift requires several modifications: the use of specialized boundary conditions (with a portion of the boundary oscillating in a piston-like manner), the adoption of a paraxial approximation in place of the full d'Alembert operator, and a transformation to retarded time [3]. This ultimately leads us to a data-driven discovery of the well-known KZK equation (Khokhlov–Zabolotskaya–Kuznetsov).

2. Theory

In line with our previous work as well as generally established approaches in the field of finite-amplitude acoustics, we first need to understand the appropriate linear basis and its weakly nonlinear extension.

Corresponding author: *filipte3@fel.cvut.cz.*

For sound propagation in two dimensions, the wave equation reads:

$$\frac{\partial^2 p}{\partial x^2} + \frac{\partial^2 p}{\partial y^2} - \frac{1}{c_0^2}\frac{\partial^2 p}{\partial t^2} = 0 \tag{1}$$

where p, c_0 denote the acoustic pressure and the adiabatic speed of sound, respectively. While the common derivation of the equation for beams (the wave equation in paraxial approximation) usually involves the Fourier transform and work with a specific limit of the Helmholtz equation, that method is fundamentally constrained to monofrequency (monochromatic) waves. For acoustic beams characterized by a variable frequency content, a more advantageous alternative is to employ the retarded time formulation [1, 3]:

$$p(x, y, t) \quad \rightarrow \quad q(x, y, \tau), \quad \tau = t - \frac{y}{c_0} \tag{2}$$

where the definition of the retarded time τ considers propagation of a sound beam along y axis. After substitution and some manipulations, it is possible to arrive at the following:

$$\frac{\partial^2 q}{\partial x^2} + \frac{\partial^2 q}{\partial y^2} - \frac{2}{c_0}\frac{\partial^2 q}{\partial y \partial \tau} = 0\,. \tag{3}$$

The final form of the linear wave equation for beams is obtained by realizing that for waves along y with slowly spatially-varying amplitude it holds that $|\frac{\partial^2 q}{\partial y^2}| \ll |\frac{2}{c_0}\frac{\partial^2 q}{\partial y \partial \tau}|$ and hence:

$$\frac{2}{c_0}\frac{\partial^2 q}{\partial y \partial \tau} - \frac{\partial^2 q}{\partial x^2} = 0\,. \tag{4}$$

Following the methodology established in our previous work, we will use the linear form of the equation as a starting point for equation discovery – only this time for the case of beams.

To generalize the discovery process, we introduce a set of dimensionless variables:

$$\begin{aligned} X &= \frac{x}{l} \qquad & Y &= \frac{y}{l} \\ P &= \frac{q}{\rho_0 c_0^2} \qquad & \Theta &= \frac{c_0}{l}\tau \end{aligned} \tag{5}$$

where l is the characteristic spatial dimension and ρ_0 is the ambient density. By substituting these relations into the previous equation 4, we get the dimensionless form of our target:

$$2\frac{\partial^2 P}{\partial Y \partial \Theta} - \frac{\partial^2 P}{\partial X^2} = f(P) \tag{6}$$

where $f(P)$ represents an unknown target function consisting of P and its partial derivatives, which is to be identified from a library of candidate terms during the equation discovery process.

Since this benchmark study is intended to verify the concept, we know the expected theoretical result corresponding to the two-dimensional lossless KZK equation, namely $f(P) = \beta\frac{\partial^2 P^2}{\partial \Theta^2}$:

$$2\frac{\partial^2 P}{\partial Y \partial \Theta} - \frac{\partial^2 P}{\partial X^2} = \beta\frac{\partial^2 P^2}{\partial \Theta^2} \tag{7}$$

where β denotes the nonlinearity coefficient. For air modeled as an ideal gas ($\gamma = 1.4$), the theoretical benchmark value is $\beta = (\gamma + 1)/2 = 1.200$.

3. Data generation

The dataset for the equation-discovery algorithm was generated by numerically solving the full compressible Navier-Stokes equations using a custom in-house solver [10, 11]. For the solution, we employed the Kurganov-Tadmor central scheme, which is specifically designed for the stable capturing of discontinuities and steep gradients in the pressure profile [12]. The simulation models a 2D nonlinear acoustic beam in the air, driven by a piston source with a Gaussian aperture profile, creating a field in which the beam propagates primarily in one direction (along the y-axis).

Since this is a proof-of-concept study formulated in dimensionless quantities, the dataset includes ten different scenarios with the non-dimensional excitation frequency Ω varying by $\pm 20\%$ (from 60 to 110, while all other parameters remained constant). This range was determined based on practical numerical constraints. It needed to be sufficiently diverse, while at the same time ensuring that even higher harmonic frequencies remained well resolved on the computational grid.

The spatial evolution of the acoustic pressure field P, as shown in Fig. 1, illustrates the transition from a linear source to a nonlinearly distorted beam. For the discovery algorithm, the raw data $P(X, Y, T)$ were transformed into retarded time $\Theta = T - Y$.

