DATA DRIVEN SOLUTIONS AND DISCOVERIES IN MECHANICS
USING PHYSICS INFORMED NEURAL NETWORK *

QI ZHANG†, YILIN CHEN †, AND ZIYI YANG‡

Abstract. Deep learning has achieved remarkable success in diverse computer science applications, however, its use in other traditional engineering fields has emerged only recently. In this project, we solved several mechanics problems governed by differential equations, using physics informed neural networks (PINN). The PINN embeds the differential equations into the loss of the neural network using automatic differentiation. We present our developments in the context of solving two main classes problems: data driven solutions and data driven discoveries, and we compare the results with either analytical solutions or numerical solutions using finite element method. The remarkable achievements of PINN model shown in this report suggest the bright prospect of the physics-informed surrogate models that are fully differentiable with respect to all input coordinates and free parameters. More broadly, this study shows that PINN provides an attractive alternative to solve traditional engineering problems.

Key words. Conservation laws, Data inference, Data discovery, Dimensionless form, PINN

1. Introduction. It is well-known that many mechanics phenomenons are governed by conservation laws which can be converted to differential equations for most of the applications. Solving these differential equations and obtaining parameter estimations from given observations are always intriguing topics, and significant research has been done to develop many advanced (semi-)analytical or numerical algorithms. While in the last decade, machine learning especially neural networks have yielded revolutionary results across diverse disciplines, including image and pattern recognition, natural language processing, genomics, and material constitutive modeling [15, 23, 26, 34], among which a fair amount of research has also been done related to differential equations [1, 9, 12, 13, 17–19, 22, 24, 25, 27–33, 35, 36, 38], owing to the dramatic increase in the computing resources. Therefore, it will be interesting to adopt neural networks as an important alternative to traditional mathematical methods to approximate the solutions to differential equations through iterative update of the network weights and biases.

2. Related work. Early researches that use neural network in scientific computing could date back to [12, 22], in which a single hidden layer neural network is adopted

\[ N = \tilde{u} + \sum_{i=1}^{H} v_{i} \sigma_{i} \left( \sum_{j=1}^{n} w_{ij} x_{j} + u_{i} \right), \]

where \( N \) is the network output which is assumed to be a scalar for illustration purposes, \( H \) is the width of the hidden layer, and \( n \) is the dimension of the input vector \( x \in \mathbb{R}^{n} \), \( w_{ij} \) and \( v_{i} \) are called weights while \( u_{i} \) and \( \tilde{u} \) are always denoted as the biases, \( \sigma_{i}(\cdot) \) is called activation function. More recently, the development of automatic differentiation [2] has led to emerging literature on the use of deep neural networks to solve differential equations [17, 33, 38, 39] even in complex geometries, and Berg

*Stanford CS229 project final report, category of physical sciences
†Department of Civil and Environmental Engineering, Stanford University, CA 94305 (qzhang94@stanford.edu, yilinc2@stanford.edu).
‡Department of Mechanical Engineering, Stanford University, CA 94305 (zy99@stanford.edu).
et al. [3] provides details of this back propagation algorithm for advection and diffusion equations. Unlike traditional machine learning methods, deep neural networks sometimes can overcome the curse of dimensionality [17]. In the work done by Raissi et al. [30–32], they named such strong form approach for differential equation as the physics-informed neural network (PINN) for the first time. Today, the PINN becomes more and more popular in many engineering fields such as hydrogeology [7, 18, 35], geomechanics [19], cardiovascular system [21] and so on. One attractive feature of PINN is that it is fully differentiable with respect to all the input coordinates and free parameters, in other words, the trial solution (via the trained PINN) represents a smooth approximation that can be evaluated and differentiated continuously on the domain [9]. Another important feature of PINN is that it can be used to solve inverse (data discovery) problems [16] with minimum change of the code from the forward problems [24].

It is also worth to mention that there are also other machine learning methods that didn’t rely on the neural network to solve differential equations [28, 29], but in this report, we will focus on the PINN.

