



## **A Feedforward Neural Network for Modeling of Average Pressure Frequency Response**

Downloaded from: <https://research.chalmers.se>, 2024-04-25 15:52 UTC

Citation for the original published paper (version of record):

Pettersson, K., Karzhou, A., Pettersson, I. (2022). A Feedforward Neural Network for Modeling of Average Pressure Frequency Response. *Acoustics Australia*, 50(2): 185-201.

<http://dx.doi.org/10.1007/s40857-021-00259-w>

N.B. When citing this work, cite the original published paper.



# A Feedforward Neural Network for Modeling of Average Pressure Frequency Response

Klas Pettersson<sup>1</sup> · Andrei Karzhou<sup>2</sup> · Irina Pettersson<sup>1,3</sup>

Received: 8 August 2021 / Accepted: 10 November 2021  
© The Author(s) 2022

## Abstract

The Helmholtz equation has been used for modeling the sound pressure field under a harmonic load. Computing harmonic sound pressure fields by means of solving Helmholtz equation can quickly become unfeasible if one wants to study many different geometries for ranges of frequencies. We propose a machine learning approach, namely a feedforward dense neural network, for computing the average sound pressure over a frequency range. The data are generated with finite elements, by numerically computing the response of the average sound pressure, by an eigenmode decomposition of the pressure. We analyze the accuracy of the approximation and determine how much training data is needed in order to reach a certain accuracy in the predictions of the average pressure response.

**Keywords** Frequency response · Sound pressure · Helmholtz equation · Machine learning · Feedforward dense neural network

## 1 Introduction

Modeling such acoustics problems as building acoustics, vehicle interior noise problems, noise reduction, insertion and transmission loss, often requires computing average sound pressure, which in its turn is based on the computation of natural frequencies and the response to a dynamic excitation. There are two main approaches to modeling of acoustic systems under small frequency excitation: To model acoustics in the time domain or in the frequency domain. Sound waves, as vibrations, are described by a time-dependent wave equation, which can be reduced to a time-independent Helmholtz equation by assuming harmonic dependence on time [30]. If the geometry of the domain is complex, it is decomposed into subdomains, in each subdomain, the analysis is performed. Given an acoustic system, one needs to solve the Helmholtz equation repeatedly for given frequency, which might be very costly. In addition, the most important and costly part in a FEM-analysis is the computation of eigenfrequencies and eigenfunctions in each subdomain.

We propose a feedforward dense neural network (multi-layer perceptron) for computing the average sound pressure in cylindrical cavities with polygonal boundary.

Some motivation for studying the average pressure response over a range of frequencies can be found in engineering applications. For example, standardized frequency ranges can be found in the ISO standard [9].

For an overview of deep learning in neural networks, we refer to [28] and for overview of basic mathematical principles to [11,31] and literature therein. Application of machine learning methods in acoustics has made significant progress in recent years. A comprehensive overview of the recent advances is given in [8]. The frequency response problem, being the basis in modeling of acoustic problems, is not specifically addressed in [8], as any other combination of machine learning techniques and modeling with partial differential equations (PDEs).

There are many work devoted to solving PDEs, forward as well as inverse problems, by means of machine learning techniques. We mention just some of them. The work [18] propose an algorithm to solve initial and boundary value problems using artificial neural networks. The gradient descent is used for optimization. In [23] the authors propose an algorithm to solve an inverse problem associated with the calculation of the Dirichlet eigenvalues of the anisotropic Laplace operator. The finite elements are used to generate

✉ Klas Pettersson  
klas.pettersson@gmail.com

<sup>1</sup> Chalmers University of Technology, Gothenburg, Sweden

<sup>2</sup> University of Tromsø, Tromsø, Norway

<sup>3</sup> Gothenburg University, Gothenburg, Sweden

the training data. The main goal is to characterize the material properties (coefficient matrix) through the eigenvalues.

In [29] the authors approximate solutions to high-dimensional PDEs with a deep neural network which is trained to satisfy the differential operator, initial condition, and boundary conditions. The convergence of the neural network to the solution of a PDE is proved. In contrast to [18], the algorithm in [29] is mesh-free.

In [1] a partially learned approach is employed for the solution of ill-posed inverse problems. The paper contains also a good literature overview for inverse problems. In [7] the authors propose deep feedforward artificial neural networks (mesh-free) to approximate solutions to partial differential equations in complex geometries. The paper [25] focuses on the nonlinear partial differential equations. Some other network inspired approaches in the study of PDEs are [5,6,21,22].

In [27], the authors propose an iterative solver for the Helmholtz equation which combines traditional Krylov-based solvers with machine learning. The result is a reduced computational complexity.

In the present paper, we use feedforward fully connected neural networks with the ReLU activation function in hidden layers in order to approximate the average pressure function originated in frequency response problems. We choose to use three hidden layers, 128 nodes in each, and ADAM optimizer. The step size in the gradient descent is scheduled to have polynomial decay. A more detail description of the neural network is provided in Sect. 4.

The feedforward neural network is designed to directly learn the average pressure, in contrast to the works cited above, where the neural networks are tailored to predict the coefficients of the inverse problem or solutions to PDEs.

It is known that a neural network can approximate any continuous function to an arbitrary accuracy [15]. We focus on the frequency response problem (low frequencies) in two-dimensional polygonal cylinders. Assuming harmonic load on a part on the boundary, we arrive at a time-independent Helmholtz equation for the sound pressure. The mean value of the average pressure over a given frequency range is an important quantity for characterizing the sound attenuation, insertion and transmission losses. The numerical solution of this problems implies solving the Helmholtz equation for many different values of the spectral parameter, which is a costly problem. Besides, the pressure function is singular near the eigenvalues of the Laplace equation, and the standard quadrature schemes cannot be applied in order to compute the average  $\Psi$  over a frequency range (see Sect. 2.3 and Fig. 3b for the explanation). Instead, we represent the average pressure  $\Psi$  (objective function) in terms of a Hilbert basis, the eigenmodes of the Laplace operator. We generate data sets containing around 700,000 randomly generated points which define polygonal cylinders and the corresponding objective

functions  $\Psi$  computed using finite elements. A feedforward neural network with five input nodes (coordinates defining cylinders), three hidden layers, and one output node (scalar objective function  $\Psi_{ml}$ ) is then constructed to approximate the objective function.

We analyze the performance of the model, and show the dependency of the mean-squared error (MSE) on the training set size. Moreover, we analyze how many samples is needed to reach a desired approximation accuracy. For example, for polygonal cylinders defined by five randomly generated points, on average over 95% are predicted with mean absolute error less than 0.01 when the training set contains 200,000 data points. The data used for machine learning in this paper is available at [24].

The sound pressure as a function of frequency is nonlinear, and thus the linear regression methods perform poorly. In Sect. 6, we show the results of the approximation of the objective function by means of linear regression vs. feedforward fully connected neural network with ReLU nonlinearity. The proposed method performs much better, as expected.

For machine learning, we have used Tensorflow [3], and the stochastic gradient descent optimizer ADAM [16]. For the numerical computation of the average pressure, we used primarily the SLEPc [14], with user interfaces and numerical PDE tools FreeFem [13] and FEniCS [4] to the standard numerical packages.

Our method can be applied for analyzing frequency response in elastic bodies and fluid–structure interaction problems. In three-dimensional case the frequency response problems become computationally heavy, and the effectiveness of the stochastic gradient descent gives some hope for significant reduction of data needed for training.

The rest of this paper is organized as follows. In Sect. 2, the numerical method for computing the average sound pressure response is described. Using the numerical method, the data sets for polygonal cylinders are generated and the data sets are described in Sect. 3. In Sect. 4, we specify the feedforward dense neural network and the choices of for the training procedure. The model obtained after training is evaluated in Sect. 5, and compared to a linear model in Sect. 6.

## 2 Frequency Response Problem and Average Pressure

Assume that a domain  $\Omega$  is occupied by a inviscid, homogeneous, compressible fluid (liquid or gas). There are several options for choosing a primary variable for small amplitude vibrations: fluid displacement, acoustic pressure, or fluid velocity potential. We are going to use a description in terms of the scalar pressure function  $P$ . Let  $c$  be the speed of sound in the fluid, and  $\rho$  be the mass density of the fluid, both assumed not to depend on the pressure  $P$ . The equation of

motion without taking in account the damping is the wave equation for the acoustic pressure [30]:

$$\rho \frac{\partial^2 P(t, x)}{\partial t^2} - c^2 \Delta P(t, x) = F(t, x). \tag{1}$$

Here  $F$  is the applied load. The solution of the last equation is by linearity a sum of a particular solution to a non-homogeneous equation (forced motion) and the general solution of the homogeneous equation (natural motion). If the excitation is harmonic  $F(t, x) = f(x) \cos(\omega t) = f(x) \Re e^{i\omega t}$ , the forced motion is called the steady-state response. The real-valued pressure together with the phase angle is then called the dynamic frequency response of the system [10]. To eliminate the time dependency in the wave equation, we substitute  $P(t, x) = \Re(p(x)e^{i\omega t})$  into it and obtain a time-independent Helmholtz equation for the amplitude  $p$ :

$$-\Delta p(x) - \frac{\omega^2 \rho}{c^2} p(x) = f(x).$$

If the acoustic medium is contained in a bounded domain  $\Omega$ , we will need to impose boundary conditions on the boundary  $\partial\Omega$ . We will impose a harmonic load  $\cos(\omega t)$  on the part of the boundary  $\Gamma_D$ , which results in the non-homogeneous Dirichlet boundary condition  $p = 1$ . On the rest of the boundary  $\Gamma_N = \partial\Omega \setminus \Gamma_D$  is assumed to be sound hard (zero-flux condition). The problem in the frequency domain takes the form

$$\begin{aligned} -\Delta p - \frac{\omega^2 \rho}{c^2} p &= 0 && \text{in } \Omega, \\ p &= 1 && \text{on } \Gamma_D, \\ \nabla p \cdot \nu &= 0 && \text{on } \Gamma_N, \end{aligned} \tag{2}$$

where  $\nu$  is the exterior unit normal.

We are going to solve (2) analytically for cylinders  $\Omega$  with constant and non-constant cross section. We will obtain expressions for the mean value of the solution to (2) with respect to the spatial variable  $\langle p \rangle = |\Omega|^{-1} \int_{\Omega} p \, dx$ , and its average with respect to the spectral parameter

$$\lambda = \frac{\omega^2 \rho}{c^2}.$$

The domain may be an open set in Euclidean space  $\mathbf{R}^n$ . In what follows, we will work with  $\Omega$  a bounded Lipschitz domain in  $\mathbf{R}^2$ , that is a bounded open connected subset of  $\mathbf{R}^2$  with Lipschitz continuous boundary. Specifically,  $\Omega$  will be a finite cylinder with polygonal boundary.

We will in the sequel assume the quantities and variables to be scaled in such a way they are nondimensionalized, and

thereby also suppress units from both manipulations and figures.

### 2.1 Uniform Cylinders

We start with cylinders with constant cross section, where one can find explicit formulas for eigenfunctions and eigenvalues for the Laplace operator and therefore solve the frequency response problem analytically.

