

Physics-informed neural network for ultrasound nondestructive quantification of surface breaking cracks

Khemraj Shukla · Patricio Clark Di Leoni ·
James Blackshire · Daniel Sparkman ·
George Em Karniadakis (*)

Received: date / Accepted: date

Abstract We introduce an optimized physics-informed neural network (PINN) trained to solve the problem of identifying and characterizing a surface breaking crack in a metal plate. PINNs are neural networks that can combine data and physics in the learning process by adding the residuals of a system of partial differential equations to the loss function. Our PINNs is supervised with realistic ultrasonic surface acoustic wave data acquired at a frequency of 5 MHz. The ultrasonic surface wave data is represented as a deformation on the top surface of a metal plate, measured by using the method of laser vibrometry. The PINN is physically informed by the acoustic wave equation and its convergence is sped up using adaptive activation functions. The adaptive activation function uses a trainable hyperparameter, which is optimized to achieve the best performance of the network. The adaptive activation function changes dynamically, involved in the optimization process. The usage of the adaptive activation function significantly improves the convergence, evidently observed in the current study. We use PINNs to estimate the speed of sound of the metal plate, which we do with an error of 1%, and then, by allowing the speed of sound to be space dependent, we identify and characterize the crack as the positions where the speed of sound has decreased. Our study also shows the effect of sub-sampling of the data on the sensitivity of sound speed estimates. More broadly, the resulting model shows a promising deep neural network model for ill-posed inverse problems.

Keywords PINN · Activation function · Adaptive · Wavefield · Nondestructive

K. Shukla
Division of Applied Mathematics
Brown University
Providence, RI 02906
E-mail: khemraj_shukla@example.com

P. Leoni
Department of Mechanical Engineering
John Hopkins University
Baltimore, MD 2121
E-mail: pato@jhu.edu

J. Blackshire
Air Force Research Laboratory
OH 45433
E-mail: james.blackshire@us.af.mil

D. Sparkman
Air Force Research Laboratory
OH 45433
E-mail: daniel.sparkman.1@us.af.mil

G. Karniadakis
Division of Applied Mathematics
Brown University
Providence, RI 02906
E-mail: george.karniadakis@brown.edu

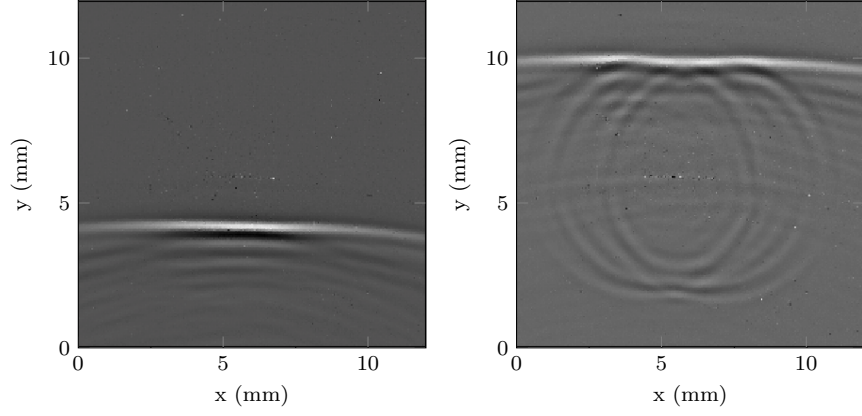
(*) Corresponding author

1 Introduction

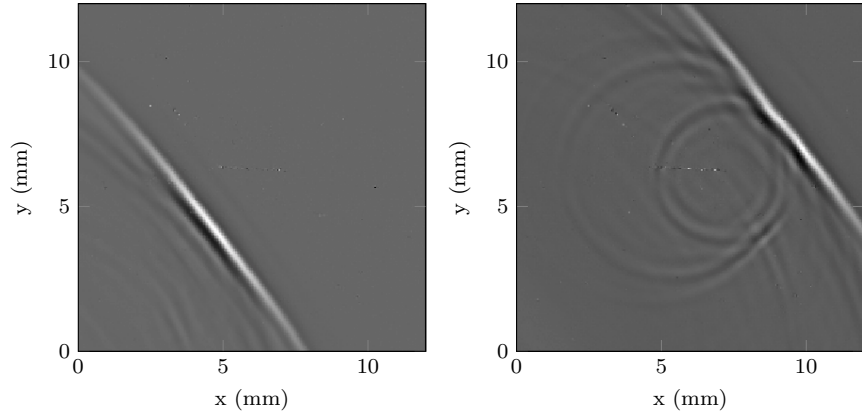
The recent advances in the machine learning algorithms along with the growth in data and computing resources [23] have transformative results across various scientific disciplines including but not limited to image recognition [12], cognitive sciences [13], genomics [2] and bioinformatics [14]. Unlike these sciences, cost of acquisition of data becomes prohibitive while analyzing the physical and biological systems. The prohibitive nature of the data results in failures of a vast majority of state-of-the-art machine-learning algorithms. The failures are typically due to an unconstrained search of non-linear mapping between high dimensional (a large ratio between number of features per sample) input-output pairs of the data, resulting in a very large space of admissible solutions. To circumvent the effect of the data sparsity, Raissi et al. [19] proposed a physics informed neural network (PINN), which uses the physical laws governing the dynamics of a time dependent system as a regularization agent resulting in a reduction of the space of admissible solutions to a manageable size. Since the advent of the PINN, applicability of PINN is being explored for various real world problems and based on this premise, our work in this paper presents applicability of PINN to solve the problems of wavefield imaging at ultrasonic frequencies. Wavefield imaging methods have acquired recent popularity and becoming a standard tool for ultrasonic non-destructive evaluation research [6, 16] gradually. In practice, wavefield imaging methods use a scanning laser vibrometry system to detect motions on a material surface generated by a stationary ultrasound excitation source. Repetitions of the ultrasonic excitation for different vibrometry detection points (x, y) enables the collection of time series data (displacement amplitude vs time signals) at discrete (x, y) spatial locations. This results in a set of measurement data captured at regularized grid points. These measurements are arranged as a collection of time-evolving snapshots of ultrasound wavefield data showing important details of the ultrasound-crack interaction i.e., crack scattering and back-scattering events. Researchers have successfully used wavefield imaging methods to quantitatively study ultrasonic wave dispersion (variation of phase velocity with frequency), attenuation (due to material property heterogeneity), damage scattering events, and time history details of elastic waves propagating in complex material systems [8, 5]. Wavefield imaging methods have, for example, been successfully used to study the crack morphology details, e.g. direct visualization of ultrasonic wave interactions with realistic damage features. This results in an improved understanding of the crack morphology and the crack closure effects on the transmitted and reflected ultrasound wave characteristics [4]. In a related effort, frequency-wavenumber analysis was used to discriminate between bulk and surface wave modes scattered from a surface-breaking crack [22]. Once these results are combined with the detailed information regarding the damage features e.g. size, shape and morphology of a crack, the scattering process of ultrasonic waves can be correlated with the locally developed damage features.

Thus far, studies conducted to determine the detailed characteristics of a crack and its morphology using ultrasonic surface acoustic waves have been primarily based on direct visualization of the wavefields or an analysis of its time series. In the present study, we have used the wavefield data to completely characterize the incipient damage in an aerospace material using a PINN. The detection of a crack is based on the premise that the presence of a crack will result in attenuation of the wave energy, thus reducing the local sound speed. In this study, the sound speed, a function of space $v(x, y)$, is learned from the PINN. We performed computational experiments on three sets of wavefield data acquired at different incidence angles (angle between incident wavefield and axis of the crack, measured from normal) of 0° , 45° , and 90° . The speed of sound recovered from these data sets represents the location and extent of the fracture with a very good accuracy.

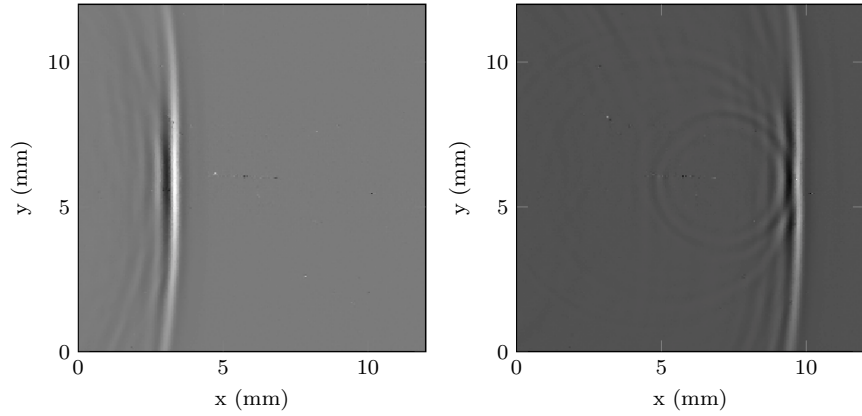
The objective of this study is two fold: First, to compute the speed of the wave in the material with ultrasonic wavefield data, which will indirectly represent the the effect of heterogeneities and properties of the material in terms of bulk modulus and density. Second, to determine systematically the minimum data size required for computing the sound speed in the material accurately. The existing methodologies [4, 22] mandate the acquisition of the wavefield data by following the Nyquist theorem [7] to avoid the temporal and spatial aliasing of the ultasonic wavefield. This results into a very dense grid points in time and space domain and thus prohibiting the acquisition of the data in real applications due to the cost and logistics.



(a) $t = 11.58 \mu s$, at $= 0^\circ$ without the crack. (b) $t = 13.58 \mu s$, at $= 0^\circ$ with the crack.



(c) $t = 9.98 \mu s$, at $= 45^\circ$ with out the crack. (d) $t = 11.98 \mu s$, at 45° with the crack.



(e) $t = 8.78 \mu s$, at $= 90^\circ$ without the crack. (f) $t = 10.98 \mu s$, at $= 90^\circ$ without the crack.

Fig. 1: Representative snapshots of particle displacement in metal plate, where (a) and (b) corresponds to 0° incidence angle and showing the displacement with and without the effects of the crack on the wavefield, respectively. Similar representations for incidence angles of 45° and 90° are shown by pairs of [(c), (d)], and [(e), (f)], respectively. A detailed description of the data is presented by Blackshire [5].

2 Ultrasonic surface acoustic wave data

Wavefield imaging methods involve a quantitative measurement of out-of-plane particle displacements on the surface of a material caused by elastic wave motions. A typical wavefield imaging system utilizes a contact transducer to introduce transient ultrasonic waves on a material substrate. These motions are detected by a focused laser vibrometry beam at various spatial positions (x, y) along the material's surface, where optical interferometric principles are used to relate phase and amplitude changes in the detected laser light to surface displacement motions [5]. A detailed study discussing the wavefield imaging method for acquiring ultrasonic surface acoustic wavefield data is provided by Blackshire [5]. The data used in this study was acquired at three incidence angles of 0° , 45° and 90° , and it includes 2D time snapshots of the surface acoustic waves representing the particle displacements at (x, y) regularized spatial grid locations for out-of-plane motions in the z -direction.