This transformation "freezes" the wave profile in the moving frame which is a crucial step for the algorithm to identify the governing operators. As illustrated in Fig. 2, the wavefronts that appear tilted in the physical domain (Y, T) due to the propagation velocity, become vertically aligned in the domain (Y, Θ). This alignment is crucial for the algorithm to smooth the data and effectively identify the paraxial operator.

4. Overview of the algorithm

The approach employed belongs to a class of supervised algorithms that utilize a library of candidate terms, such as the well-known SINDy [6–8]. A key feature of our approach is the use of a weak formulation of the equations, which, through integration by parts, allows us to transfer derivatives from the data to sufficiently smooth test functions [13].

A more comprehensive description and formulation of the discovery algorithm is provided in our previous work [10]. In that study, the weak formulation was advantageous mainly due to the occurrence of shock waves. In the present study, we also take advantage

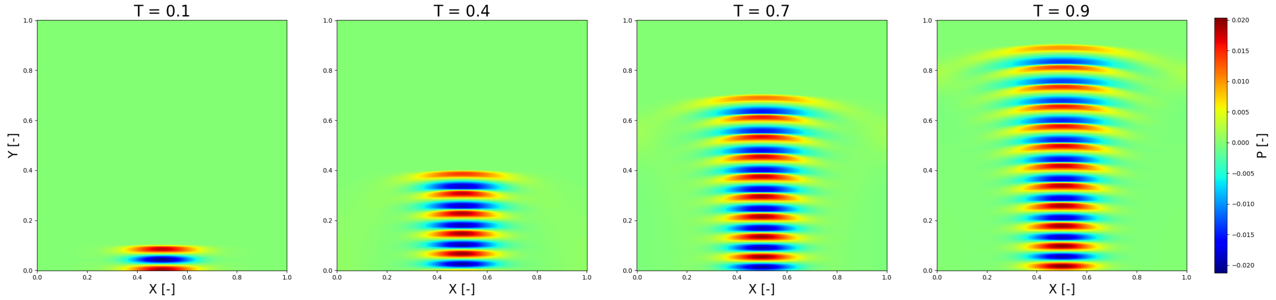

Figure 1: Evolution of the 2D acoustic pressure field P in dimensionless X, Y coordinates at selected time instants T. Particularly on the more developed beam ($T = 0.7$ or 0.9), the combination of diffraction and wave steepening is evident.

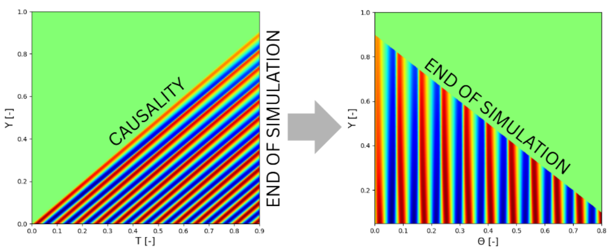

Figure 2: Space-time plot (Y, T) on the beam axis and its transformation to retarded time (Y, Θ). Notice how, on the one hand, causality comes into play (which is guaranteed in retarded time) and, on the other hand, the simple technical fact of the finiteness of numerical simulation.

of the effective smoothing of data transformed into retarded time.

The core of the algorithm has not changed significantly from the previously mentioned article: data from individual cases are fed into the model separately, and features from the candidate library are tested on them using L1-regularized least squares (LASSO) with an appropriate shrinkage parameter selected using the Bayesian information criterion. However, this is merely a step used to identify a relevant subset of candidates. The rule is that only those candidate terms with a 65% occurrence within the studied cases are considered suitable. Subsequently, the final numerical coefficients of the discovered partial differential equation (PDE) model are obtained by means of ordinary least squares and a greater number of cases allows us to determine basic statistics (mean and standard deviation of the coefficients).

For this proof of concept, the candidate terms included derivatives of pressure and its square with respect to the retarded time (from first to third order in all combinations).

The physical limits of the simulation constrain the possibilities of data sampling during the equation discovery. To ensure the integrity of the weak formulation, each test function is centered such that its entire support remains within the computational domain. In layman's terms, it wouldn't be difficult to reach somewhere in the transformed data where the values haven't been calculated (note the causality and the end of simulation lines in Fig. 2). Specifically, we enforce the condition $\Theta + Y \leq T_{max}$ to prevent the test functions from extending beyond the temporal limits of the available simulation data. Adhering to these constraints ensures that the integration by parts used in the formulation remains mathematically valid.

5. Results

For ten different sound beams with various frequencies, the algorithm led to the following equation:

$$2\frac{\partial^2 P}{\partial Y \partial \Theta} - \frac{\partial^2 P}{\partial X^2} = (1.218 \pm 0.079)\frac{\partial^2 P^2}{\partial \Theta^2} \qquad (8)$$

where the interval in parentheses represents the 95% confidence interval.