Let us denote  $\Omega = (0, 1) \times (-a, a)$  for  $r_{\min} \leq a \leq r_{\max}$  a uniform cylinder. The boundary of  $\Omega$  consists of two parts, and we denote  $\Gamma_D = \{(x_1, x_2) : x_1 = 0\}$  (the part where a Dirichlet boundary condition will be imposed) and  $\Gamma_N = \partial\Omega \setminus \Gamma_D$  (with a Neumann boundary condition). Consider the frequency response problem (2) in  $\Omega$ . By the Fredholm alternative, there exists a unique solution  $p_{\lambda} \in H^1(\Omega)$  to (2) if and only if  $\lambda$  is not an eigenvalue of the Laplace operator in the cylinder:

$$\begin{aligned} -\Delta \psi &= \lambda \psi && \text{in } \Omega, \\ \psi &= 0 && \text{on } \Gamma_D, \\ \nabla \psi \cdot \nu &= 0 && \text{on } \Gamma_N. \end{aligned} \tag{3}$$

By the Hilbert–Schmidt and the Riesz–Schauder theorems, the spectrum of (3) is positive, discrete, countably infinite, and each eigenvalue of finite multiplicity,

$$0 < \lambda_1 < \lambda_2 \leq \lambda_3 \leq \dots \leq \lambda_n \rightarrow \infty, \quad n \rightarrow \infty.$$

Moreover, the eigenfunctions  $\psi_i$  form an orthonormal basis under a proper normalization. By separation of variables, choosing a convenient enumeration, the eigenvalues  $\lambda_{i,k,l}$  to (3) are given by

$$\lambda_{i,k,l} = \mu_k + \eta_{i,l}, \quad i = 1, 2, \quad k, l = 0, 1, \dots, \tag{4}$$

where

$$\mu_k = \frac{(2k + 1)^2 \pi^2}{4}, \quad k = 0, 1, \dots, \tag{5}$$

$$\eta_{1,l} = \frac{l^2 \pi^2}{a^2}, \quad \eta_{2,l} = \frac{(2l + 1)^2 \pi^2}{4a^2}, \quad l = 0, 1, \dots \tag{6}$$

The sequences  $\mu_k > 0$  and  $\eta_{i,l} \geq 0$  are the Dirichlet–Neumann eigenvalues of the Laplace operator on  $(0, 1)$ , and the Neumann eigenvalues of the Laplace operator on  $(-a, a)$ , respectively. The eigenfunctions  $\psi_{i,k,l}$  to (3) corresponding to the eigenvalues  $\lambda_{i,k,l}$  are

$$\begin{aligned} \psi_{1,k,l} &= a_{1,k,l} \sin(\sqrt{\mu_k} x_1) \cos(\sqrt{\eta_{1,l}} x_2), \\ \psi_{2,k,l} &= a_{2,k,l} \sin(\sqrt{\mu_k} x_1) \sin(\sqrt{\eta_{2,l}} x_2), \end{aligned} \tag{7}$$

where  $a_{i,k,l}$  are  $L^2(\Omega)$  normalization factors defined by

$$\int_{\Omega} \psi_{i,k,l} \psi_{j,p,q} \, dx = \begin{cases} 1 & \text{if } (i, k, l) = (j, p, q), \\ 0 & \text{otherwise.} \end{cases}$$

Explicitly,  $a_{1,k,0} = \sqrt{1/a}$ , and otherwise  $a_{i,k,l} = \sqrt{2/a}$ .

Let  $\lambda \in \mathbf{R}$  not be an eigenvalue to (3). Then the method of separation of variables gives a solution  $p_{\lambda}$  to (2) in the case of uniform cylinder:

$$\begin{aligned} p_{\lambda} &= 1 + \sum_{k=0}^{\infty} \frac{\lambda}{\mu_k - \lambda} \frac{2}{\sqrt{\mu_k}} \sin(\sqrt{\mu_k} x_1) \\ &= \cos(\sqrt{\lambda} x_1) + \tan(\sqrt{\lambda}) \sin(\sqrt{\lambda} x_1). \end{aligned} \tag{8}$$

Remark that  $p_{\lambda}$  is constant in the  $x_2$ -direction for this particular choice of harmonic load  $p(0, x_2) = 1$ .

When analyzing the acoustic response, one could be interested in the average pressure defined for a frequency sweep, namely the average of  $p_{\lambda}$  with respect to  $x$  and  $\lambda$ . Let us first compute the average pressure  $\langle p_{\lambda} \rangle$  with respect to  $x$ :

$$\langle p_{\lambda} \rangle = \frac{1}{|\Omega|} \int_{\Omega} p_{\lambda} \, dx = \begin{cases} \frac{\tan(\sqrt{\lambda})}{\sqrt{\lambda}} & \text{if } \lambda > 0, \\ 1 & \text{if } \lambda = 0. \end{cases} \tag{9}$$

The pressure  $p_{\lambda}$  and its mean-value with respect to the space variable as a function of  $\lambda$  is shown in Fig. 1.

The form of the response  $\langle p_{\lambda} \rangle$  in (9) indicates that it could be challenging to numerically evaluate an integral of  $\langle p_{\lambda} \rangle$  in  $\lambda$  over an interval  $(\lambda_{\min}, \lambda_{\max})$  that contains a pole, because the computation of the Cauchy principal value of the integral requires both the location of the poles and their orders. For the uniform cylinder, we obtain the following explicit formula for the objective function:

$$\begin{aligned} \Psi &= \frac{1}{\lambda_{\max} - \lambda_{\min}} \text{p.v.} \int_{\lambda_{\min}}^{\lambda_{\max}} \langle p_{\lambda} \rangle \, d\lambda \\ &= \frac{1}{\lambda_{\max} - \lambda_{\min}} \text{p.v.} \int_{\lambda_{\min}}^{\lambda_{\max}} \frac{\tan(\sqrt{\lambda})}{\sqrt{\lambda}} \, d\lambda \\ &= \frac{2}{\lambda_{\max} - \lambda_{\min}} \log \left| \frac{\cos(\sqrt{\lambda_{\min}})}{\cos(\sqrt{\lambda_{\max}})} \right|, \end{aligned} \tag{10}$$

which is defined as long as both  $\lambda_{\min}, \lambda_{\max}$  are not eigenvalues of (3). More precisely, if  $\lambda = \lambda_{i,k,l}$  is an eigenvalue to (3), the response  $p_{\lambda}$  exists if and only if  $\int_{\Omega} \psi_{i,k,l} \, dx = 0$  for all eigenfunctions corresponding to  $\lambda_{i,k,l}$ , by the Fredholm alternative. One notes that  $\int_{\Omega} \psi_{1,k,l} \, dx = 0$  for  $l \geq 1$ , and  $\int_{\Omega} \psi_{2,k,l} \, dx = 0$  for  $l \geq 0$ . Thus for  $\lambda = \lambda_{1,k,l}$  with  $l \geq 1$ , and for  $\lambda = \lambda_{2,k,l}$  with  $l \geq 0$ , the solution  $p_{\lambda}$  is unique modulo a linear combination of the corresponding eigenfunctions. Such eigenfunctions do not contribute to the

mean value  $\langle p_{\lambda} \rangle$  and therefore also not to  $\Psi$ . It follows that (9) holds for  $\lambda \neq \mu_k$ , and (10) holds for  $\lambda_{\min}, \lambda_{\max} \neq \mu_k$ .

## 2.2 Cylinders with Varying Cross section

In this section, we will show how a Hilbert basis can be used to compute the average in  $\lambda$  of  $\langle p_{\lambda} \rangle$  (the objective function) in the case when cylinders have varying cross-section. The explicit formulas for the eigenvalues and eigenfunctions like (4)–(7) are not available any more, and we will use the finite elements to compute the eigenpairs of the Laplace operator.

Let  $p$ , as before, for a given  $\lambda$  solve the frequency response problem

$$\begin{aligned} -\Delta p - \lambda p &= 0 && \text{in } \Omega, \\ p &= 1 && \text{on } \Gamma_D, \\ \nabla p \cdot \nu &= 0 && \text{on } \Gamma_N. \end{aligned} \tag{11}$$

The cylinder  $\Omega$  is not uniform any more, and can be described by  $\Omega = \{x = (x_1, x_2) : x_1 \in (0, 1), x_2 \in I(x_1)\}$ , where  $I(x_1) = (-a(x_1), a(x_1))$  is an interval such that  $r_{\min} \leq a(x_1) \leq r_{\max}$ .

We will represent the solution  $p_{\lambda}$  of (11) in terms of the eigenpairs of the Laplace operator

$$\begin{aligned} -\Delta \psi &= \kappa \psi && \text{in } \Omega, \\ \psi &= 0 && \text{on } \Gamma_D, \\ \nabla \psi \cdot \nu &= 0 && \text{on } \Gamma_N. \end{aligned} \tag{12}$$

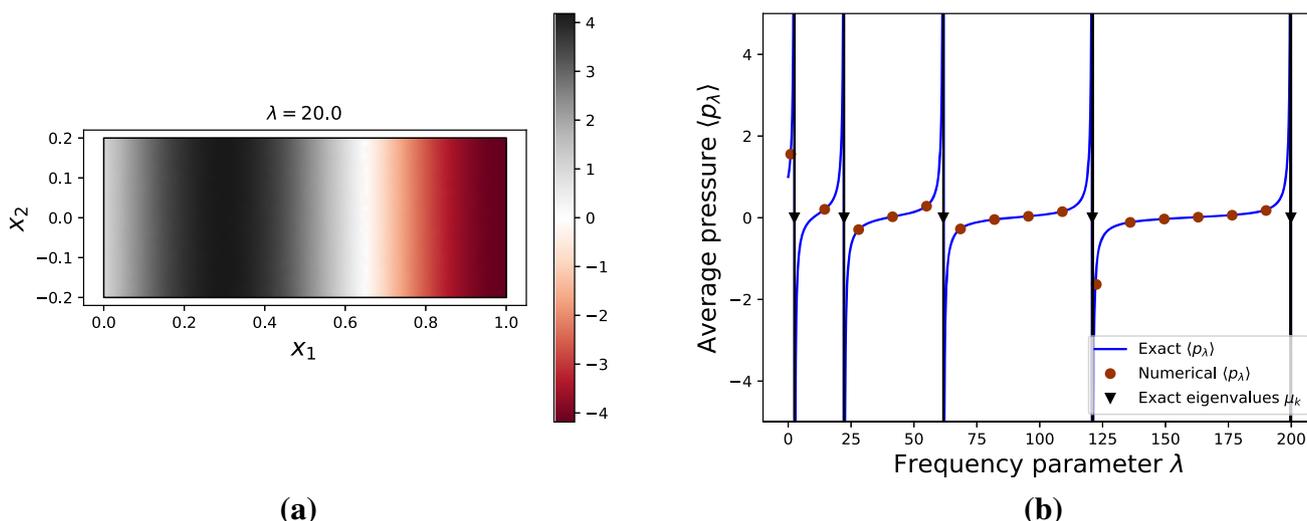
As before, the spectrum  $0 < \kappa_1 < \kappa_2 \leq \dots \leq \kappa_j \rightarrow \infty$  is discrete, and the eigenfunctions  $\psi_i$  form a Hilbert basis in  $L^2(\Omega)$ , and we assume that they are orthonormalized by  $\int_{\Omega} \psi_i \psi_j \, dx = \delta_{ij}$ . Writing  $p_{\lambda} = 1 + \sum_{i=1}^{\infty} \beta_i \psi_i$  and substituting into (11) one gets

$$\begin{aligned} p_{\lambda}(x) &= 1 + |\Omega| \sum_{i=1}^{\infty} \frac{\lambda}{\kappa_i - \lambda} \langle \psi_i \rangle \psi_i(x), \\ \langle \psi_i \rangle &= \frac{1}{|\Omega|} \int_{\Omega} \psi_i \, dx. \end{aligned} \tag{13}$$

The mean value of  $p_{\lambda}$  in  $\Omega$  is

$$\begin{aligned} \langle p_{\lambda} \rangle &= \frac{1}{|\Omega|} \int_{\Omega} p_{\lambda} \, dx = 1 + |\Omega| \sum_{i=1}^{\infty} \frac{\lambda}{\kappa_i - \lambda} \langle \psi_i \rangle^2 \\ &= 1 - |\Omega| \sum_{i=1}^{\infty} \langle \psi_i \rangle^2 + |\Omega| \sum_{i=1}^{\infty} \frac{\lambda}{\kappa_i - \lambda} \langle \psi_i \rangle^2. \end{aligned} \tag{14}$$

The pressure  $p_{\lambda}$  in a polygonal cylinder, and its mean-value  $\langle p_{\lambda} \rangle$  with respect to the space variable as a function of  $\lambda$  is shown in Fig. 2.