The data was acquired using a National Institute of Standards and Technology (NIST) surface-breaking crack reference standard (RM 8458) in a $7\text{ cm} \times 7\text{ cm} \times 2\text{ cm}$ block of 7075-T651 aluminum alloy substrate material, with wavefield imaging measurements taken over a $12\text{ mm} \times 12\text{ mm}$ region centered over the crack with a scanning step resolution of 50 microns (240×240 $x - y$ scan points), and a time-step data acquisition interval of $0.02\text{ }\mu\text{s}$ (1024 total time-step points or $20.48\text{ }\mu\text{s}$). A wave packet of 5 MHz was propagated as initial conditions and data were arranged in a dimension $[N_x, N_y, N_t]$ of $240 \times 240 \times 1024$, with N_x, N_y being the grid points in x and y directions and N_t is number of time samples. These datasets were acquired for three sets of incidence angles 0° , 45° and 90° . The angle of incidence is the angle between the wave vector and the axis of the fracture and measured from the normal. Examples of snapshots of wavefield representing the particle displacement are shown in Figure 1. Figures 1a and 1b represent the snapshots of wavefield at $11.58\text{ }\mu\text{s}$ and $13.58\text{ }\mu\text{s}$, before and after the interaction of the wavefield with the crack at 0° of incidence angle, respectively. Similarly, Figures 1c and 1d represent the snapshots of the wavefield at 45° (at $9.98\text{ }\mu\text{s}$) with and without the effects of crack on the wavefield, respectively. Figures 1e and 1f represent the snapshots for incidence angle of 90° . It is quite evident that the back-scattering phenomena are more prominent for angles 0° and 45° than those of recorded at angle of 90° .

3 Problem setup and Physics Informed Neural Network (PINN)

To characterize the crack in terms of location and extent, we postulate that the speed, $v(x, y)$, of the wave will be the key feature. As ultrasonic acoustic waves propagate in a medium with crack(s), the speed of wave decreases due to inelastic attenuation and near-field scattering of the wave energy [7]. Thus, a local change in the wave speed from the background will show the evidence of the presence of crack. In this study, we consider the linear second-order partial differential equation governing the propagation of acoustic wave equations, which is expressed as

$$u_{tt} = v^2(x, y)\Delta u, \quad x \in \Omega, \text{ and } t \in [0, T], \quad (1)$$

where $u(t, x, y)$ is the solution, $v(x, y)$ is the sound speed and $\Omega \in \mathbb{R}^2$. The subscript t denotes the partial differentiation in time domain. Here, we are given measurements u and want to learn $v(x, y)$ that best describes equation (1). These problems of learning are known as system identification or data driven discovery of partial differential of equations [18, 21]. Neural networks (NNs) are successfully used to obtain the approximate solution of partial differential equations (PDEs). One can also construct the physics-informed machine learning using systematically structured prior information about the solutions. The work in [17, 18, 20] have successfully demonstrated the use of PDEs as prior information to constrain the minimization process in the context of system identification (forward problem) and solution inferences (inverse problem). In particular, PINN can solve the forward and inverse problem accurately. In forward problems, the approximate solutions of the PDEs are obtained, whereas the inverse problem computes the parameters or even unknown functions involved in the PDEs. The problem of characterization of a crack falls into the second category i.e. computation of $v(x, y)$ with $u(t, x, y)$ given by the real data. The loss function constructed in the PINN algorithm incorporates the residual term from the governing equation(s), which act as a regularization term and thus constrains the space of admissible solutions.

3.1 Neural network

We consider a neural network (NN) of depth D corresponding to a network with an input layer, $D - 1$ hidden layers and an output layer. In the k^{th} hidden layer, N_k number of neurons are present. Each hidden layer of the network receives an output $x^{k-1} \in \mathbb{R}^{N_{k-1}}$ from the previous layer where an affine transformation is performed. The transformation is expressed as

$$\mathcal{L}_k(x^{k-1}) := w^k x^{k-1} + b^k. \quad (2)$$

The network weights w^k and bias term b^k for k^{th} are initially chosen from *independent and identically distributed (iid)* samplings. The nonlinear activation function $\sigma(\cdot)$ is applied to $\mathcal{L}_k(x^{k-1})$ component wise prior to sending it as input to next layer. The activation function at the output layer is an identity function. Thus, the final neural network representation is expressed as

$$u_\Theta(x) = (\mathcal{L}_k \circ \sigma \circ \mathcal{L}_{k-1} \circ \dots \circ \sigma \circ \mathcal{L}_1)(x), \quad (3)$$

where the operator \circ is the composition operator, $\Theta = \{w^k, b^k\}_{k=1}^D$ represents the trainable parameters in the network, u is the output and $x := x^0$ is the input to the neural network.

3.2 PINN for ultrasonic surface acoustic wave equation

We define the residual $f(t, x)$ by (1) i.e.,

$$f := u_{tt} - v^2(x, y)\Delta u, \quad (4)$$

and proceed with approximating $u(t, x)$ with a deep neural network as expressed in equation (3). This assumption along with equation (4) results in the *physics informed neural network* $f(t, x)$. This network is derived by applying the chain rule for differentiating composition of functions using automatic differentiation [18, 3] and has the same parameters representing the $u(t, x)$ but with different activation function due to the Δ operator in (1). The shared parameters between the neural networks $u(t, x)$ and $f(t, x)$ can be learned by minimizing the mean squared loss error expressed as

$$MSE = \lambda MSE_u + MSE_f, \quad (5)$$

where $MSE_u = \frac{1}{N_u} \sum_{i=1}^{N_u} |u(t_u^i, x_u^i, y_u^i) - u^i|^2$, and $MSE_f = \frac{1}{N_f} \sum_{i=1}^{N_f} |f(t_f^i, x_f^i, y_f^i)|^2$. The parameter $\lambda > 0$ is a penalty parameter, which helps in achieving the fast convergence. Here, $\{t_u^i, x_u^i, y_u^i\}$ denote the initial and boundary training data on $u(t, x)$ and $\{t_f^i, x_f^i\}$ specify the residual points for $f(t, x)$. The loss MSE_u here corresponds to the data and MSE_f enforces the structure imposed by (4).

In the present study, the characterization of the crack is provided by spatially varying speeds of the ultrasonic acoustic waves i.e $v(x, y)$. Thus, to estimate the speed, we used another neural network to learn the $v(x, y)$ and subsequently fed to PINN. A block diagram representing the PINN model used in this study is shown in Figure 2a.

3.3 Loss function and optimization algorithm

Our aim is to find the optimal weight for which the loss function defined in (5) is minimized. Thus, the definition of the resulting optimization problem is expressed as

$$w^* = \arg \min_{w \in \Theta} (J(w)); \quad b^* = \arg \min_{b \in \Theta} (J(b)) \quad v^* = \arg \min_{v \in \Theta} (J(v)). \quad (6)$$

One can approximate the solutions to optimization problems, defined in (6), using an iterative method by using one of the forms of gradient descent algorithm. The stochastic gradient descent (SGD) algorithm is ubiquitously used by machine learning community [21]. In SGD method, the weights are updated as

$$w^{m+1} = w^m - \eta_l \nabla_w J^m(w), \quad (7)$$

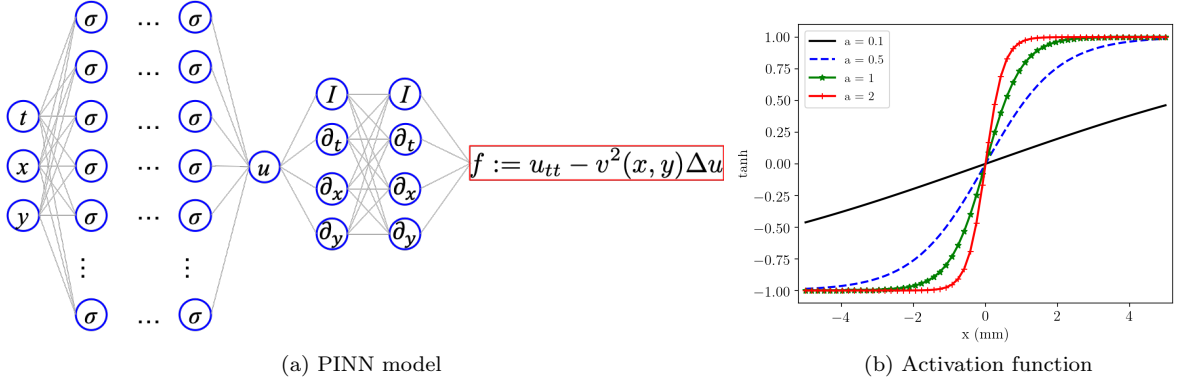


Fig. 2: (a) represents physics informed neural network model informed by the acoustic wave equation and (b) shows comparison of activation function \tanh at various values of a .

where $\eta_l > 0$ is the learning rate and J^m is the loss function at m^{th} iteration. SGD methods could be initialized with some starting value w^0 . In this work, the ADAM optimizer [11], a variant of SGD method, is used unless mentioned otherwise.

3.4 Adaptive activation function

The dependence of the derivative of the loss function on the optimization parameters defines a very important role of activation function. The regularity of the derivative of the loss function is dependent on the activation function. Various activation functions are used in PINN algorithm such as \tanh , \sin etc to solve the PDEs [19,20]. There is no criterion to choose these function as it solely depends on the problem in hand. In the present work, while performing the training on ultrasonic acoustic data with the \tanh , a very slow and non-convergence of loss function was observed. To speed up the convergence we adopted the method of the adaptive activation function [10]. In the adaptive activation method, Jagtap et al. [10] introduced the hyper-parameter $a > 0$ in the activation function as $\sigma(a\mathcal{L}_k(x^{k-1}))$, where a is subsequently learned by minimizing the loss function along with the weights and biases. Thus, the definition of optimization problem for a is expressed as

$$a^* = \arg \min_{a \in \mathbb{R}^+ \setminus \{0\}} (J(a)).$$

The parameter a is updated as

$$a^{m+1} = a^m - \eta_l \nabla_a J^m(a).$$

To accommodate the effect of learning factor [9], responsible for convergence to global minima, the hyper-parameter a is multiplied by a scale factor $n \geq 1$ and the final activation function is recovered as

$$\sigma(na\mathcal{L}_k(x^{k-1})).$$

The effect of a on the activation function defined by hyperbolic tangent (\tanh) is shown in Figure 2b. It is to be noted from Figure 2b that introducing a steepens the activation function, which eventually helps in achieving the fast convergence. The effect parameter a on achieving the fast convergence is shown in detail by Jagtap et al. [10] for problems dealing with non-linear PDEs such as the Burger's equation, Klein-Gordon equation, as well as other standard machine learning benchmarks.