The derived expression formally corresponds to the KZK equation without the thermoviscous losses (Eq. (7), sometimes referred to simply as the KZ equation). From the relations for dimensionless quantities (Eq. (5)) and the nonlinearity coefficient β for air as an ideal gas ($\beta = 1.200$), we can conclude that the result is a very good fit of the KZK equation seen from the perspective of sparse supervised regression. A relative error of the nonlinear term is ca. 1% compared to the theoretical value.

In the air, absorption is much weaker than diffraction and nonlinear effects. These effects are so low that their values are numerically undetectable by the discovery algorithm. However, the high accuracy of the determined nonlinearity coefficient suggests that this model is well-suited even for media with higher thermoviscous losses, such as biological tissues.

6. Conclusion

In this article, we have shown that the equation discovery algorithm developed for weakly nonlinear acoustics problems and tested on cases indicated for the Westervelt and Kuznetsov equation is capable, after certain modifications, of identifying the KZK equation as well. This demonstrates that working with nonlinear acoustic beams using these techniques is entirely feasible.

Since this article serves as a proof of concept, broader issues (such as performance evaluation on completely unknown data configurations) are the subject of future work.
Knowing that we have met this benchmark, we can apply a similar approach in the future to media that do not have as straightforward material expressions and representation of conservation laws as the compressible Navier-Stokes equations for an ideal gas. Further possibilities also open up in data-driven linearization techniques (as an analogy to the Cole-Hopf transform for Burgers equation), which are otherwise not known for the KZK equation.

7. Acknowledgments

This work was supported by the Grant Agency of the Czech Technical University in Prague, grant No. SGS25/136/OHK3/3T/13.

References

[1] M. F. Hamilton and D. T. Blackstock, *Nonlinear Acoustics*. American Institute of Physics, 1998, ISBN: 978-3-031-58962-1.

[2] K. Naugolnykh and L. Ostrovsky, *Nonlinear wave processes* (Cambridge texts in applied mathematics), en. Cambridge, England: Cambridge University Press, 1998.

[3] M. F. Hamilton and C. L. Morfey, "Model equations," in *Nonlinear Acoustics*, M. F. Hamilton and D. T. Blackstock, Eds. Cham: Springer Nature Switzerland, 2024, pp. 39–61, ISBN: 978-3-031-58963-8. DOI: `10.1007/978-3-031-58963-8_3`.

[4] M. S. Canney, M. R. Bailey, L. A. Crum, V. A. Khokhlova, and O. A. Sapozhnikov, "Acoustic characterization of high intensity focused ultrasound fields: A combined measurement and modeling approach," *The Journal of the Acoustical Society of America*, vol. 124, no. 4, pp. 2406–2420, Oct. 2008, ISSN: 0001-4966. DOI: `10.1121/1.2967836`.

[5] W.-S. Gan, J. Yang, and T. Kamakura, "A review of parametric acoustic array in air," *Applied Acoustics*, vol. 73, no. 12, pp. 1211–1219, Dec. 2012, ISSN: 0003-682X. DOI: `10.1016/j.apacoust.2012.04.001`.

[6] S. L. Brunton, J. L. Proctor, and J. N. Kutz, "Discovering governing equations from data by sparse identification of nonlinear dynamical systems," *Proceedings of the National Academy of Sciences*, vol. 113, no. 15, pp. 3932–3937, 2016.

[7] S. H. Rudy, S. L. Brunton, J. L. Proctor, and J. N. Kutz, "Data-driven discovery of partial differential equations," *Science Advances*, vol. 3, no. 4, e1602614, 2017.

[8] S. L. Brunton and J. N. Kutz, *Data-Driven Science and Engineering: Machine Learning, Dynamical Systems, and Control*, 2nd. Cambridge University Press, 2022.

[9] e. a. Xu, "Data-driven discovery of partial differential equations (title according to xu et al. 2025)," *Nature Communications*, vol. 16, p. 10 255, 2025.

[10] V. Hruška, A. Furmanová, T. Filipská, and J. Valášek, "Data-driven discovery of weakly nonlinear acoustic wave equations," *Journal of Sound and Vibration*, vol. 626, p. 119 625, 2026, ISSN: 0022-460X. DOI: `10.1016/j.jsv.2025.119625`.

[11] T. Filipská, *Numerical solutions of nonlinear acoustic wave propagation under parameters appropriate to medical ultrasound applications*, Bachelor's thesis, 2025. [Online]. Available: `http://hdl.handle.net/10467/122397`.

[12] A. Kurganov and E. Tadmor, "New high-resolution central schemes for nonlinear conservation laws and convection–diffusion equations," *Journal of Computational Physics*, vol. 160, no. 1, pp. 241–282, May 2000, ISSN: 0021-9991. DOI: `10.1006/jcph.2000.6459`.

[13] R. Stephany and C. Earls, "Weak-PDE-LEARN: A weak form based approach to discovering PDEs from noisy, limited data," *Journal of Computational Physics*, vol. 506, p. 112 950, Jun. 2024, ISSN: 0021-9991. DOI: `10.1016/j.jcp.2024.112950`.