**Fig. 1** The response  $p_\lambda$  for  $\lambda = 20$  (a), and the average pressure  $\langle p_\lambda \rangle$  (b) for a uniform cylinder

Let us now average (14) over  $(\lambda_{\min}, \lambda_{\max})$  to get the objective function:

$$\begin{aligned} \Psi &= \frac{1}{\lambda_{\max} - \lambda_{\min}} \text{p.v.} \int_{\lambda_{\min}}^{\lambda_{\max}} \langle p_\lambda \rangle d\lambda \\ &= 1 + |\Omega| \sum_{i=1}^{\infty} \left[ \frac{\kappa_i}{\lambda_{\max} - \lambda_{\min}} \log \left| \frac{\kappa_i - \lambda_{\min}}{\lambda_{\max} - \kappa_i} \right| - 1 \right] \langle \psi_i \rangle^2. \end{aligned} \tag{15}$$

As we have seen above, for the case of a uniform cylinder, the right-hand side of (15) sums up to (10).

### 2.3 Numerical Computation of the Average Pressure Response

In order to compute the objective function  $\Psi$  (15) in the case of a non-uniform cylinder, we compute the eigenpairs  $(\kappa_j, \psi_j)$  of (12) using the first order Lagrange finite elements.

The variational formulation for (12) reads: Find  $\kappa \in \mathbf{R}$  and  $\psi \in H^1(\Omega) \setminus \{0\}$ ,  $\psi = 0$  on  $\Gamma_D$ , such that

$$\int_{\Omega} \nabla \psi \cdot \nabla v \, dx = \kappa \int_{\Omega} \psi v \, dx, \tag{16}$$

for any  $v \in H^1(\Omega)$ ,  $v = 0$  on  $\Gamma_D$ . For a triangulation mesh  $\mathcal{T}_h$  of  $\Omega$ , we consider Lagrange triangular finite elements of order 1 as a basis for the finite-dimensional subspace

$$V_{0h} = \left\{ v \in C(\bar{\Omega}) : v|_K \in \mathbb{P}_1 \text{ for all } K \in \mathcal{T}_h, v = 0 \text{ on } \Gamma_D \right\}. \tag{17}$$

The internal approximation for the variational formulation (16) is

$$\int_{\Omega} \nabla \psi_h \cdot \nabla v_h \, dx = \kappa_h \int_{\Omega} \psi_h v_h \, dx, \tag{18}$$

for all  $v_h \in V_{0h}$ . The eigenvalues of (18) form a finite increasing sequence

$$0 < \kappa_{h,1} \leq \kappa_{h,2} \leq \dots \leq \kappa_{h,n_{dl}}, \quad \text{with } n_{dl} = \dim V_{0h},$$

and there exists a basis in  $V_{0h}$  consisting of corresponding eigenfunctions which is orthonormal in  $L^2(\Omega)$ . A proof of this statement can be found in [2, Ch. 7.4].

We look for a solution of (18) in the form  $\psi_h(x) = \sum_{i=1}^{n_{dl}} U_i^h \phi_i(x)$ , where  $(\phi_i)_{1 \leq i \leq n_{dl}}$  is the basis in  $V_{0h}$ . Introducing the mass matrix  $\mathcal{M}_h$  and the stiffness matrix  $\mathcal{K}_h$ ,

$$\begin{aligned} (\mathcal{M}_h)_{ij} &= \int_{\Omega} \phi_i \phi_j \, dx, \\ (\mathcal{K}_h)_{ij} &= \int_{\Omega} \nabla \phi_i \cdot \nabla \phi_j \, dx, \quad 1 \leq i, j \leq n_{dl}, \end{aligned} \tag{19}$$

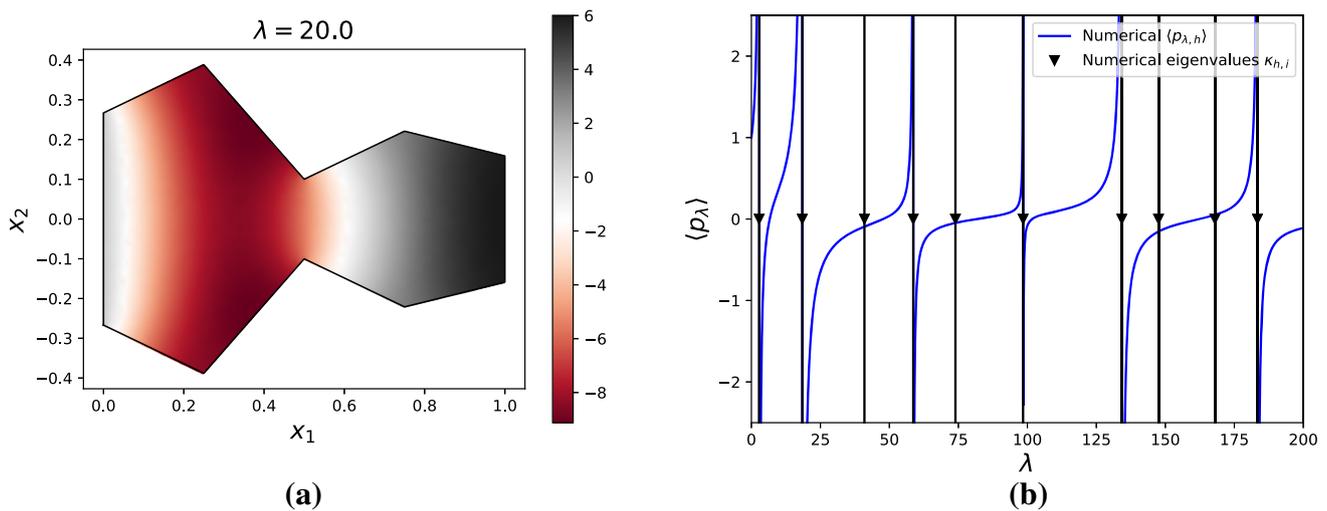
we get the following discrete finite-dimensional spectral matrix problem:

$$\mathcal{K}_h \psi_h = \kappa_h \mathcal{M}_h \psi_h. \tag{20}$$

The matrices  $\mathcal{M}_h$  and  $\mathcal{K}_h$  are symmetric and positive definite.

The error estimate for the eigenvalues corresponding to eigenfunctions in  $H^2(\Omega)$ , which is, for instance, the case if  $\Omega$  is convex in  $\mathbf{R}^2$ , is

$$|\kappa_i - \kappa_{h,i}| \leq C_i h^2,$$



**Fig. 2** The response  $p_\lambda$  for  $\lambda = 20$  (a), and the average pressure  $\langle p_\lambda \rangle$  (b) for the polygonal cylinder shown in Fig. 6b. The vertical lines in **b** indicate the poles

where  $C_i$  does not depend on  $h = \max\{\text{diam}(K) : K \in \mathcal{T}_h\}$ , but does depend on the number of the eigenvalue, that is why is it important to take a sufficiently fine mesh to get a good approximation for  $\kappa_i$  with large  $i$  (see, e.g., [2]). More precisely, if  $\kappa_i$  is an eigenvalue with eigenfunctions in  $H^{k+1}(\Omega)$ ,  $\Omega \subset \mathbf{R}^n$ , and  $2(k+1) > n$ , then  $|\kappa_i - \kappa_{h,i}| \leq C_i h^{2k}$ .

In the numerical method, we truncate the series in (15) at  $i = N$ ,

$$\Psi_h = 1 + |\Omega| \sum_{i=1}^N \left[ \frac{\kappa_{h,i}}{\lambda_{\max} - \lambda_{\min}} \log \left| \frac{\kappa_{h,i} - \lambda_{\min}}{\lambda_{\max} - \kappa_{h,i}} \right| - 1 \right] \langle \psi_{h,i} \rangle^2, \quad (21)$$

where  $N$  is chosen such that the sum ranges over the eigenvalues up to at least  $10\lambda_{\max}$ , and the number of degrees of freedom  $\dim V_{0h}$  is at least 10 times greater than the greatest eigenvalue  $\kappa_{h,N}$  used in the computation. This ensures that the eigenvalues  $\kappa_{h,i}$  and the eigenfunctions  $\psi_{h,i}$  of the discrete eigenvalue problem (20) are correct approximations to the exact eigenvalues  $\kappa_{h,i}$  and exact eigenfunctions  $\psi_{h,i}$  of (16).

In order to evaluate the accuracy of the method, in the case of uniform cylinders, we can compare the exact objective function (10) (the blue curve) with its numerical approximation (21) (the dots). The result is presented in Fig. 3a. The peaks of the objective function are located at the eigenvalues  $\mu_k$  since  $p_\lambda$  has poles at these points. The graph is valid for a uniform cylinder of arbitrary radius, because the pressure does not depend on the transverse variable. It is important to note that numerical integration by means of the trapezoidal rule of the exact response  $\langle p_\lambda \rangle$  given by (9) with respect to  $\lambda$  does not give a good approximation for  $\Psi$ . In Fig. 3b one can see that the numerical integration fails after the first eigen-

value. The reason for this is the singular behavior of  $p_\lambda$  near the eigenvalues  $\mu_k$  (Fig. 4).

In the case of non-uniform cylinders we do not have any explicit formulas any more, so we investigate numerically the rate of convergence for the approximation of  $\Psi$  by  $\Psi_h$ . For the sake of completeness, we present the convergence rate for both uniform and non-uniform cylinders.

In Fig. 5a, one can see a clear quadratic decay of  $|\Psi - \Psi_h|$  with respect to mesh size  $h$  for uniform cylinders (convex). The objective function  $\Psi_h$  (21) is computed as the average over  $(0, \lambda_{\max})$  for several  $\lambda_{\max}$  and for uniform mesh refinements. The quadratic decay of the error with respect to the mesh size  $h$  is expected for first order polynomial approximations of a smooth function in  $L^2(\Omega)$ . In  $\mathbf{R}^2$ , the number of degrees of freedom  $\dim V_{0h}$  grows as  $h^{-2}$  for uniform mesh refinement, which suggests an expected rate of decay  $(\dim V_{0h})^{-1}$  for non-degenerate uniform mesh refinement.

In Fig. 5b, we present the rate of convergence while refining the mesh for several cylinders with polygonal boundary. We observe a subquadratic convergence rate with respect to the mesh size.