3.5 Data preconditioning

Wavefield data are acquired through an experimental set up and prone to random and environmental noise. To filter out these random noises, we have used the method of principal component analysis (PCA). PCA is an unsupervised method for extracting the variance structure from a high dimensional data and

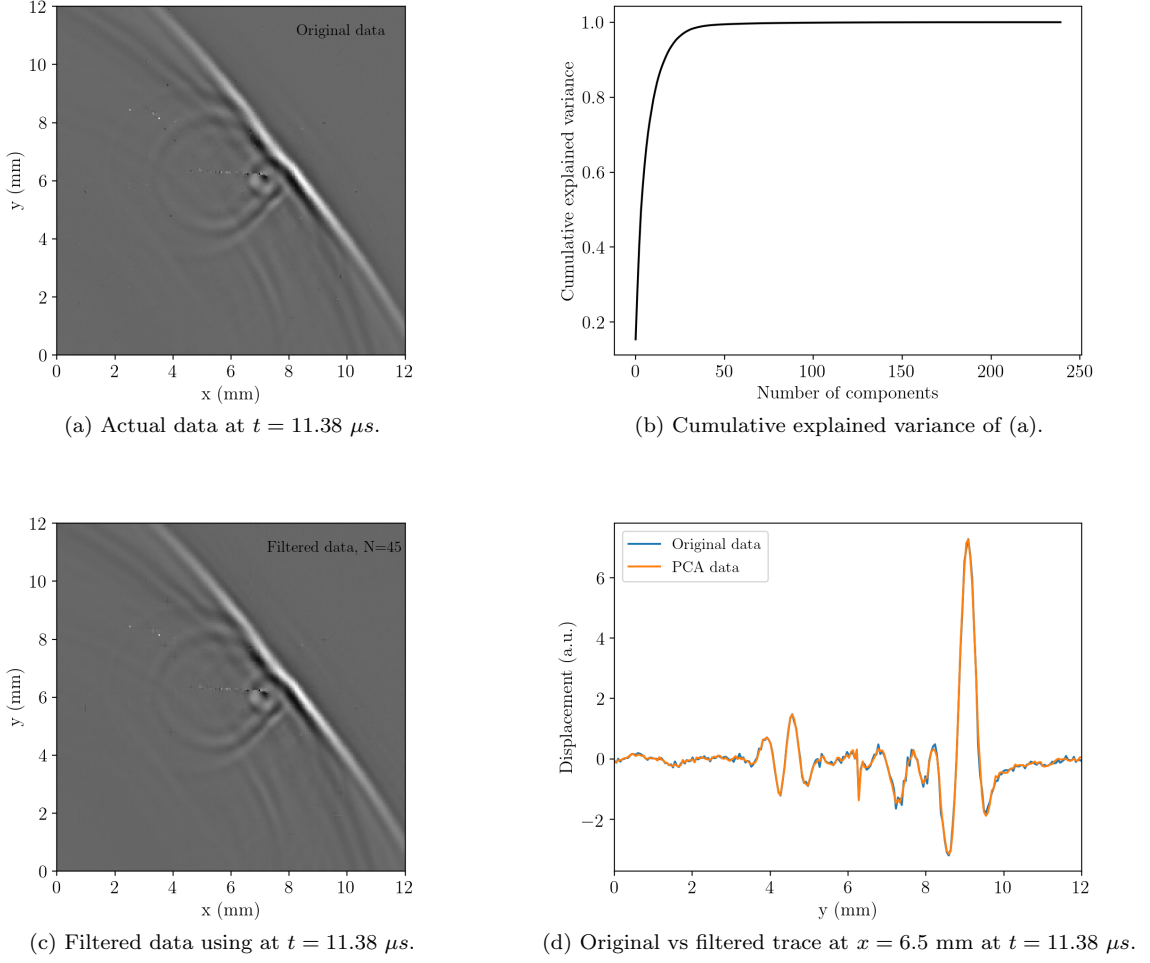
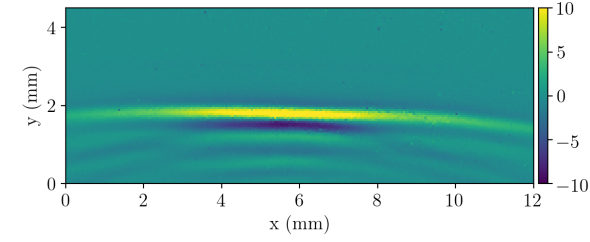
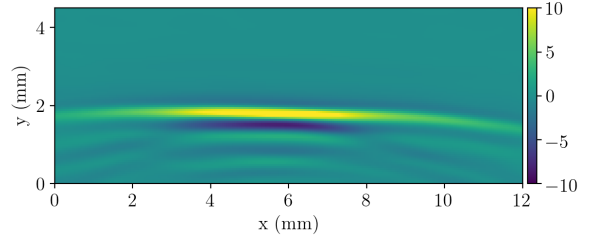


Fig. 3: Filtering of wavefield data acquired at 45° , using the principle component analysis (PCA). PCA requires only the first 45 components to construct the original data by zeroing out the insignificant components.

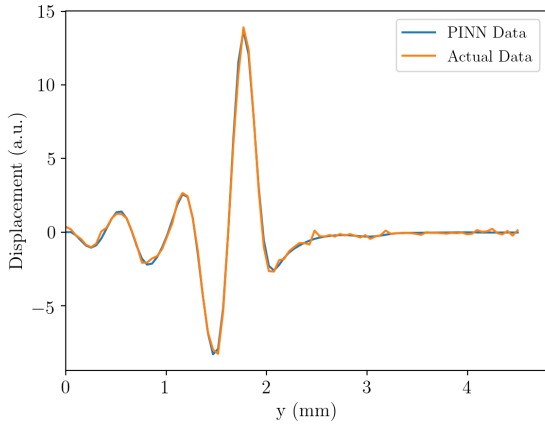
project the data into a subspace such that variance of projected data is maximized. Application of PCA to the data provides two types of information: principal components and explained variance ratio, which is the ratio between the variance of that principal component and the sum of variances of all individual principal components. Thus, the filtering process involves the zeroing out one or more smallest principle component by preserving the maximum data variance. We used the PCA module of sklearn package [15], and a snippet of routine used to filter out the data is given in A. Results obtained after application of PCA on data set for incidence angle 45° are shown in Figure 3. Figure 3a represents a snapshot of the raw data at $11.38 \mu s$ for 45° data set. Figure 3b shows a plot of cumulative explained variance and principle components. This quantifies how much of the total, 240-dimensional variance is contained within the first N components. Figure 3b clearly shows that the first 45 components contain approximately 95% of the variance. Figure 3c represent the filtered data obtained after discarding the components close to zero. Figure 3d represents a comparison between raw and filtered trace extracted from Figures 3a and 3c at $x = 7 \text{ mm}$, respectively. Figure 3d clearly shows that the filtered data is smoother in comparison to the raw data. Results for datasets acquired at incidence angle of 45° and 90° are shown in appendix B as Figure B1 and Figure B2, respectively. To construct the filtered data for 0° case, it requires first 15 principle components, less than those for 45° (45 components) and 90° (20 components) cases. This is due to the fact that back scattering effects are more prominent for 45° than 0° and 90° .



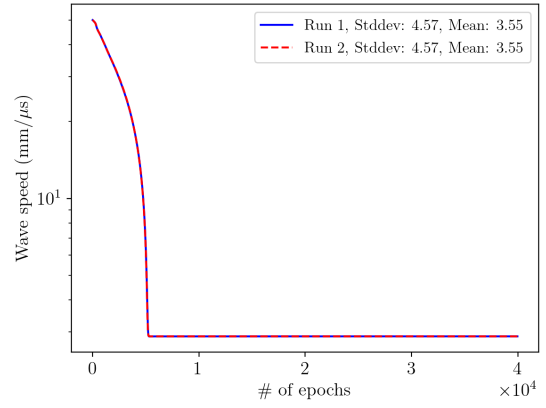
(a) Actual data at $t = 10.8 \mu s$.



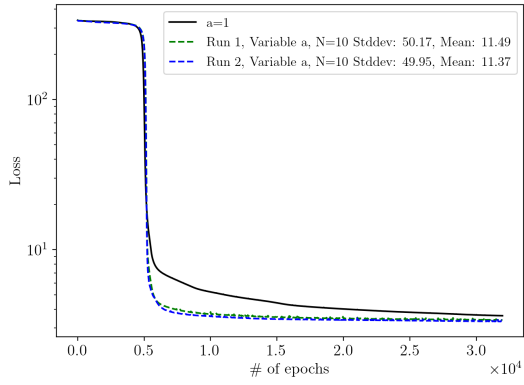
(b) Data recovered from PINN at $t = 10.8 \mu s$.



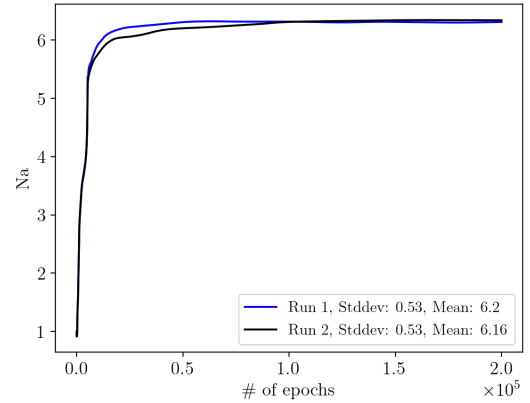
(c) Traces of (a) and (b) at $x = 6 \text{ mm}$ at $t = 10.8 \mu s$.



(d) Wave speed discovered from PINN simulation as a function of epoch.

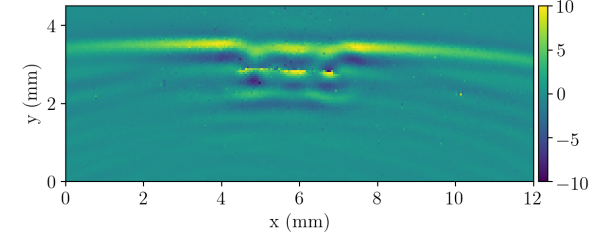


(e) Loss function of neural network model

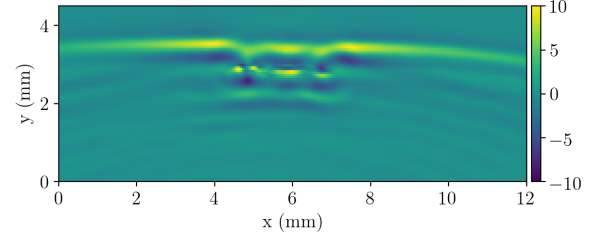


(f) Variation in a with $N = 10$.

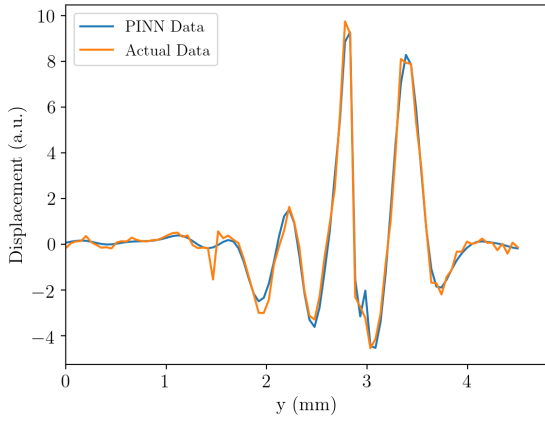
Fig. 4: Results from training of neural network model for data acquired at incidence angle of 0° and devoid of the crack, where (a) and (b) represent the snapshot of the particle displacement obtained from actual and PINN simulated data, respectively. (c) shows a comparison of traces obtained from actual and PINN simulated data. (d) represents the speed of wave discovered from two different runs of PINN simulation as a function of epoch. (e) shows a comparison of the loss functions for fix and adaptive activation function. The results are shown for two runs of the PINNs with two different initializations and (f) shows variation in a with $N = 10$ for two different initializations.



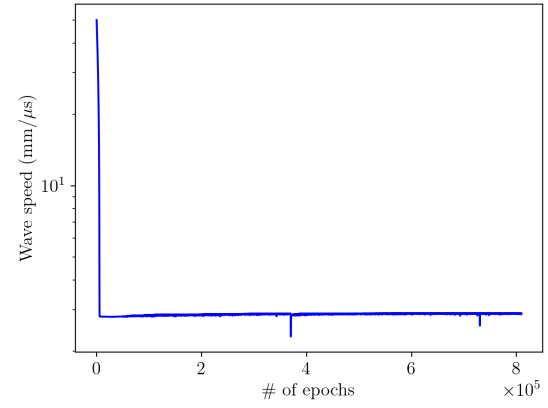
(a) Actual data at $t = 12.38 \mu s$.