## 2.4 Shape Derivative of the Average Pressure Response

In this section, we compute the derivative  $\Psi'$  of the objective function

$$\Psi = \frac{1}{\lambda_{\max} - \lambda_{\min}} \int_{\lambda_{\min}}^{\lambda_{\max}} \langle p_\lambda \rangle d\lambda, \quad (22)$$

with respect to certain variations of the convex cylindrical domains  $\Omega$  in  $\mathbf{R}^2$ . The purpose of this is twofold. One, it ascertains that the accuracy of our trained model on Lip-

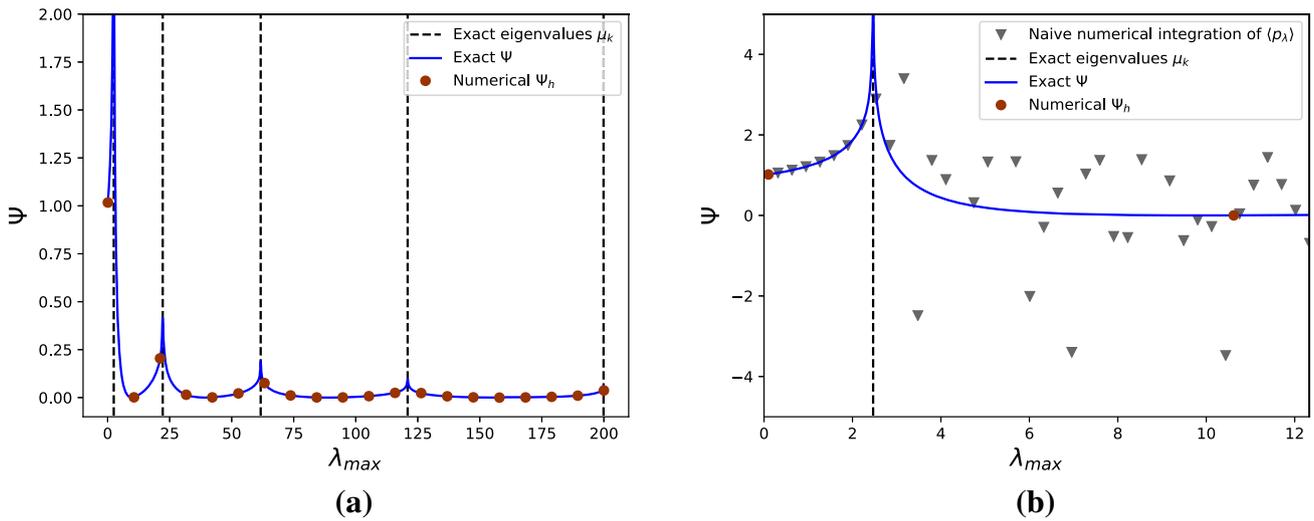


Fig. 3 Objective function in a uniform cylinder for intervals  $(0, \lambda_{\max})$  with  $\lambda_{\max}$  on the horizontal axis

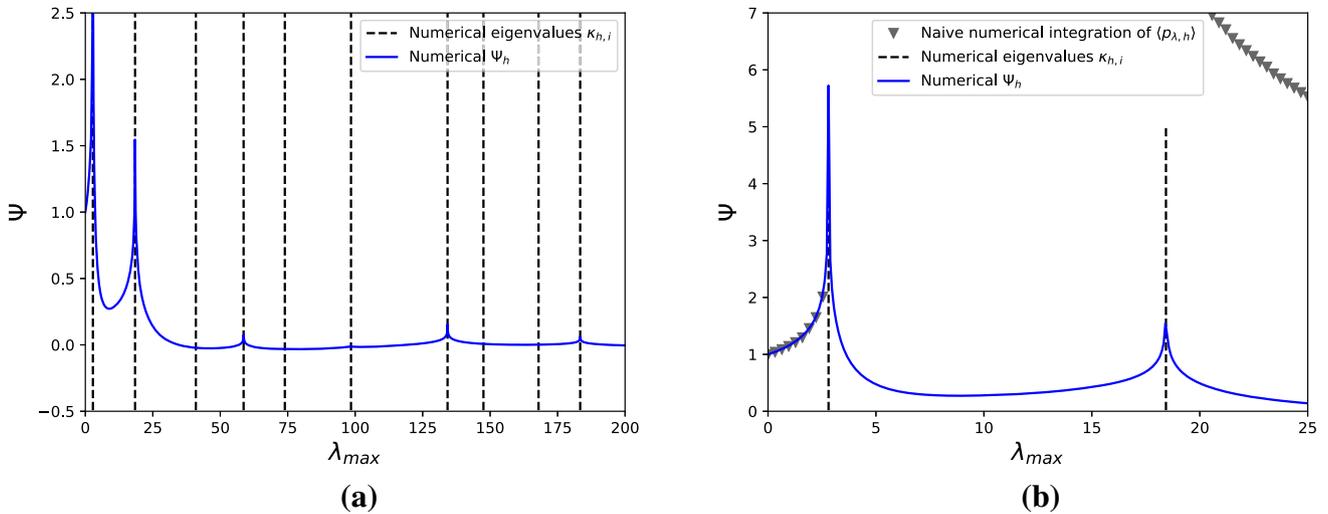


Fig. 4 The objective function in a non-uniform cylinder in Fig. 6b for intervals  $(0, \lambda_{\max})$  with  $\lambda_{\max}$  on the horizontal axis

schitz domains that are close in a precise sense to certain polygonal domains in our evaluation sets. Two, it enables for some boosting of the training sets that are otherwise somewhat costly to generate by the method we have chosen.

For a vector field  $V$  and a parameter  $t$ , we introduce the bi-Lipschitz transformation  $T(x) = x + tV(x)$ . We denote by  $\Omega_t$  the image of  $\Omega$  under  $T$ . Let  $\Psi_t$  be the value of (22) for the domain  $\Omega_t$ ,  $\Psi = \Psi_0$ . With  $\Psi'_0$ , the derivative of  $\Psi$  at  $t = 0$ ,  $\Psi$  is linearized as

$$\Psi_t = \Psi_0 + t\Psi'_0 + o(t),$$

as  $t$  tends to zero.

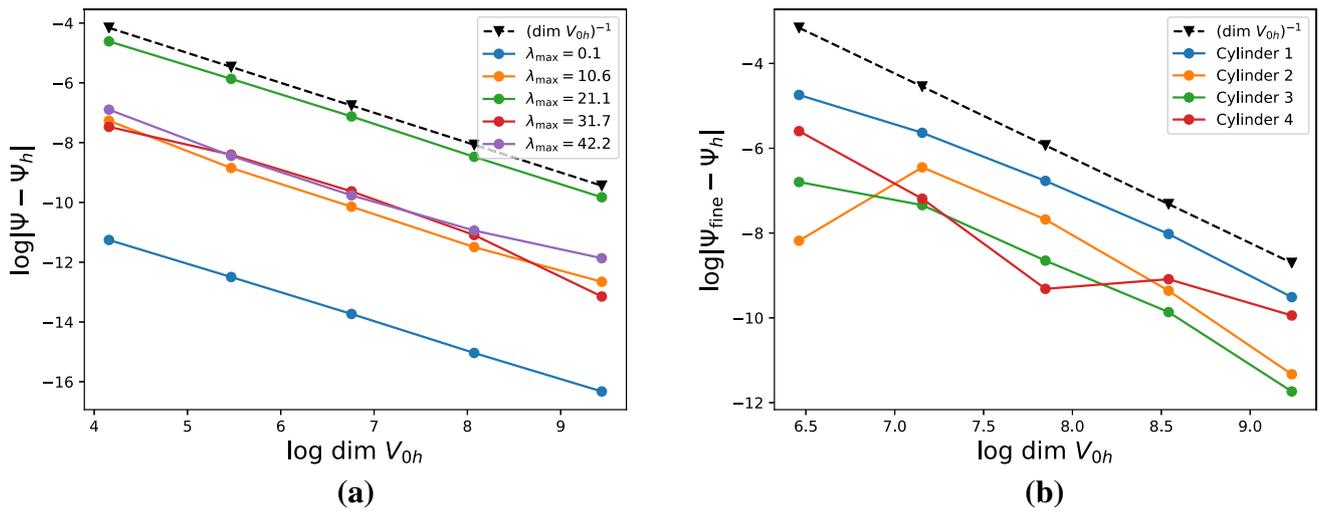
In this section, we employ standard techniques of domain variations in the theory of elliptic equations. We refer to [12, 17,20] for expositions.

**Lemma 2.1** *Let  $V \in W^{1,\infty}(\Omega)$  be a solenoidal vector field on convex  $\Omega$ . Suppose that the interval  $[\lambda_{\min}, \lambda_{\max}]$  does not contain any eigenvalue  $\kappa_i$  for which  $\langle \psi_i \rangle \neq 0$ . Then the shape derivative of the objective function (22) is given by*

$$\Psi' = \lim_{t \rightarrow 0} \frac{\Psi_t - \Psi_0}{t} = \sum_{i,j=1}^{\infty} c_{i,j} \langle \nabla V \nabla \psi_i \cdot \nabla \psi_j \rangle \langle \psi_i \rangle \langle \psi_j \rangle,$$

where for  $\kappa_i = \kappa_j$ ,

$$c_{i,j} = \frac{2|\Omega|^2}{\lambda_{\max} - \lambda_{\min}} \left[ \log \left| \frac{\lambda_{\max} - \kappa_i}{\kappa_i - \lambda_{\min}} \right| - \frac{\kappa_i}{\lambda_{\max} - \kappa_i} - \frac{\kappa_i}{\kappa_i - \lambda_{\min}} \right],$$



**Fig. 5** Rate of convergence of the finite element approximation of the objective function  $\Psi$ . **a** The absolute error for a uniform cylinder. **b** The estimated rate of convergence for four samples of non-uniform cylinders, and fixed  $\lambda_{\max} = 60$

and for  $\kappa_i \neq \kappa_j$ ,

$$c_{i,j} = \frac{2|\Omega|^2}{\lambda_{\max} - \lambda_{\min}} \left[ \frac{\kappa_i}{\kappa_i - \kappa_j} \log \left| \frac{\lambda_{\max} - \kappa_i}{\kappa_i - \lambda_{\min}} \right| - \frac{\kappa_j}{\kappa_i - \kappa_j} \log \left| \frac{\lambda_{\max} - \kappa_j}{\kappa_j - \lambda_{\min}} \right| \right].$$

**Proof** One notes that by elliptic regularity  $p_\lambda \in H^2(\Omega)$ . Denote by  $\dot{p}_\lambda \in H^1(\Omega, \Gamma_D)$  the material derivative of  $p_\lambda$ :  $\dot{p}_\lambda = p'_\lambda + \nabla p_\lambda \cdot V$ , where  $p'_\lambda \in H^1(\Omega)$  denotes the shape derivative of the response with respect to  $V$ . By the regularity of  $V$ , there exist a Sobolev extension, and thereby the shape derivatives with respect to  $V$  of the response and the associated linear and bilinear forms exist in the sense of Fréchet with respect to the parameter  $t$ .