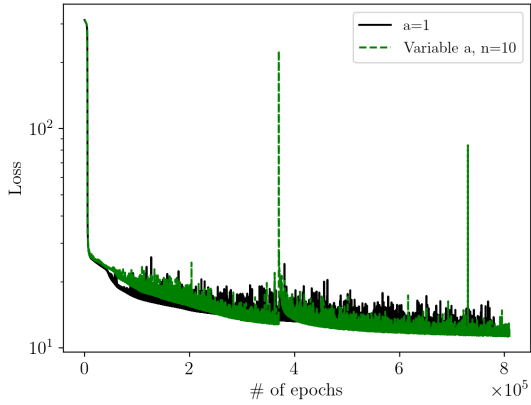
(b) Data recovered from PINN simulation at $t = 12.38 \mu s$.



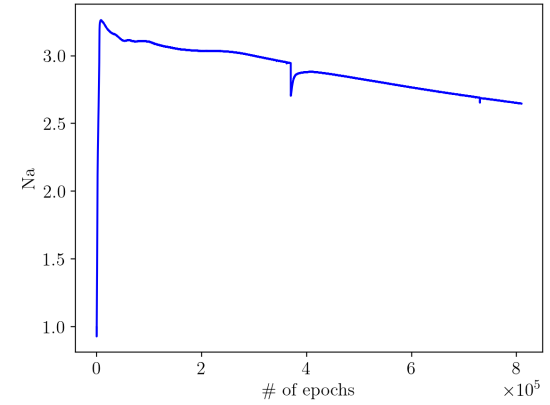
(c) Traces of (a) and (b) at $x = 6 (mm)$ at $t = 12.38 \mu s$.



(d) Wave speed discovered from PINN simulation as a function of epoch.



(e) Loss function of neural network model.



(f) Variation in a with $N = 10$.

Fig. 5: Results from training of neural network model for data acquired at incidence angle of 0° including the crack, where (a) and (b) represents the snapshots of the particle displacement from actual and PINN simulated data, respectively. (c) shows a comparison of traces obtained from actual and PINN simulated data and (d) represents the speed of wave discovered from PINN simulation as a function of epoch (e) shows a comparison of loss function for fix and adaptive activation function and (f) shows variation in a with $N = 10$.

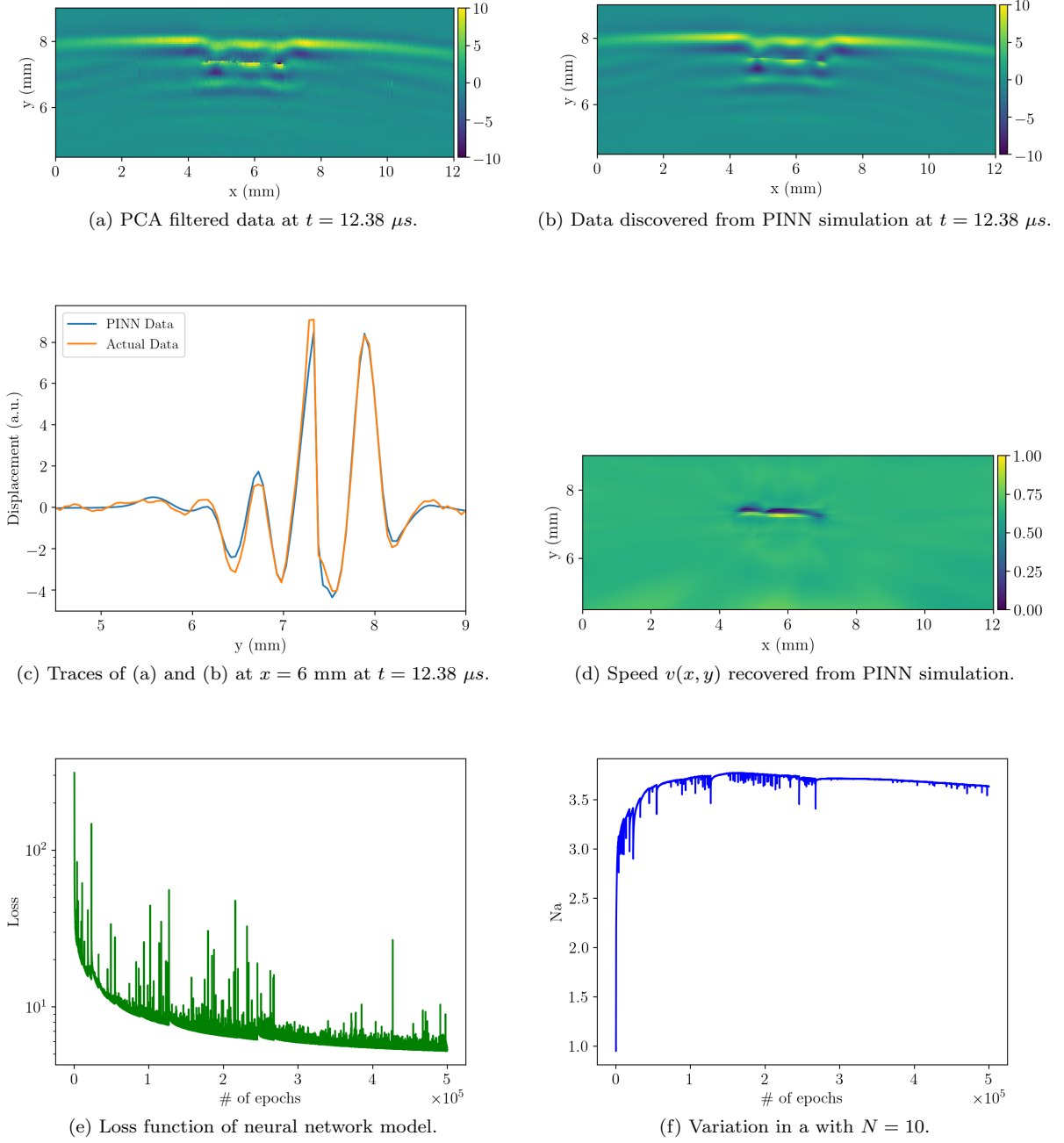


Fig. 6: Results from training of neural network model for data acquired at incidence angle of 0° , where (a) and (b) represents the snapshot of the particle displacement from actual and PINN simulated data, respectively. (c) shows a comparison of traces obtained from actual and PINN simulated data. (d) represents the normalized speed of the wave as a function of space, where crack is reflected as low speed zone. (e) shows the loss function corresponding to neural network model and (f) shows variation in a with $N = 10$.

4 Results

4.1 Wave speed in the homogeneous and heterogeneous parts of the material for 0° angle: Independent of spatial variable (x, y)

To compute the speed of sound in the 7075-T651 aluminum alloy substrate material, we started the training experiments with the wavefield data acquired at 0° angle and trained the network for the region where the wavefield had not encountered the crack. We used 30 out of the 1000 snapshots (we removed those where the wave was not present or where the wave had already hit the crack), and randomly picked about 20 % of the locations at each snapshot for the training procedure. The neural network in this case had 2 layers each with 32 neurons with a fully connected architecture and an initial learning rate of $5e - 4$. Figures 4a and 4b represent snapshots of the wavefield at $10.8 \mu s$, obtained from real and PINN simulated data. Figure 4c shows a comparison between traces of the data (real and PINN recovered), extracted at $x = 6$ mm from Figure 4a and 4b. This shows that traces recovered from the PINN simulated data are in a very good agreement with the real data. Figure 4d provides a plot of the recovered wave speed and the number of training epochs. The sound speed converges to $2.9 \text{ mm}/\mu s$, which is basically the speed of the surface acoustic wave in an Aluminium alloy, computed from non-destructive testing and reported in [1]. In Figure 4d, a comparison between the speeds obtained for two runs of the PINNs with different initialization is also presented. The standard deviation and mean of the speed for both runs are 4.57 and 3.55, respectively. This ensures the predictability of results obtained from the PINNs. A comparison of loss functions computed with fix and variable a for two different runs with different initialization is shown in Figure 4e. Figure 4e clearly shows that usage of adaptive activation function accelerates the convergence for both the runs. The tuples of standard deviation and mean of the loss value for both runs are (50.17, 11.49) and (49.95, 11.37), respectively and thus ensuring a very good certainty in reproducibility of the results. Figure 4f shows a plot of a against the number of epochs for two runs corresponding to different initializations. The plots indicate a similar convergence trend for both runs with standard deviation and mean of Na being (0.53, 6.2) and (0.53, 6.16). We perform the modeling of the data incorporating the back-scattering coming off the crack. The input to the network includes 40 snapshots and again uses only 40% of the total points in each snapshot. A fully connected network with width of 96 neurons and depth of 4 layers is used along with an initial learning rate of $5e - 4$. Figure 5a and 5b show snapshots from actual and PINN simulated data, respectively. To show the accuracy of the PINN predicted results, a trace along $x = 6$ mm is extracted from actual (Figure 5a) and PINN predicted model (Figure 5b) and plots of these traces are shown in Figure 5c. The prominent mode, represented by maximum amplitude in Figure 5c, is basically back-scattered events coming off the crack. This mode takes maximum computation time to get reconstructed. Figure 5d provides a plot of the wave speeds, recovered as a global variable, against number of training epochs. Similar to the case shown in Figure 4, the speed of the sound converges to $2.9 \text{ mm}/\mu s$, whereas the sensitivity of the speed of the crack is expressed by a decrease in the speed, shown in the plot Figure 5d. This phenomena motivates the training of the model to recover the speed dependence on space i.e $v(x, y)$ in 1. Figure 5e represents a comparison of loss function computed with fix and variable a , which shows that implementation of adaptive activation function helps in accelerating the convergence. Figure 5f show a plot of a against number of epochs.