A direct computation of the Gateaux derivative gives

$$\begin{aligned} \Psi' &= \frac{1}{\lambda_{\max} - \lambda_{\min}} \int_{\lambda_{\min}}^{\lambda_{\max}} \langle p_\lambda \rangle' d\lambda \\ &= \frac{1}{\lambda_{\max} - \lambda_{\min}} \int_{\lambda_{\min}}^{\lambda_{\max}} (\langle \dot{p}_\lambda \rangle + \langle p_\lambda \text{div} V \rangle - \langle p_\lambda \rangle \langle \text{div} V \rangle) d\lambda. \end{aligned}$$

To compute  $\langle \dot{p}_\lambda \rangle$ , we note that  $\dot{p}_\lambda$  is an admissible test function in the variational form of the equation for  $p_\lambda$ :

$$\int_{\Omega} \nabla(p_\lambda - 1) \cdot \nabla v \, dx - \lambda \int_{\Omega} (p_\lambda - 1)v \, dx = \lambda \int_{\Omega} v \, dx.$$

Therefore,

$$\begin{aligned} \lambda \int_{\Omega} \dot{p}_\lambda \, dx &= \int_{\Omega} \nabla(p_\lambda - 1) \cdot \nabla \dot{p}_\lambda \, dx \\ &\quad - \lambda \int_{\Omega} (p_\lambda - 1)\dot{p}_\lambda \, dx \end{aligned}$$

$$\begin{aligned} &= 2 \int_{\Omega} \nabla(\nabla p_\lambda \cdot V) \cdot p_\lambda \, dx \\ &\quad - \int_{\Omega} \text{div}(|\nabla p_\lambda|^2 V) \, dx \\ &\quad - 2\lambda \int_{\Omega} (\nabla p_\lambda \cdot V)p_\lambda + \lambda \int_{\Omega} \nabla p_\lambda \cdot V \, dx \\ &\quad + \lambda \int_{\Omega} \text{div}(p_\lambda(p_\lambda - 1)V) \, dx, \end{aligned}$$

where one in the second step has differentiated the variational form of the equation for  $p_\lambda$  and in that way eliminated  $\dot{p}_\lambda$  by using  $p_\lambda - 1$  as a test function. Indeed,

$$\begin{aligned} &\int_{\Omega} \nabla \dot{p}_\lambda \cdot \nabla v \, dx - \lambda \int_{\Omega} \dot{p}_\lambda v \, dx \\ &= - \int_{\Omega} \nabla(\nabla p_\lambda \cdot V) \cdot \nabla v \, dx \\ &\quad - \int_{\Omega} \nabla p_\lambda \cdot (\nabla v \cdot V) \, dx \\ &\quad + \int_{\Omega} \text{div}((\nabla p_\lambda \cdot \nabla v)V) \, dx \\ &\quad + \lambda \int_{\Omega} (\nabla p_\lambda \cdot V)v \, dx + \lambda \int_{\Omega} p_\lambda(\nabla v \cdot V) \, dx \\ &\quad - \lambda \int_{\Omega} \text{div}(p_\lambda v V) \, dx, \end{aligned}$$

for any  $v \in H^2(\Omega) \cap H^1(\Omega, \Gamma_D)$ . After some manipulation of terms, one concludes that

$$\begin{aligned} \langle p_\lambda \rangle' &= \langle \dot{p}_\lambda \rangle + \langle p_\lambda \text{div} V \rangle - \langle p_\lambda \rangle \langle \text{div} V \rangle \\ &= -\langle p_\lambda \rangle \langle \text{div} V \rangle + \langle p_\lambda^2 \text{div} V \rangle \\ &\quad - \frac{1}{\lambda} \langle |\nabla p_\lambda|^2 \text{div} V \rangle + \frac{2}{\lambda} \langle \nabla V \nabla p_\lambda \cdot \nabla p_\lambda \rangle, \end{aligned}$$

which for solenoidal  $V$  reduces to

$$\langle p_\lambda \rangle' = \frac{2}{\lambda} \langle \nabla V \nabla p_\lambda \cdot \nabla p_\lambda \rangle.$$

By substituting the expansion

$$p_\lambda = 1 + \sum_{i=1}^\infty \beta_i \psi_i, \quad \beta_i = |\Omega| \frac{\lambda}{\kappa_i - \lambda} \langle \psi_i \rangle,$$

and integrating in  $\lambda$  the desired formula is obtained, by the Fubini theorem.  $\square$

For instance, in the case of a uniform cylinder,

$$p_\lambda = \cos(\sqrt{\lambda}x_1) + \tan(\sqrt{\lambda}) \sin(\sqrt{\lambda}x_1).$$

For solenoidal  $V \in W^{1,\infty}(\Omega)$ , the shape derivative of the averaged pressure response is

$$\begin{aligned} \langle p_\lambda \rangle' &= \frac{2}{\lambda} \langle (\partial_1 p_\lambda)^2 \partial_1 V_1 \rangle \\ &= 2 \langle (\sin(\sqrt{\lambda}x_1) - \tan(\sqrt{\lambda}) \cos(\sqrt{\lambda}x_1))^2 \partial_1 V_1 \rangle, \end{aligned}$$

and

$$\begin{aligned} \Psi' &= \frac{1}{\lambda_{\max} - \lambda_{\min}} \left\langle \left[ \frac{2 \sin(\sqrt{\lambda}) \cos(\sqrt{\lambda}x_1)^2}{\sqrt{\lambda} \cos(\sqrt{\lambda})} \right. \right. \\ &\quad \left. \left. - \frac{2 \sin(\sqrt{\lambda}x_1) \cos(\sqrt{\lambda}x_1)}{\sqrt{\lambda}} \right]_{\lambda_{\min}}^{\lambda_{\max}} \frac{\partial V_1}{\partial x_1} \right\rangle \\ &+ \frac{1}{\lambda_{\max} - \lambda_{\min}} \left\langle \left[ \frac{\sin(\sqrt{\lambda}x_1) \cos(\sqrt{\lambda}x_1)}{\sqrt{\lambda} \cos(\sqrt{\lambda})^2} \right. \right. \\ &\quad \left. \left. + \frac{x_1}{\cos(\sqrt{\lambda})^2} \right]_{\lambda_{\min}}^{\lambda_{\max}} \frac{\partial V_1}{\partial x_1} \right\rangle. \end{aligned}$$

A somewhat more direct proof of Lemma 2.1 goes as follows.

**A second proof of Lemma 2.1** Let  $\tilde{p}_\lambda^t$  be such that  $\tilde{p}_\lambda^t - 1 \in H^1(\Omega_t, T(\Gamma_D))$  and

$$\int_{\Omega_t} \nabla \tilde{p}_\lambda^t \cdot \nabla v \, dx - \lambda \int_{\Omega_t} \tilde{p}_\lambda^t v \, dx = 0,$$

for all  $v \in H^1(\Omega_t, T(\Gamma_D))$ , supposing that  $t$  is small enough. Then, by the Lipschitz transform of the Sobolev space, using that  $T$  and  $T^{-1}$  are Lipschitz,

$$p_\lambda^t = \tilde{p}_\lambda^t \circ T$$

is such that  $p_\lambda^t - 1 \in H^1(\Omega, \Gamma_D)$  and it is the solution to

$$\begin{aligned} \int_{\Omega} (\nabla T^{-T} \nabla p_\lambda^t \cdot \nabla T^{-T} \nabla v) |\det \nabla T| \, dx \\ - \lambda \int_{\Omega} p_\lambda^t v |\det \nabla T| \, dx = 0, \end{aligned} \quad (23)$$

for all  $v \in H^1(\Omega, \Gamma_D)$ . Here,  $\nabla T^{-T}$  denotes the transpose of the inverse of the gradient of  $T$ . Using that  $\psi_i$  form a Hilbert basis, let

$$p_\lambda^t = \sum_i \gamma_i(t) \psi_i.$$

Recall that

$$p_\lambda = \sum_i \beta_i \psi_i.$$

Using that

$$\begin{aligned} \nabla T^{-T} &= 1 - t \nabla V + o(t), \\ |\det \nabla T| &= 1 + t \operatorname{div} V + o(t), \end{aligned}$$

as  $t$  tends to zero, and expanding the coefficients  $\gamma_i(t)$  as

$$p_\lambda^t = \sum_i (\gamma_i^0 + \gamma_i^1 t) \psi_i + o(t),$$

give by equation (23) that  $\gamma_i^0 = \beta_i$  and

$$p_\lambda^t = p_\lambda + t \sum_i \gamma_i^1 \psi_i + o(t),$$

where

$$\begin{aligned} \gamma_i^1 &= \sum_j \frac{\beta_j}{\kappa_i - \lambda} \left[ \int_{\Omega} \nabla V \nabla \psi_i \cdot \nabla \psi_j \, dx \right. \\ &\quad \left. + \int_{\Omega} \nabla \psi_i \cdot \nabla V \nabla \psi_j \, dx \right. \\ &\quad \left. - \int_{\Omega} (\nabla \psi_i \cdot \nabla \psi_j) \operatorname{div} V \, dx + \lambda \int_{\Omega} \psi_i \psi_j \operatorname{div} V \, dx \right] \\ &\quad + \frac{\lambda}{\kappa_i - \lambda} \int_{\Omega} \psi_i \operatorname{div} V \, dx. \end{aligned}$$

The shape derivative may then be computed as follows:

$$\begin{aligned} \Psi' &= \frac{1}{\lambda_{\max} - \lambda_{\min}} \left( \lim_{t \rightarrow 0} \frac{\int_{\lambda_{\min}}^{\lambda_{\max}} \langle p_\lambda^t - p_\lambda \rangle \, d\lambda}{t} \right. \\ &\quad \left. - \int_{\lambda_{\min}}^{\lambda_{\max}} \langle p_\lambda \rangle \langle \operatorname{div} V \rangle \, d\lambda \right) \end{aligned}$$

$$\begin{aligned}
&= \frac{1}{\lambda_{\max} - \lambda_{\min}} \lim_{t \rightarrow 0} \frac{\int_{\lambda_{\min}}^{\lambda_{\max}} \langle p_{\lambda}^t - p_{\lambda} \rangle d\lambda}{t} - \langle \operatorname{div} V \rangle \Psi \\
&= \frac{1}{\lambda_{\max} - \lambda_{\min}} \int_{\lambda_{\min}}^{\lambda_{\max}} \sum_i \gamma_i^1 \langle \psi_i \rangle d\lambda - \langle \operatorname{div} V \rangle \Psi.
\end{aligned}$$

For solenoidal  $V$ , the computation results in

$$\begin{aligned}
\Psi' &= \frac{2|\Omega|^2}{\lambda_{\max} - \lambda_{\min}} \\
&\quad \sum_{i,j} \int_{\lambda_{\min}}^{\lambda_{\max}} \frac{\lambda}{(\kappa_i - \lambda)(\kappa_j - \lambda)} d\lambda \\
&\quad \langle \nabla V \nabla \psi_i \cdot \nabla \psi_j \rangle \langle \psi_i \rangle \langle \psi_j \rangle.
\end{aligned}$$

After evaluation of the integrals, the desired formula is again obtained.

The condition that  $V$  is solenoidal in Lemma 2.1 is only for presentation purpose. The case of non-solenoidal  $V \in W^{1,\infty}(\Omega)$  is covered by both of the above proofs, except for the last step of integration in  $\lambda$ , which then results in lengthier formulas.

### 3 Data Sets

The data sets consist of the randomly generated coordinates defining polygonal cylinders and the corresponding objective function  $\Psi$ . The coordinates are generated in such a way that the radius of a cylinder varies between 0.1 and 0.5. The coordinates were sampled as independent and identically uniformly distributed random variables, using a pseudo-random number generator. The number of points defining the polygonal boundary might be 1 (uniform cylinder, as in Fig. 6a), 2 (cone segment), 3, 4, and 5 (as shown in Fig. 6b). The objective function is computed, as described in the previous section, with finite elements. In total, we have about 700,000 data points in the main data set, which we call Random 5. We have also generated some smaller data sets for evaluation purpose, as well as a set of 100,000 data points for uniform cylinders. The uniform cylinder set is important because it is a set for which we have very high accuracy in the numerical value of the objective function.

For the non-uniform cylinders we have no guarantee that the error is small, as we are doing non-rigorous numerics with finite elements and floating point arithmetic without tracing or bounding the round off errors. In the choice of mesh sizes we have employed standard indicators such as numerically observing what happens to the solution and the objective function under mesh refinement. By means of Lemma 2.1, we can guarantee for certain intervals and modulo round off errors that the error stays below a threshold  $\tau$

for small enough domain perturbations of the convex cylinders if the error on the reference is bounded by  $2\tau$ . This can be exemplified with perturbations of uniform cylinders under bi-Lipschitz mappings close to the unit. For the method of validated numerics, bounding the round off errors, we refer to [32].