4.2 Wave speed $v(x, y)$ in the entire domain for the 0° angle dataset

Now, we train the PINN to recover $v(x, y)$ to detect the location the crack in the aluminum substrate. To achieve this we used two neural networks, the first one to approximate the speed of wave and then this speed is fed into equation (1) while computing the loss function defined in (5) for PINN. The input to the PINN includes 40 snapshots, and again we use only 20% of the total points in each snapshot. The PINN and velocity networks are fully connected networks with width of 64 neurons and depth of 4 layers, and an initial learning rate of $5e - 4$. Figures 6a and 6b represent snapshots at $t = 12.38 \mu s$ obtained from real data and PINN model. A video (movie_ds1.mp4) showing the real and PINN simulated data against the time stamp is uploaded as an additional material. A comparison of traces extracted from Figures 6a

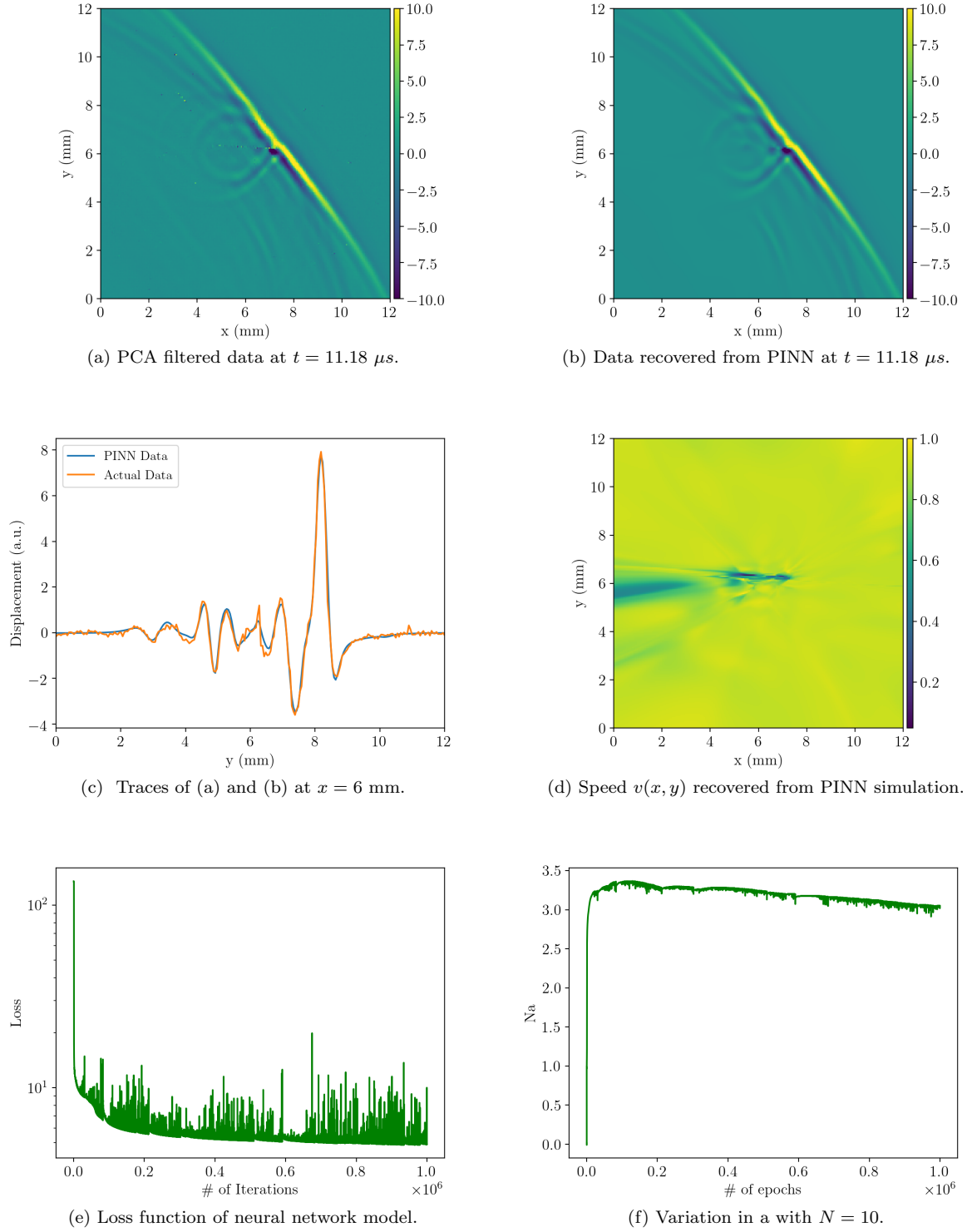
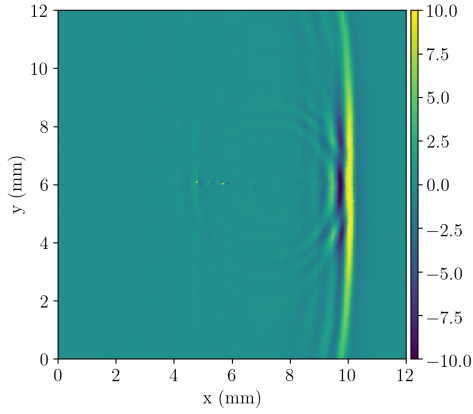
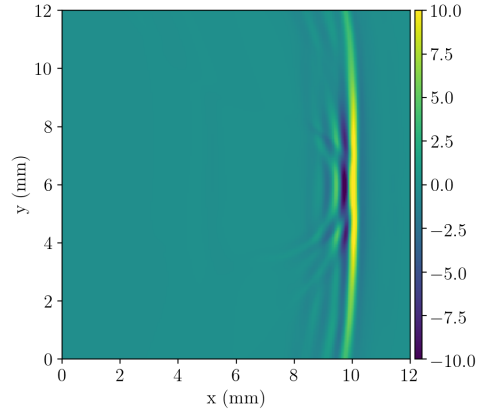


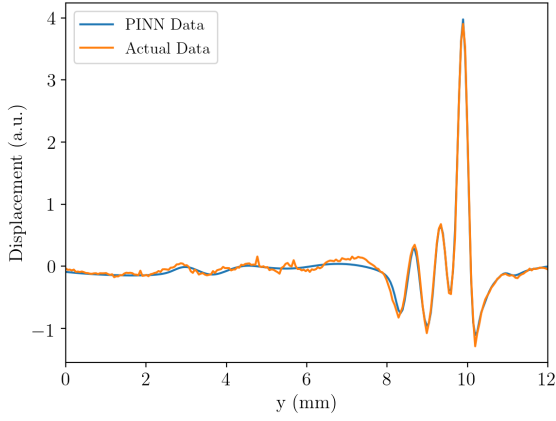
Fig. 7: Results from training of neural network model for data acquired at incidence angle of 45° , where (a) and (b) represents the snapshots of the particle displacement from actual and PINN simulated data, respectively. (c) shows a comparison of traces obtained from actual and PINN simulated data and (d) represents the normalized speed of the wave as a function of space, where crack is reflected as low speed zone. (e) shows the loss function corresponding to neural network model and (f) shows variation in a with $N = 10$.



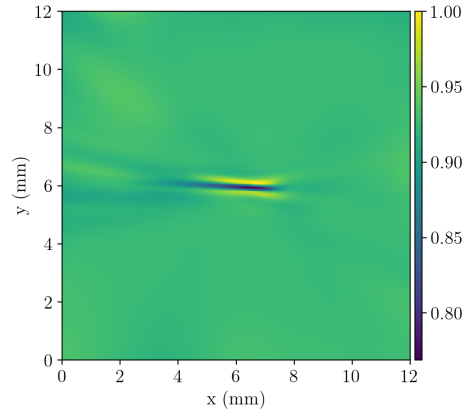
(a) PCA filtered data at $t = 11.08 \mu s$.



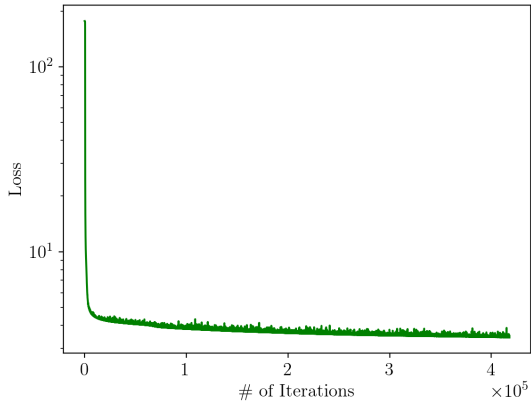
(b) Data recovered from PINN at $t = 11.08 \mu s$.



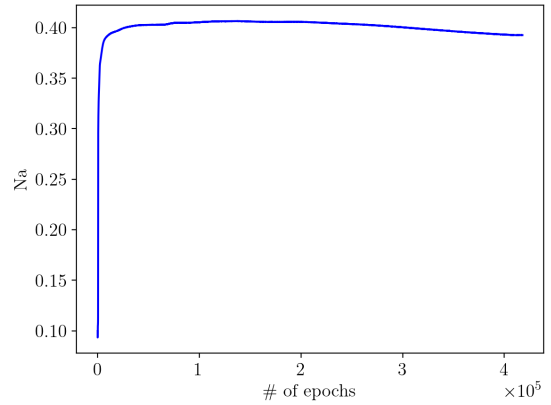
(c) Traces of (a) and (b) at $y = 11.25 \text{ mm}$.



(d) Speed, $v(x, y)$ discovered from PINN simulation.



(e) Loss function of neural network model.



(f) Variation in a with $N = 10$.

Fig. 8: Results from training of the neural network model for data acquired at incidence angle of 90° , where (a) and (b) represents the snapshot of the particle displacement from actual and PINN recovered data, respectively. (c) shows a comparison of traces obtained from actual and PINN simulated data and (d) represents the normalized speed of the wave as a function of space, where crack is reflected as low speed zone. (e) shows the loss function corresponding to neural network model and (f) shows variation in a with $N = 10$.

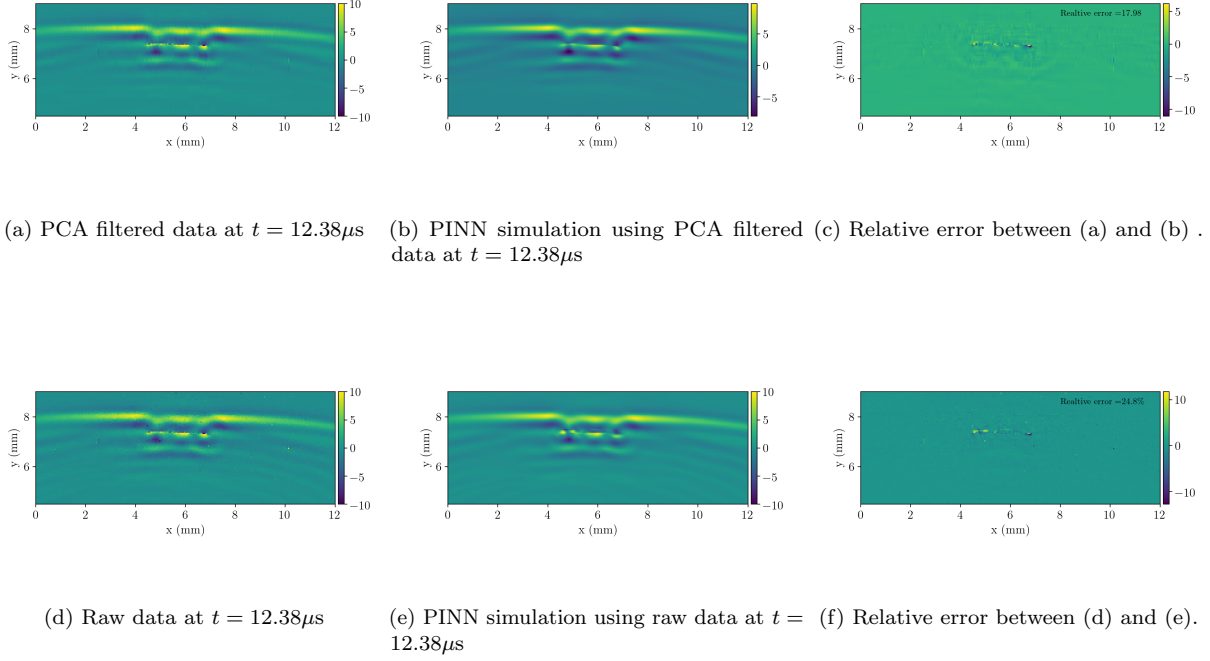


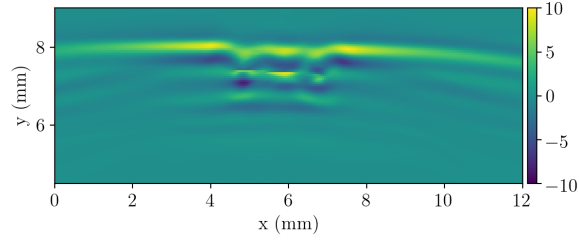
Fig. 9: Application of PCA on the data where (a) and (b) PCA filtered data and PINN simulated data using filtered data, respectively. (c) represents the error between (a) and (b) with a relative error of 17.98%. (d) and (e) represent raw and PINN simulated data using raw data, respectively. (f) represents the error between (d) and (e) with a relative error of 24.80%.

and 6b at $x = 6$ mm is shown in Figure 6c. Figure 6d shows the variation of the normalized speed of the wave $v(x, y)$. The crack is characterized by the zone of low speed with a decrease of 17 % from the background. Figure 6e represents the loss function showing that the error reduces to 1.2 % even with the 20% data, which also includes the effect of crack. Figure 6f represents the variation of a against number of epochs during the training process.