In Table 1, an overview of the data sets is provided, where we have indicated the size of different data subsets used for training and evaluation (test), as well as the statistics of mean, variance, minimum, and maximum of the objective function  $\Psi_h$ .

The data and the code for data generation is available on the GitHub [24].

## 4 Feedforward Dense Neural Network for Approximation of Average Pressure

As a base model for the average pressure we will use a feedforward fully connected neural network, with the radii of the cylinder at a discrete set of points (1, 2, 3 or 5) as input. The base model is nonlinear, and it consists of three hidden layers, each with ReLU activation. We will compare this with a linear model as a point of reference.

### 4.1 Structure of the Neural Network

The main goal of the paper is to construct a learned algorithm which for a given polygonal cylinder outputs the corresponding average pressure. In this section, we describe the main ideas and principles underlying the dense neural network which is used for the prediction of the average pressure level  $\Psi$  over a given frequency range. For a rigorous and at the same time concise description of the the deep neural networks construction we refer to [19,31].

Let us call  $\Psi_{\text{ml}}$  the function that for given cylinders outputs the objective function  $\Psi_h$ . Inputs to this function are the radial coordinates of the points defining the boundary, 5 along a uniform segmentation of the interval  $[0, 1]$ . The output is one real number  $\Psi_{\text{ml}}$ , that is we have a regression type of problem. Assume that we have a data set containing values of  $\Psi$  for  $N$  polygonal cylinders. We will train a learning function on a part of this set. Assigning weights to the inputs, we create a function so that the error in the approximation of  $\Psi$  is minimized. Then we evaluate the performance of our function  $\Psi_{\text{ml}}$  by applying it to the unseen data and measure the accuracy of the predicted average pressures.

The simplest learning function is affine, but it is usually too simple to give a good result. In Sect. 6, for the sake of illustration, we compare the results for linear regression and the proposed algorithm, and show that linear regression gives a poor result for nonuniform cylinders. A widely used choice of nonlinearity is a composition of linear func-

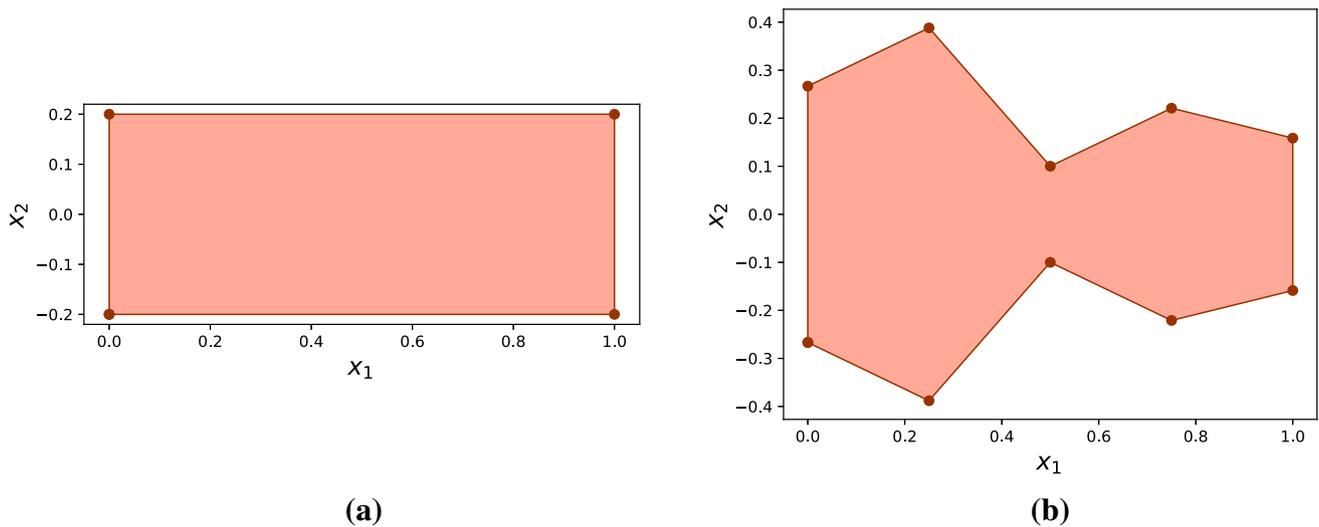


Fig. 6 A uniform cylinder (a) and a non-uniform polygonal cylinder (b)

Table 1 Data sets split into training and test categories

Data set	Size	Category	Mean	Variance	Min	Max
Random 5	200,000	Training	0.0638	0.00860	-0.0918	0.882
Random 5	500,000	Test	0.0636	0.00864	-0.0909	0.991
Random 5 (fine)	20,000	Test	0.0632	0.00869	-0.0876	0.734
Random 3	10,000	Test	0.0833	0.00494	-0.0508	0.631
Random 2	10,000	Test	0.0769	0.00137	-0.0267	0.186
Uniform	100,000	Test	0.0742	0	0.0742	0.0742

The statistics mean, variance, min, and max of  $\Psi_h$  are truncated

tions with so-called sigmoidal functions (having S-shaped graph). A smooth sigmoidal function has been a popular choice, but after that numerous numerical experiments indicated that this might not be an optimal choice. In many examples, it has turned out that a piecewise linear function  $\text{ReLU}(x) = \max\{0, x\}$  (the positive part  $x^+$  of the linear function  $x$ , sometimes called a rectified linear unit) performs better [31]. Specifically, we consider a learning function  $\Psi_{ml}$  in the form of a composition

$$\Psi_{ml}(v) = L_M(R(L_{M-1}(R \cdots (L_1 v)))) \tag{24}$$

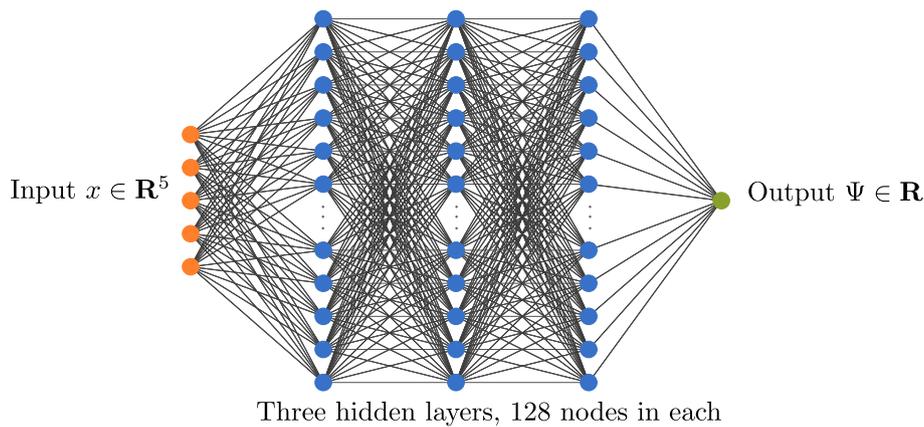
where  $L_k v = A_k v + b_k$  are affine functions, and  $Rx = \text{ReLU}(x)$  is the nonlinear ramp function (rectifies linear unit), the activation function. In this way, the output is a recursively nested composition function of inputs: input to the first hidden layer, input from the first to the second hidden layer, ..., input from the last hidden layer to output layer. Each hidden layer in Fig. 7 contains both the linear  $L_k$  and the nonlinear activation function  $R$ . For our purpose seems sensible to have three hidden layers with 128 nodes in each layer. In this sense, we use a what could be called a deep neural network. The elements of the matrices  $A_k$  and the bias vectors  $b_k$  are

weights in our learning function. Note that to have 128 nodes in the first hidden layer, the first matrix  $A_1$  should have 128 rows and 5 columns. The goal of the learning is to choose the weights to minimize the error over training sample, such that it generalize well to unseen data.

### 4.2 Hyperparameters and Training

The choice of hyperparameters is important for the learning of the model. We choose the following:

- **Nonlinearity:** ReLU.
- **Hidden layers:** Three hidden layers, 128 nodes in each.
- **Optimizer:** ADAM.
- **Learning rate:** The step size  $s_k$  in the gradient descent is scheduled to have polynomial decay from  $s_0 = 0.001$  to 0.0001 according to  $s_k = s_0/\sqrt{k}$  in 10,000 steps.
- **Loss function:** Mean squared error (MSE).
- **Validation split:** 20% of the training set.
- **Early stopping:** In order to avoid overtraining, the change of the MSE for the validation set  $10^{-5}$  counts as an improvement. If we have 25 iterations without improve-



**Fig. 7** A feedforward network with three hidden layers

ment, we stop and use the weights that give the minimum MSE up to this point.

- **Initializer:** GlorotUniform.

In Appendix A, we provide a hyperparameter grid that indicates together with the results of Sect. 5, that the performance of the model is not that sensitive to the values of the parameters around the chosen ones.

## 5 Performance of the Feedforward Neural Network Model

Our aim is to construct a ML algorithm to approximate  $\Psi$  based on the data for  $\Psi_h$  (computed with finite elements) with the same accuracy on the unseen data as the numerical error  $\Psi - \Psi_h$ . Here, we present the measured performance on our data sets.

In Fig. 8, we present the dependence of the error on the size of the training set. For each training set, we train the model ten times and take the mean of the mean squared error (mean MSE). The purple curve in Fig. 8a shows the MSE for polygonal cylinders with five random points defining the boundary. The objective function  $\Psi_h$  is computed on a mesh with density approximately three times higher than regular (referred to as "fine"). In Fig. 8b, we present the percentage of the unseen data used for the test that gives the mean absolute error less than 0.01. Again, here we take the mean value of the percentage after ten training sessions, the reason being the stochastic gradient descent algorithms used which results in some nonzero variance.

The choice of the threshold 0.01 is based on numerical indication of what is a bound on the error for almost all data points. We do not guarantee this bound on the error in the numerical data. In spite of that, we believe it serves as an illustrative example in that similar behavior in the accuracy

of the machine learning model on unseen data is expected if this threshold is increased, or if the error in the data had been zero.

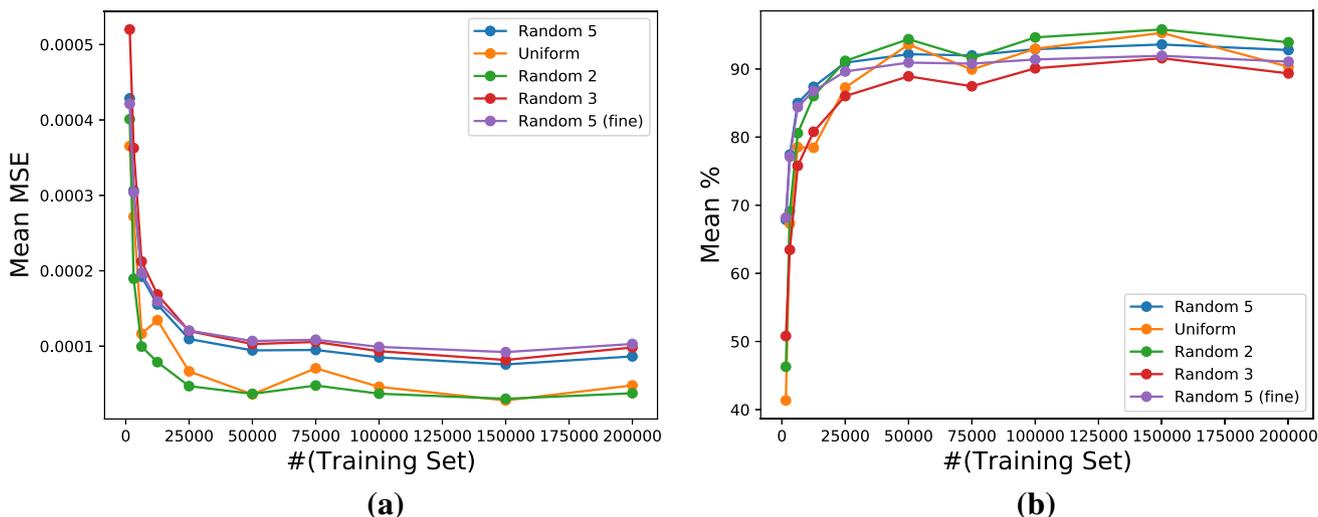
Numerical values for the best model are presented in Tables 2 and 3 in Sect. 6. For example, for polygonal cylinders defined by five randomly generated points, over 95% are predicted with mean absolute error 0.01 (the accuracy of the numerical data) if the training set contains 200,000 data points. The MSE for our model trained on 200,000 data points is  $2.31 \cdot 10^{-5}$  for uniform cylinders and  $5.5 \cdot 10^{-5}$  for polygonal cylinders. Clearly, the MSE is much smaller than the variance in the data for the test set Random 5 (Table 1).

Since we present the mean-value of the MSE, we need to analyze the standard deviation. In Fig. 9, we present the MSE error and the percentage of unseen data with absolute error below threshold (red curve) together with the standard deviation for polygonal cylinders with five random points (shadowed region). Figure 10 illustrates the MSE for different epoch numbers for training and validation sets. We observe that the error on the training set, on the validation set, and on the test set are close.

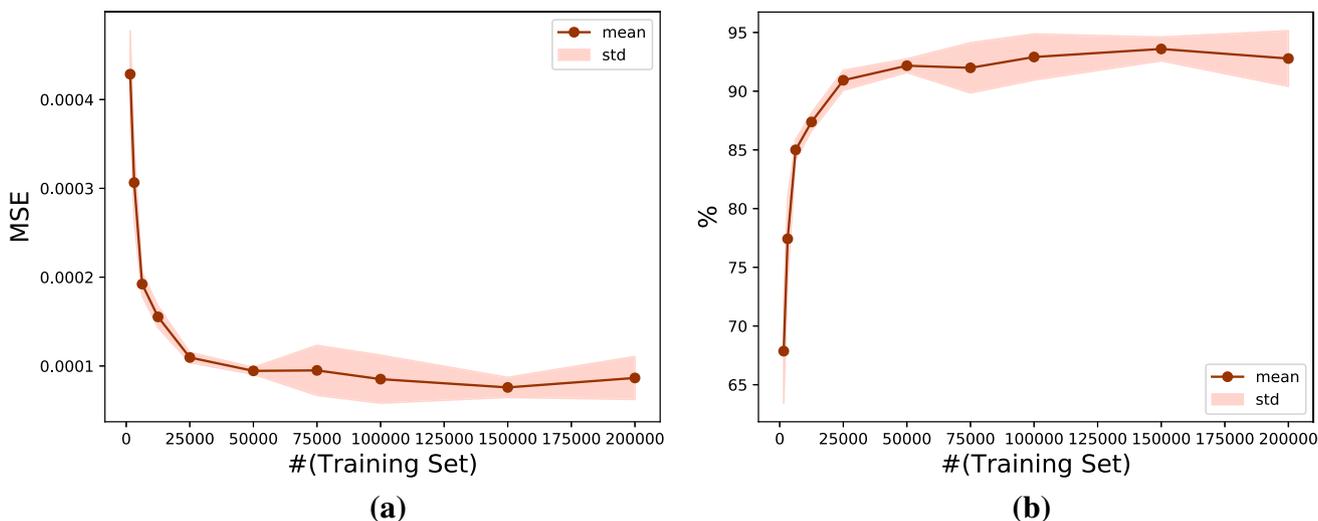
When it comes to the choice of hyperparameters, the numerical experiments have shown that reducing number of layers to two shows poor result, while increasing the number of layers and number of nodes in each layer does not improve much the result. We have also tested YOGI [26], but we did not manage to tune it to perform any better than ADAM. The decaying learning rate gives better accuracy than a constant one.

### 5.1 Performance on an Out of Sample Set

To complement the evaluation of the trained models on the unseen data, we here include an out of sample set. Specifically, we choose a one-parameter family of convex symmetric cylinders defined as follows, and illustrated in Fig. 11. For



**Fig. 8** **a** Dependence of the MSE on the training set size. **b** Percentage of the unseen data with absolute error less than 0.01 as a function of the training set size



**Fig. 9** **a** MSE and the standard deviation for polygonal cylinders with 5 random points. **b** Percentage of the unseen data and the standard deviation with absolute error less than 0.01 as a function of the training set size

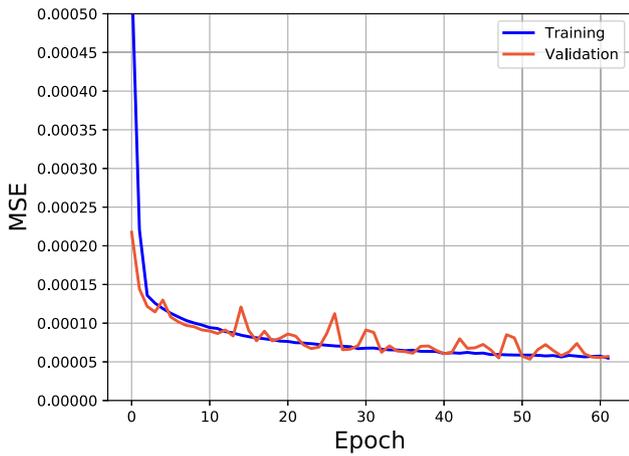
the model parameters of minimal and maximal radii 0.1 and 0.5, respectively, and center axis between 0 and 1, we let the midpoint radius  $r(1/2)$  be a parameter varying between 0.1 and 0.5. For each midpoint radius  $r(1/2)$ , we let  $r(0) = r(1) = 0.1$  and construct a circular arc connecting the points  $(0, 0.1)$ ,  $(1/2, r(1/2))$ ,  $(1, 0.1)$ . In this way, by mirroring the arc, a symmetric convex cylinder is constructed.

As justified by Lemma 2.1, we may compute approximations of the averaged response  $\Psi$ , using piecewise linear interpolations of the cylinder boundary in local charts. We do this with 19 uniform grid points on each arc, as well as the downsampled 5-point uniform grid arcs. The squared error between 19- and 5-point numerically computed values of

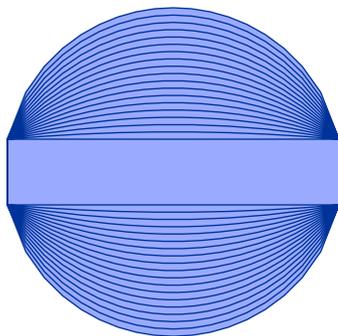
$\Psi$ , and the variation of the error over the parameter interval  $[0.1, 0.5]$  is shown in Fig. 12 with label FEM. The frequency range is again the interval  $(\lambda_{\min}, \lambda_{\max}) = (0, 60)$ .

We compute the predictions in  $\Psi$  of the models trained with 200,000 points from the set Random 5, as evaluated in Fig. 8. In order to do so, we let the radius parameter  $r(1/2)$  vary on a uniform grid of 100 points, and down sample the cylinder radius to 5 points on each arc. The mean squared error computed against the 19-point arc sets is shown in Fig. 12 with label ML.

Numerically, the mean squared error in  $\Psi$  between 19-point and the downsampled 5-points cylinder is truncated to  $7.55 \cdot 10^{-5}$ , while the mean error of the predictions is truncated to  $6.10 \cdot 10^{-5}$ . For comparison, we recall that the best



**Fig. 10** The training history for  $\Psi_{ml}$  with the mean squared error for the total training set split into training and validation parts. The history shown is for one trained model using a training set with 200,000 points from the set Random 5

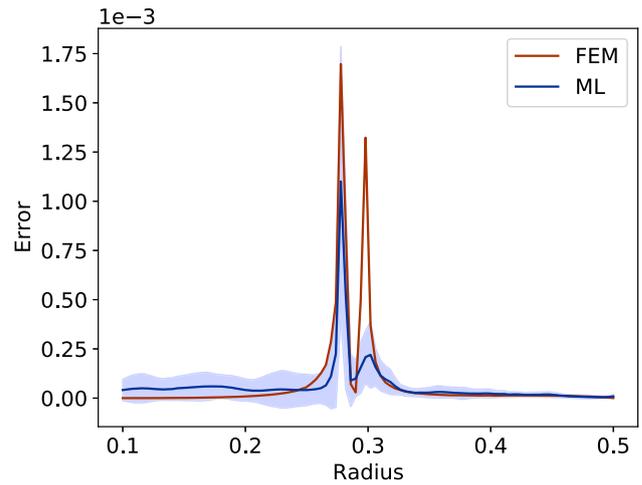


**Fig. 11** The one-parameter out of sample family of convex cylinders

DNN model evaluated to a mean squared error on the set Random 5 truncated to  $5.50 \cdot 10^{-5}$ , according to Table 2. This verifies that indeed the performance of the trained models of the objective function  $\Psi$  on sets of convex cylinders is indicated by the performance on our test sets when the domains are close, as described in Sect. 2.4.

### 6 Comparison with a Linear Model

One can ask why the linear regression would not be perform well in this case. To understand the nature of the nonlinearity in our problem, we look at cylinders with radius  $r(x_1)$  affine in  $x_1$ . For radius  $0.1 \leq r \leq 0.5$ , and  $x_1 \in (0, 1)$ , the set of cylinders may be parametrized by  $r(0)$  and  $r(1)$ . For the interval  $(\lambda_{min}, \lambda_{max}) = (0, 60)$ , we compute a numerical approximation  $\Psi_h$  of the objective function  $\Psi$ . In Fig. 13a,  $\Psi_h$  is shown. Of course  $\Psi$  is linear for uniform cylinders, as it is constant. On the diagonal  $r(0) = r(1)$  in the Fig. 13a, we see the value of this constant. Off the diagonal, we see that  $\Psi_h$  is clearly not the graph of a linear function. A more



**Fig. 12** The squared error between the 19- and 5-point FEM for radius  $r(1/2)$ , and the mean squared error (shaded area is standard deviation) in the predictions of the trained models ML. The models used here are the 10 obtained by training on sets with 200,000 points from the set Random 5. The two peaks come from that the 19-point cylinder and the 5-point cylinder, both hit the spectrum for exactly one value of  $r(1/2)$  in the interval  $[0.1, 0.5]$ , nonorthogonal to the data in  $L^2$