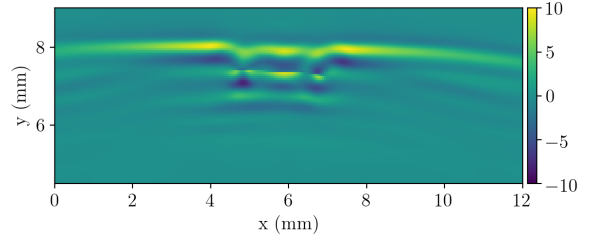
4.3 Wave speed $v(x, y)$ in the entire domain for the 45° angle dataset

Next, we train the PINN to recover $v(x, y)$ to detect the location of crack from the data acquired at incidence angle of 45° . We consider 80 snapshots and each snapshot is sampled with 10% of total points from entire domain i.e (12 mm \times 12 mm). The PINN and velocity network are a fully connected network with width of 32 neurons and depth of 6 layers, and an initial learning rate of $5e - 4$. Figures 7a and 7b represent the snapshots of the wavefields at $t = 11.18 \mu\text{s}$ extracted from real data and PINN simulated data. A video (movie_ds2.mp4) showing the real and PINN simulated data against the time stamp is uploaded as an additional material. Figure 7c shows a comparison between the traces of data extracted at $x = 6$ mm from real (Figure 7a) and PINN simulated data (Figure 7b), which shows a very good agreement. Figure 7d shows spatial variation of the normalized wave speed in the entire domain with a zone of the low speed represented by 17 % decrease in the speed from the background. This low speed zone indicates the presence of the crack. Figure 7e represents the loss function showing that the relative error reduces to 1 % even with the 20% of the data. Figure 7f represents the variation of a against number of epochs during the training process.

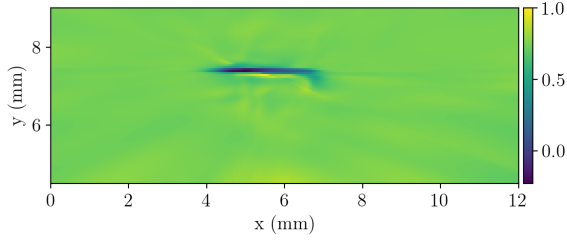
4.4 Wave speed $v(x, y)$ in the entire domain for the 90° angle dataset



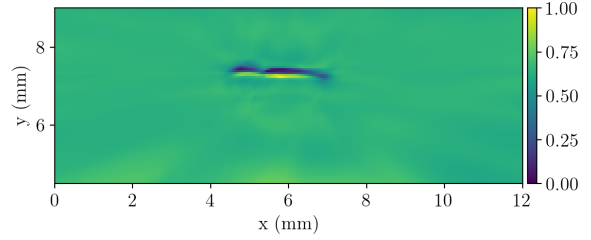
(a) Data discovered from PINN simulation using 10 % data.



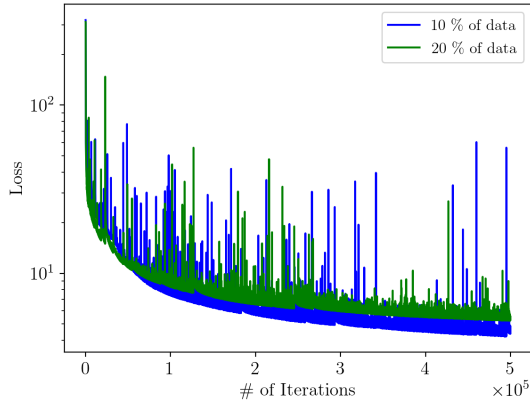
(b) Data discovered from PINN simulation using 20 % data.



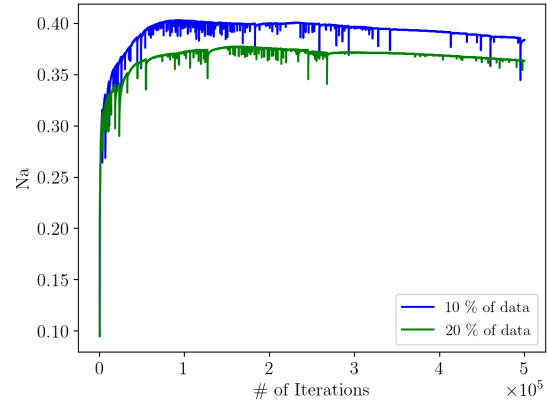
(c) $v(x, y)$ discovered from PINN simulation using 10 % of data.



(d) $v(x, y)$ discovered from PINN simulation using 20 % of data.



(e) Loss function of neural network model.



(f) Variation in a with $n = 10$.

Fig. 10: Results from training of PINN model for subsampled data acquired at incidence angle of 0° , where (a) and (b) represents the snapshot of the particle displacement from using 10 % and 20% of data, respectively. (c) and (d) represents the speed of wave discovered from PINN simulation using 10% and 20% of total data, respectively. (e) shows the comparison of loss function for data usage of 10% and 20%. (f) shows variation in a with $n = 10$ for the both sampling cases.

Finally, we train the PINN using the data acquired at 90° of incidence angle and results are presented in Figure 8. The PINN and velocity network are a fully connected network with width of 32 neurons and depth of 4 layers with an initial learning rate of $5e-4$ are used. The input data to the network is comprised of 120 snapshots and each snapshot is sampled with 20% of total points from entire domain i.e ($12\text{ mm} \times 12\text{ mm}$). The sub-figures of Figure 8 are subject to the same interpretation as those in Figure 7. A video (movie_ds3.mp4) showing the real and PINN simulated data against the time stamp is uploaded as an additional material. The location of crack is clearly represented by the streak of low speed in Figure 8d. This low speed zone is represented by a decrease of 82 % in the speed from the background, which is not as strong as the cases for 0° and 45° . This could be due to the fact that effects of back-scattering for 90° incidence angle are not as strong as those for other sets of data acquired for 0° and 45° .

4.5 Hyperparametr of PINN and effect of PCA on the convergence

Table 1 summarizes the hyperparameters used for all computational experiments, shown in Figure 4-8. To demonstrate the efficiency of PCA, we represent a PCA filtered data, PINN simulated data using filtered data and relative error between them in Figure 9a, b, and c, respectively. The relative error between filtered and PINN simulated data is 17.88%. Subsequently, we also recovered the PINN simulated data using raw (non-filtered) data. Figures 9d, 9e and 9f represent the raw, PINN simulated data and error between raw and PINN simulated, respectively. The relative error using raw data is 24.80% which is greater than error computed from the PCA filtered data. Thus, PCA helps here in achieving the greater accuracy.

4.6 Wavespeed $v(x, y)$ with submapped data

To test the requirement of minimum data size used for training of PINN, we further sub-sampled the data and used only 10% of the data from each snapshot and trained the PINN. The network architecture and hyper-parameters used are the same as those used to generate Figure 6. The results from the sub-sampling process are shown in Figure 10. Figures 10a and 10b represent the snapshots of wavefield recovered from PINN model using 10% and 20% of data, respectively. This shows a very good similarity. The speed $v(x, y)$ from 10% and 20% data is shown in Figure 10c and d, respectively. The crack in both images can be seen very convincingly. The loss functions from both data sets are plotted in Figure 10e. Figure 10e shows that the convergence with 10% and 20% of the data is very similar. The variation of a against the epoch is shown in the Figure 10f.

4.7 PINN results and dependence on initialization and training data volume

Next, we address the stochastic component involved in the training process of the PINN, which will also ensure the reproducibility of the results presented in our study. PINN optimizes the loss function by constraining it with the PDE governing the underlying physics, which results into reduction of the size of space of the admissible solutions. To prove this in the present study, we first train the PINNs with different initializations and then we vary the fraction of the input data used for the training of the PINNs. In Figure 11, we present a comparison of the loss functions and variable a of the adaptive activation functions for the PINNs trained with five different initializations. The architecture of PINNs is the same as the one used for Figure 6. Figure 11a shows the plot of loss functions for five different runs along with the corresponding standard deviation and mean. The standard deviation and mean of the losses varies from 5.62-6.73 and 7.84-8.47, respectively. The small variations in standard deviation and mean of losses ensures reproducibility of the results. Figure 11b shows the variation of variable a corresponding to the loss functions shown in Figure 11a, which also shows a very small range of variation of mean and standard deviation of a .

Next, we train the PINNs by feeding the different fraction of total input data to ensure the accuracy and predictability of wave speed $v(x, y)$. Figures 12a, 12b, 12c, and 12d represent the wavefield simulated by PINN with 10%, 20%, 30% and 40% of the total input data acquired at 0° incidence angle. Figures 13a, 13b, 13c, and 13d represent the wave speed $v(x, y)$ discovered by PINN while using 10%, 20%, 30%

Table 1: Hyper-parameters of neural network

Hyperparameters used in this study							
Reference	Depth	Width	# of epochs ($\times 10^5$)	Activation function	Learning rate	% of data used	λ
Figure 4	2	32	2	sin	$5e-4$	20%	100
Figure 5	4	96	15	sin	$5e-4$	40%	100
Figure 6	4	64	5	tanh	$5e-4$	20%	100
Figure 7	6	32	10	tanh	$5e-4$	10%	100
Figure 8	4	32	20	tanh	$5e-4$	20%	100

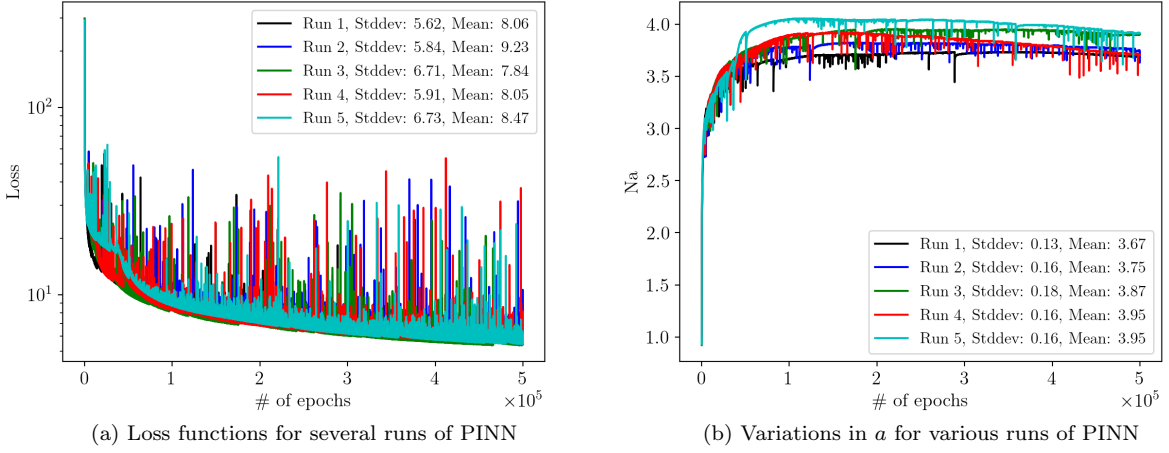


Fig. 11: Predictability of results obtained from PINNs validated by various runs with different initialization of neural network. Results are obtained for data acquired at incidence angle of 0° with same architecture as in the case for the Figure 6, where (a) and (b) represents the loss functions and variation in a .

and 40% of the total input data. Assuming the results based on the usage of 40% of the data as expected outputs, then the L2-norm of the errors corresponding to PINN simulated wavefields for 10%, 20% and 30% of the data are 12.52% (Figure 12a), 11.37% (Figure 12b) and 8.0% (Figure 12c), respectively. However, the L2-norm of the errors corresponding to wave speed recovered by PINN for 10%, 20% and 30% are 3.9% (Figure 13a), 3.8% (Figure 13b) and 2.4% (Figure 13c). The small variation of error in wavefield and wave speed clearly depicts the efficiency of PINNs for the sparse datasets.

Figure 14a presents the loss functions corresponding to the results shown in Figure 12a-d and Figure 13a-d. Figure 14a clearly shows that by increasing the fraction of the data the convergence is faster. The variation in the variable a of the adaptive activation function corresponding to losses shown in 12a is presented in 12b. Figure 12b shows that a converged to higher values as the data volume decreases.

5 Summary

In this study, we delineated the crack based on the zone marked by low sound speed. Though the data used in this study was acquired at a high frequency of 5MHz but as reported by Blackshire et al. [6] the effect of dispersion resulting into frequency dependent velocity and attenuation will not be a concern for or the aluminum alloy material system. That could be the case for much higher frequencies. The decrease in the sound speed in the crack zone could be attributed to the near-field crack feature interactions of wave field with the crack. The acoustic wave equation used in this study is primarily used to model the wave propagation in time domain and to study the effect of near-field and far-field interaction a frequency domain representation of acoustic wave equation will be more useful.

In the current paper, we hypothesize that the speed of waves, propagating on the surface of a material, could be key indicator for crack identification. To prove our hypothesis, we designed and tested a deep neural network informed by the physics of acoustic wave propagation. The PINN proposed in this study

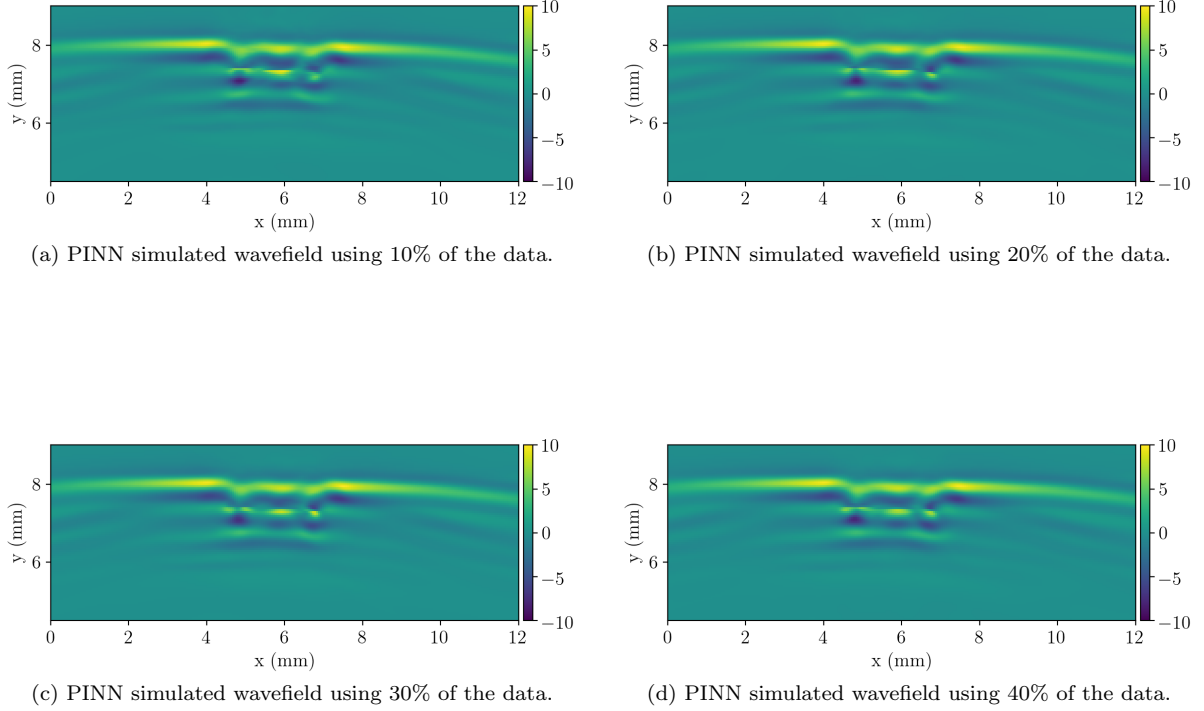
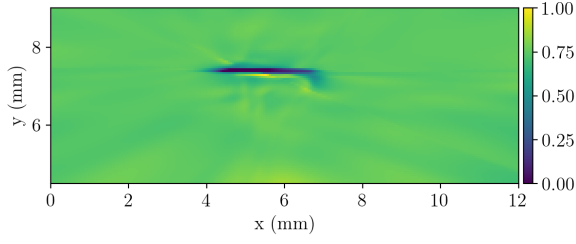


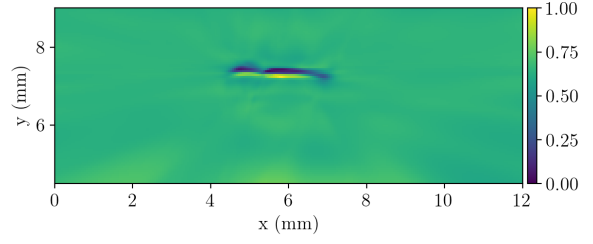
Fig. 12: PINN simulated wavefield while using (a) 10%, (b) 20%, (c) 30% and (d) 40% of the data.

is designed for inferring the system by discovering the data driven PDE. This is achieved by computing the space dependent speed of the wave in the metal plate. The data used in this study was acquired at incidence angles of 0° , 45° and 90° . The results clearly show that the presence of the crack generate lower effective speeds in the crack affected zone. This lower speed is due to the back-scattering of waves from the crack, which eventually result in a loss of wave energy. The PINN model accurately predicts the speed of the wave in metal plate, not affected by crack, as $2.9 \text{ mm}/\mu\text{s}$ which is corroborated by the data base of material properties estimated from non-destructive testing. The results shown in the present study used only 10-20% of the total data, which provided an accuracy of 1.2-1% error. In order to reduce the cost of acquiring the data, we also have shown the results from further sub-sampling by reducing the size of input data from 20% to 10%.

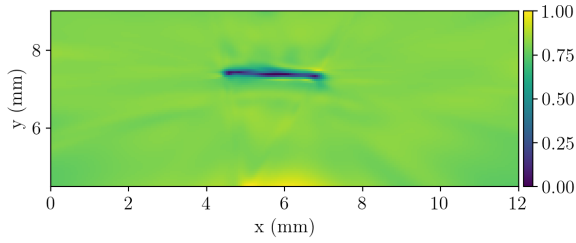
The ultrasound nondestructive quantification and analysis for various material science problems is very important. The PINN model could be very helpful to solve such problems, as acquiring the real data for such analysis is prohibited by cost and logistical issues. To support our claim, we introduced an optimized PINN model to characterize a surface breaking crack in an aluminum alloy substrate material by using only the 20% of the given input data. This also proves that full wavefield data can be recovered over the entire domain just by using a small portion of the data. In contrary to the visualization based study which requires a dense coverage of the data, PINN will only require a small amount of data points, which minimizes substantially the cost and time of the data acquisition process. The main contributions of this work is two fold: First, characterizing the presence of crack in terms of its location, and extent and second, minimizing the amount of the wave field data required to carry out the study. The wavefield data is acquired in time and space domain ($t - x$) simultaneously and it requires a very fine time and space interval to avoid temporal and spatial aliasing caused by ultrasonic frequency. With a trained PINN model, we interpolate the wave field in time and space domain by using a few data points only and moreover we compute the wave speed in the material.



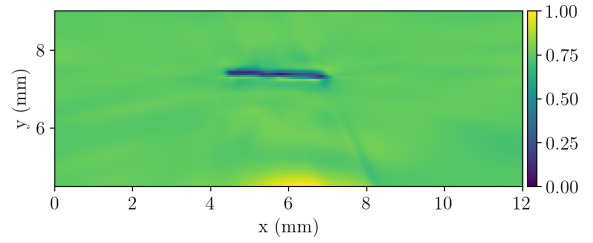
(a) PINN discovered $v(x, y)$ by using 10% of the data.



(b) PINN discovered $v(x, y)$ by using 20% of the data.

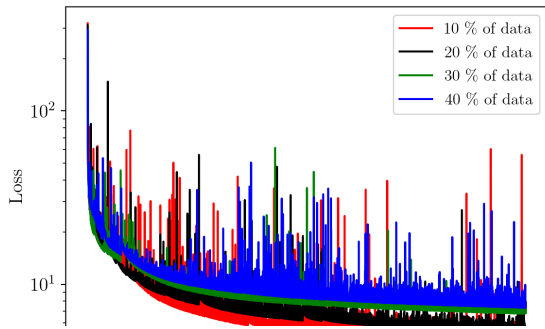


(c) PINN discovered $v(x, y)$ by using 30% of the data.

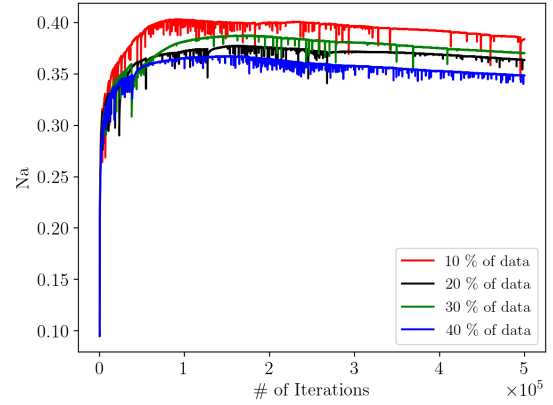


(d) PINN discovered $v(x, y)$ by using 40% of the data.

Fig. 13: $v(x, y)$ computed by PINN by using (a) 10%, (b) 20%, (c) 30% and (d) 40% of the total data size.



(a) Loss for different sampling order of the data.



(b) Variation in a for different sampling order of the data

Fig. 14: Variation of (a) loss and (b) a for different sampling order of the data i.e 10 %, 20 % 30 % and 40 %.

Acknowledgements The work is supported by DARPA-AIRA grant HR00111990025. This research was conducted using computational resources and services at the Center for Computation and Visualization, Brown University. KS would like to acknowledge Dr. Helen Kershaw from CCV, Brown University for providing the help and feedback at various stages of the research especially for problems concerning to the computation.