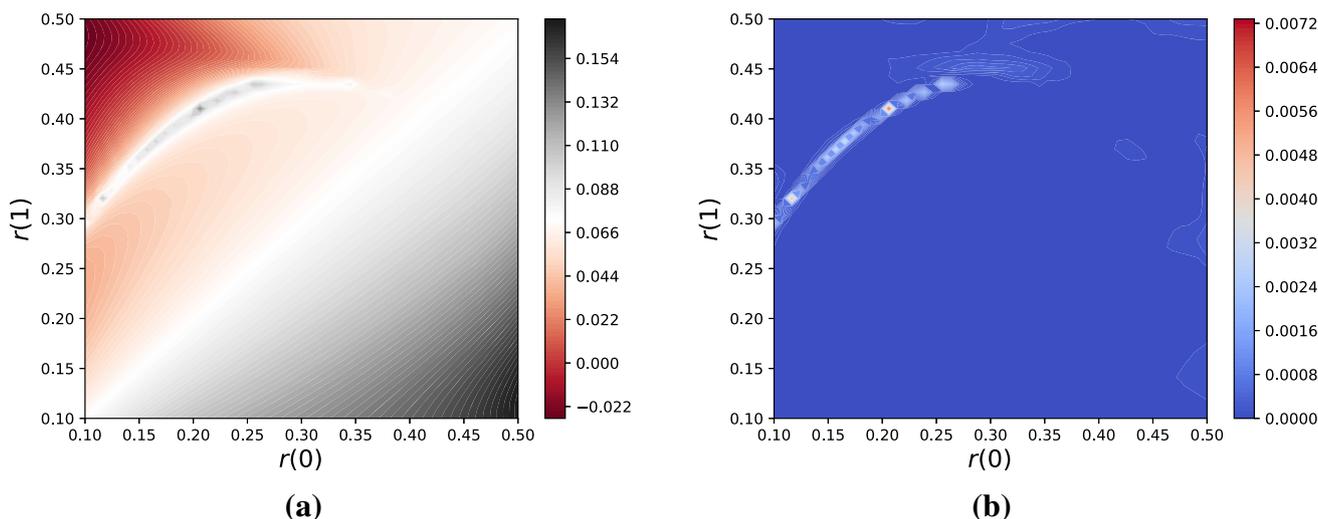
**Table 2** The performance on unseen data of the converged linear and nonlinear models

Model	TS = Training Set		Mean squared error	
	TS	#(TS)	Uniform	Random 5
Linear	Random 5	200,000	2.56e-3	1.91e-3
DNN	Random 5	200,000	2.31e-5	5.50e-5

The mean squared error on unseen data. For the dense neural network model (DNN), the performance is for the best of the sampled models in the sense of minimum mean squared error. The floating point numbers are truncated

careful inspection shows that  $\Psi_h$  is smooth and seems to be linear everywhere except in the upper left corner of the figure, where it shows rapid growth in a narrow region. We know that  $\Psi$  and  $\Psi_h$  are defined for intervals  $(0, \lambda_{max})$  for almost every  $\lambda_{max} > 0$ , but not for all. Namely,  $\Psi_h$  might show singular behavior in the vicinity of a set of positive one-dimensional measure, where  $p_\lambda$  is singular. The approximation  $\Psi_{ml}$  we get with the ML algorithm gives largest error exactly in this singularity region, as seen in Fig. 13b. This verifies the need for a nonlinear activation function in our problem even if restricted to the cylinders with affine radii as functions of  $x_1$ .

In Tables 2 and 3, the errors of the approximations on the Uniform and Random 5 test sets of unseen data are provided. The floating point numbers have been truncated. For comparison, we include both a linear model and the proposed nonlinear model DNN. We train the models for polygonal cylinders with five random points. One can see that the proposed DNN model performs much better than the linear one, both on uniform and non-uniform cylinders.



**Fig. 13** The value of  $\Psi_h$  for cylinders with affine radii and  $\lambda_{\max} = 60$  (a). The squared error between  $\Psi_{ml}$  and  $\Psi_h$  (b)

**Table 3** The performance on unseen data of the converged linear and nonlinear models

Model	TS = Training Set		% Abs Err < 0.01	
	TS	#(TS)	Uniform	Random 5
Linear	Random 5	200,000	46.9	22.1
DNN	Random 5	200,000	98.3	95.7

The percentage of unseen data with an absolute error  $|\Psi_{ml} - \Psi_h|$  less than 0.01. For the dense neural network model (DNN), the performance is for the best of the sampled models in the sense of minimum mean squared error. The floating point numbers are truncated

## 7 Conclusions

We have proposed a feedforward dense neural network for predicting the average sound pressure response over a frequency range. We have shown for polygonal cylinders that the obtained results are sufficiently accurate in that they reach the estimated accuracy of the numerical data. Although the amount of data needed in order to reach the desired accuracy could be considered as big, it is expected that the results would serve as a point of reference for more advanced machine learning models. The performance of the feedforward dense neural network has been evaluated. The dependence of the percentage accurately predicted samples and the mean squared error on the training set size is presented.

**Acknowledgements** The computations were partially performed on resources at Chalmers Centre for Computational Science and Engineering (C3SE) provided by the Swedish National Infrastructure for Computing (SNIC). We thank the referees for careful reading of the paper.

**Funding** Open access funding provided by Chalmers University of Technology.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## Appendix A: A Hyperparameter Grid

To supplement Sect. 4.2 we here provide a grid around the model and training parameter values specified. In particular, we choose the parameters (i) number of hidden layers, (ii) number of nodes in each layer, and (iii) the fraction of the training data used for the validation split. For each 3-tuple of parameters, we present in Table 4 the mean squared error on unseen data from the data sets Uniform and Random 5. The data sets used for training was 200,000 points from the Random 5 training set.

**Table 4** The mean squared errors on the unseen data Uniform and Random 5 for various values of the parameters

Hyperparameter			Mean squared error	
Hidden layers	Layer size	Validation split %	Uniform	Random 5
2	64	10	3.98e−5	1.17e−4
2	64	20	4.64e−5	1.53e−4
2	64	30	1.37e−4	1.44e−4
2	128	10	4.80e−5	8.45e−5
2	128	20	5.00e−5	8.74e−5
2	128	30	4.65e−5	9.61e−5
2	192	10	1.21e−5	7.38e−5
2	192	20	3.15e−5	7.98e−5
2	192	30	9.42e−6	8.47e−5
3	64	10	5.21e−5	8.25e−5
3	64	20	1.79e−5	7.91e−5
3	64	30	4.44e−5	9.35e−5
3	128	10	1.78e−5	6.01e−5
3	128	20	1.29e−5	5.75e−5
3	128	30	9.45e−5	5.75e−5
3	192	10	1.75e−5	5.09e−5
3	192	20	1.07e−5	5.20e−5
3	192	30	7.97e−6	5.55e−5
4	64	10	2.02e−5	7.40e−5
4	64	20	1.86e−5	6.26e−5
4	64	30	2.47e−5	8.25e−5
4	128	10	1.68e−5	5.34e−5
4	128	20	3.00e−5	5.60e−5
4	128	30	3.31e−5	5.46e−5
4	192	10	1.04e−5	4.91e−5
4	192	20	1.40e−5	5.39e−5
4	192	30	1.80e−5	5.27e−5

The numerical values of the mean squared errors are truncated

## References

1. J. Adler and O. Öktem. Solving ill-posed inverse problems using iterative deep neural networks. *Inverse Problems*, 33(12):124007, 2017
2. Allaire, G.: *Numerical Analysis and Optimization: An Introduction to Mathematical Modelling and Numerical Simulation*. Oxford University Press, Oxford (2007)
3. Abadi, M., et al.: *TensorFlow: large-scale machine learning on heterogeneous systems* (2015). Software available from tensorflow.org
4. Alnæs, M. S., et al.: The fenics project version 15. *Arch. Numer. Softw.*, 3(100), 9-23 2015
5. Baymani, M., Effati, S., Kerayechian, A.: A feed-forward neural network for solving stokes problem. *Acta Appl. Math.* **116**(1), 55 (2011)
6. Baymani, M., Effati, S., Niazmand, H., Kerayechian, A.: Artificial neural network method for solving the navier-stokes equations. *Neural Comput. Appl.* **26**(4), 765–773 (2015)
7. J. Berg and K. Nyström. A unified deep artificial neural network approach to partial differential equations in complex geometries. *Neurocomputing*, 317:28–41, 2018
8. Bianco, M.J., Gerstoft, P., Traer, J., Ozanich, E., Roch, M.A., Gannot, S., Deledalle, C.-A.: Machine learning in acoustics: Theory and applications. *The Journal of the Acoustical Society of America* **146**(5), 3590–3628 (2019)
9. BSEN ISO. 7235: 2003.: *Acoustics—Laboratory measurement procedures for ducted silencers and air-terminal units—insertion loss, flow noise and total pressure loss* (2003)
10. Choi, K. K., Kim, N.-H.: *Structural Sensitivity Analysis and Optimization I: Linear Systems*. Springer, Berlin (2006)
11. Goodfellow, I., Bengio, Y., Courville, A.: *Deep learning (adaptive computation and machine learning series)* (2016)
12. Grisvard, P.: *Elliptic Problems in Nonsmooth Domains*, volume 69. SIAM, Philadelphia (1985)
13. F. Hecht. New development in freefem++. *J. Numer. Math.*, 20(3–4):251–265, 2012
14. Hernandez, V., Roman, J.E., Vidal, V.: Slepc: A scalable and flexible toolkit for the solution of eigenvalue problems. *ACM Transactions on Mathematical Software (TOMS)* **31**(3), 351–362 (2005)
15. K. Hornik, M. Stinchcombe, and H. White. Universal approximation of an unknown mapping and its derivatives using multilayer feedforward networks. *Neural networks*, 3(5):551–560, 1990

16. Kingma, D.P., Ba, J.: Adam: a method for stochastic optimization. arXiv preprint [arXiv:1412.6980](https://arxiv.org/abs/1412.6980) (2014)
17. Komkov, V., Choi, K. K., Haug, E. J.: Design Sensitivity Analysis of Structural Systems, volume 177. Academic Press, New York, 1986
18. Lagaris, I.E., Likas, A., Fotiadis, D.I.: Artificial neural networks for solving ordinary and partial differential equations. *IEEE Trans. Neural Networks* **9**(5), 987–1000 (1998)
19. Montavon, G., Orr, G., Müller, K.-R.: *Neural Networks: Tricks of the Trade*, volume 7700. Springer, Berlin, 2012
20. Necas, J.: *Les méthodes directes en théorie des équations elliptiques* (1967)
21. S. Ossandón and C. Reyes. On the neural network calculation of the lamé coefficients through eigenvalues of the elasticity operator. *Comptes Rendus Mécanique*, 344(2), 113–118, 2016
22. S. Ossandón, C. Reyes, P. Cumsille, and C. M. Reyes. Neural network approach for the calculation of potential coefficients in quantum mechanics. *Computer Physics Communications*, 214:31–38, 2017
23. Ossandón, S., Reyes, C., Reyes, C.M.: Neural network solution for an inverse problem associated with the dirichlet eigenvalues of the anisotropic laplace operator. *Computers & Mathematics with Applications* **72**(4), 1153–1163 (2016)
24. Pettersson, K., Karzhou, A., Pettersson, I.: Frre1. <https://github.com/klaspettersson/FrRe1> (2020)
25. Raissi, M., Perdikaris, P., Karniadakis, G.E.: Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *J. Comput. Phys.* **378**, 686–707 (2019)
26. Reddi, S., Zaheer, M., Kale, S., Sachan, D., Kumar, S.: Adaptive methods for nonconvex optimization. *NIPS* (2018)
27. Rizzuti, G., Siahkoohi, A., Herrmann, F.J.: Learned iterative solvers for the Helmholtz equation. In: 81st EAGE Conference and Exhibition 2019, volume 2019(1), pp. 1–5. European Association of Geoscientists & Engineers (2019)
28. Schmidhuber, J.: Deep learning in neural networks: An overview. *Neural Netw.* **61**, 85–117 (2015)
29. Sirignano, J., Spiliopoulos, K.: Dgm: a deep learning algorithm for solving partial differential equations. *J. Comput. Phys.* **375**, 1339–1364 (2018)
30. Skudrzyk, E.: *The Foundations of Acoustics: Basic Mathematics and Basic Acoustics*. Springer, Berlin, (2012)
31. Strang, G.: *Linear Algebra and Learning from Data*. Wellesley-Cambridge Press, Cambridge (2019)
32. Tucker, W.: *Validated Numerics: A Short Introduction to Rigorous Computations*. Princeton University Press, Princeton (2011)

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.