Supplementary materials

Supplementary materials (movie_ds1.mp4, movie_ds2.mp4, and movie_ds3.mp4,) for this paper is available at https://github.com/rajexplo/videos_NDT.

Conflict of interest

The authors declare that they have no conflict of interest.

References

1. Acoustic properties for metals in solid: From NDT Resource Center. https://www.nde-ed.org/GeneralResources/MaterialProperties/UT/ut_matlprop_metals.htm. Accessed: 2019-11-20
2. Alipanahi, B., Delong, A., Weirauch, M.T., Frey, B.J.: Predicting the sequence specificities of DNA-and RNA-binding proteins by deep learning. *Nature Biotechnology* **33**(8), 831 (2015)
3. Baydin, A.G., Pearlmutter, B.A., Radul, A.A., Siskind, J.M.: Automatic differentiation in machine learning: a survey. *Journal of Machine Learning Research* **18**(153) (2018)
4. Blackshire, J.L.: Ultrasonic scattering from complex crack morphology features. In: 2012 IEEE International Ultrasonics Symposium, pp. 248–251. IEEE (2012)
5. Blackshire, J.L.: Enhanced damage characterization using wavefield imaging methods. In: AIP Conference Proceedings, vol. 1806, p. 090008. AIP Publishing (2017)
6. Blackshire, J.L., Sathish, S.: Near-field ultrasonic scattering from surface-breaking cracks. *Applied Physics Letters* **80**(18), 3442–3444 (2002)
7. Carcione, J.M.: Wave fields in real media: Wave propagation in anisotropic, anelastic, porous and electromagnetic media, vol. 38. Elsevier (2007)
8. Dawson, A.J., Michaels, J.E., Michaels, T.E.: Isolation of ultrasonic scattering by wavefield baseline subtraction. *Mechanical Systems and Signal Processing* **70**, 891–903 (2016)
9. Goyal, P., Dollár, P., Girshick, R., Noordhuis, P., Wesolowski, L., Kyrola, A., Tulloch, A., Jia, Y., He, K.: Accurate, large minibatch sgd: Training imagenet in 1 hour. *arXiv preprint arXiv:1706.02677* (2017)
10. Jagtap, A.D., Kawaguchi, K., Karniadakis, G.E.: Adaptive activation functions accelerate convergence in deep and physics-informed neural networks. *Journal of Computational Physics* **404**, 109136 (2020)
11. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014)
12. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: *Advances in neural information processing systems*, pp. 1097–1105 (2012)
13. Lake, B.M., Salakhutdinov, R., Tenenbaum, J.B.: Human-level concept learning through probabilistic program induction. *Science* **350**(6266), 1332–1338 (2015)
14. Nielsen, M., Lund, O.: Nn-align. an artificial neural network-based alignment algorithm for MHC class II peptide binding prediction. *BMC Bioinformatics* **10**(1), 296 (2009)
15. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E.: Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research* **12**, 2825–2830 (2011)
16. Pohl, J., Mook, G.: Laser-vibrometric analysis of propagation and interaction of Lamb waves in cfrp-plates. *CEAS Aeronautical Journal* **4**(1), 77–85 (2013)
17. Raissi, M., Karniadakis, G.E.: Hidden physics models: Machine learning of nonlinear partial differential equations. *Journal of Computational Physics* **357**, 125–141 (2018)
18. Raissi, M., Perdikaris, P., Karniadakis, G.E.: Machine learning of linear differential equations using gaussian processes. *Journal of Computational Physics* **348**, 683–693 (2017)
19. 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. *Journal of Computational Physics* **378**, 686–707 (2019)
20. Raissi, M., Wang, Z., Triantafyllou, M.S., Karniadakis, G.E.: Deep learning of vortex-induced vibrations. *Journal of Fluid Mechanics* **861**, 119–137 (2019)
21. Rudy, S.H., Brunton, S.L., Proctor, J.L., Kutz, J.N.: Data-driven discovery of partial differential equations. *Science Advances* **3**(4), e1602614 (2017)
22. Tian, Z., Yu, L.: Lamb wave frequency-wavenumber analysis and decomposition. *Journal of Intelligent Material Systems and Structures* **25**(9), 1107–1123 (2014)
23. Wang, Y.E., Wei, G.Y., Brooks, D.: Benchmarking TPU, GPU, and CPU platforms for deep learning. *arXiv preprint arXiv:1907.10701* (2019)

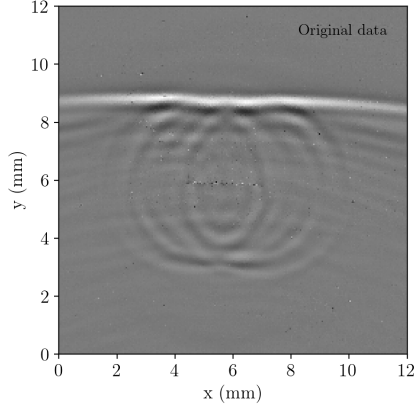
A Python code snippet of PCA process

Listing 1: PCA of wave field data

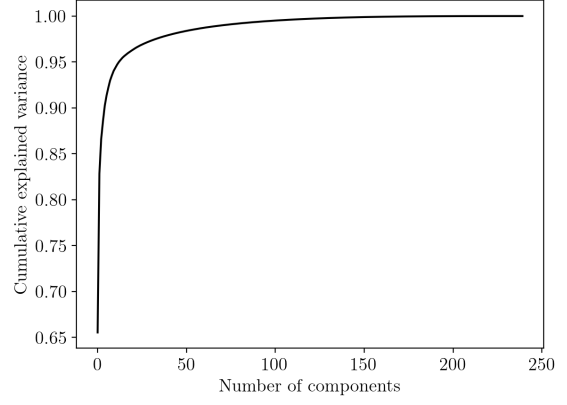
```
import numpy as np
import scipy
from sklearn.decomposition import PCA
# Routine for choosing the number of components
def nComponent(d, maxTh):
    wvf=scipy.io.loadmat('wvf.mat')['wvf']
    pca=PCA(n_components=240)
    pca.fit(d)
    variance=np.cumsum(pca.explained_variance_ratio_)
    nComp=0
    for i in range(0, len(variance)):
        if (variance[i]-0.95*max(variance)) < eps:
            nComp +=1
    return nComp

# Routine for filtering the data and constructing the filtered data
def pcaWd(d, nComponent):
    nComp=nComponent(d, maxTh)
    pca = PCA(n_components=nComp)
    d1_pca = pca.fit_transform(d1)
    inv_pca = pca.inverse_transform(d1_pca)
    return inv_pca
```

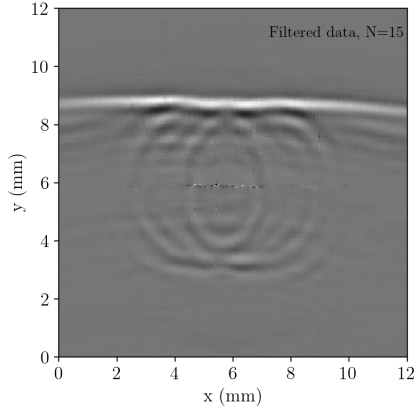
B PCA filtered data acquired at 0° and 90°



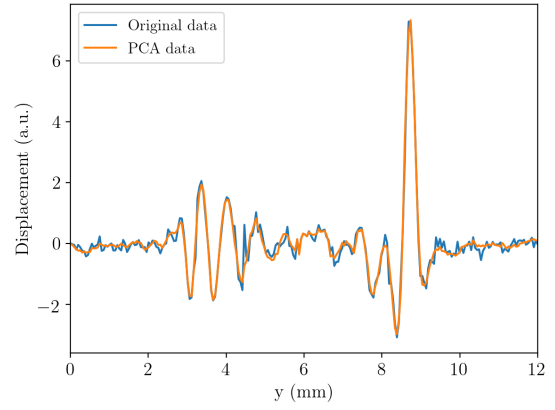
(a) Actual data at $t = 13.2 \mu s$.



(b) Cumulative explained variance of (a).

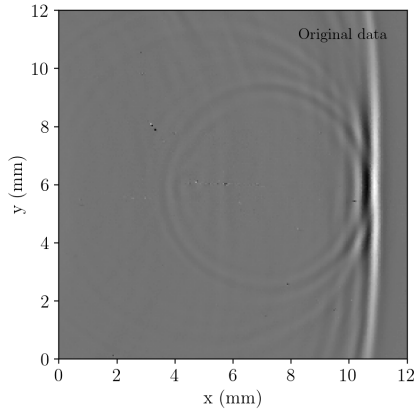


(c) Filtered data at $t = 13.2 \mu s$.

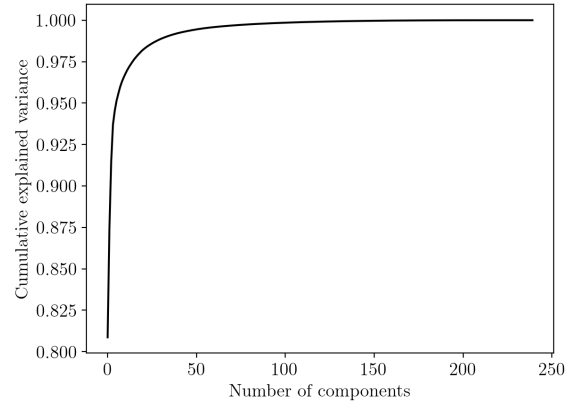


(d) Original vs filtered trace at $x = 7 \text{ mm}$ at $t = 13.2 \mu s$.

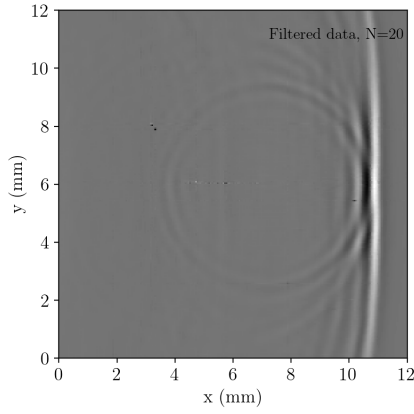
Fig. B1: Filtering of wavefield data acquired at 0° , using principle component analysis (PCA). PCA requires only the first 15 components to construct the original data by zeroing out the insignificant components.



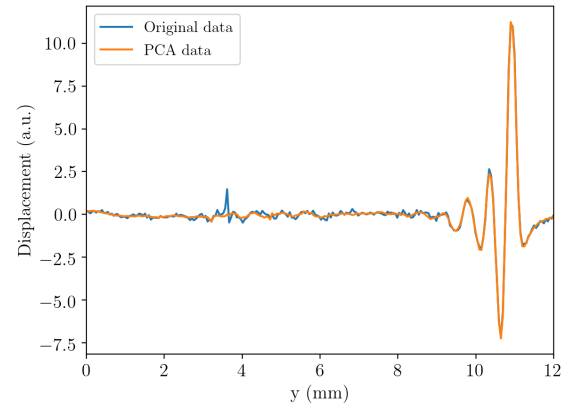
(a) Actual data at $t = 11.38 \mu s$.



(b) Cumulative explained variance of (a).



(c) Filtered data using at $t = 11.38 \mu s$



(d) Original vs filtered trace at $x = 4.5 \text{ mm}$ at $t = 11.38 \mu s$.

Fig. B2: Filtering of the wavefield data acquired at 90° , using principle component analysis (PCA). PCA requires only first 20 components to construct the original data by zeroing out the insignificant components.