3. Methods. In this section, we will give a very brief overview of PINN in order to demonstrate how the loss function $\mathcal{L}$ is defined in such context. We refer interested readers to look at [32] for more details. Consider a system of differential equations defined on the domain $\Omega$ with boundary $\partial\Omega$:

$$
\mathcal{D}(u(x)) = 0 \quad x \in \Omega,
$$

$$
\mathcal{B}(u(x)) = 0 \quad x \in \partial\Omega,
$$

where $u$ is the unknown solution vector, $\mathcal{D}$ denotes the abstract (non-linear) differential operator (e.g., $\partial/\partial x$, $u \circ \partial/\partial x$, $x \circ \partial/\partial x$, higher order derivatives, non-linear functions of $u$ or $x$, their combinations, etc.), and the operator $\mathcal{B}$ expresses arbitrary boundary conditions (e.g., Dirichlet, Neumann, Robin, etc.) associated with the problem. Note here for time-dependent problems, we consider $t$ as a special component of $x$, i.e., the $\Omega$ contains the temporal domain. The initial condition can be simply treated as a special type of Dirichlet boundary condition on the spatio-temporal domain [24].

Now we construct a neural network with $L$ layers, or $L - 1$ hidden layers to approximate $u(x)$, the input to neural network is $x$, and we will denote the output of the neural network as $\hat{u}(x; \theta)$ (we want $\hat{u}(x; \theta) \approx u(x)$), where $\theta$ represents the collection of weight matrix $W[l] \in \mathbb{R}^{n_l \times n_{l-1}}$ and bias vector $b[l] \in \mathbb{R}^{n_l}$ for each layer $l$ with $n_l$ neurons. In the inverse or data discovery problems, there are some unknown parameters $\gamma$ in the original differential equations, thus we will rewrite our approximation as $\hat{u}(x; \theta, \gamma)$. Now we can define the physics-informed loss $\mathcal{L}$ as

$$
\mathcal{L} (\theta, \gamma) = w_f \mathcal{L}_f (\theta, \gamma) + w_b \mathcal{L}_b (\theta, \gamma) + w_d \mathcal{L}_d (\theta, \gamma),
$$

where $w_f$ is the weight for the loss due to mismatch with the governing equations

$$
\mathcal{L}_f (\theta, \gamma) = \frac{1}{|\Gamma_f|} \sum_{x \in \Gamma_f \subset \Omega} \| \mathcal{D}(\hat{u}(x; \theta, \gamma)) \|_2^2,
$$

and $w_b$ is the weight for the loss due to mismatch with the boundary conditions

$$
\mathcal{L}_b (\theta, \gamma) = \frac{1}{|\Gamma_b|} \sum_{x \in \Gamma_b \subset \partial\Omega} \| \mathcal{B}(\hat{u}(x; \theta, \gamma)) \|_2^2.
$$

This manuscript is for review purposes only.
and $w_d$ is the weight for the loss due to mismatch with the given data observations $u^*(x)$

$$L_d(\theta, \gamma) = \frac{1}{|\Gamma_d|} \sum_{x \in \Gamma_d \subset \Omega} \| \hat{u}(x; \theta, \gamma) - u^*(x) \|^2_2.$$  

In practice, $\Gamma_f$ and $\Gamma_b$ are always known as the set of locations of the “residual” points, and $\Gamma_d$ is known as the set of measurement locations. We then optimize $\theta$ and $\gamma$ together, and our solution is

$$\theta^*, \gamma^* = \arg\min_{\theta, \gamma} L(\theta, \gamma).$$

Above process is approximately summarized in Fig. 1. Due to the network architecture and incorporation of differential equations, the loss will be highly non-linear and also non-convex, thus we will need some gradient-based optimizers, such as steepest descent [4], Newton’s method [4], Adam [20], and L-BFGS (limited-memory quasi-Newton method) [5]. In this report, we will adopt purely Adam or Adam combined with L-BFGS optimizers (i.e., we first use Adam for a certain number of iterations, and then switch to L-BFGS until convergence).

**Fig. 1.** Two schematics of the physics-informed neural network (PINN) [24, 25].

For estimation of the test error, we will evaluate the loss $L_f$ on a given test set $\Gamma_{test} \subset \Omega$. In addition, if the true solution is available, we will also calculate the mean square error (similar to the form of $L_d$) on the same test set.
4. Experiments, results and discussions.

4.1. One dimensional consolidation problem. In soil mechanics, one of the most important processes is the consolidation process, while under some specific assumptions made by Prof. Terzaghi and Prof. Biot [6, 8, 37], we can mathematically quantify this process as a one dimensional diffusion problem [14] with specific boundary conditions and initial conditions. Suppose our problem domain is $0 \leq x \leq L$ where $L$ can be understood as the thickness of the aquifer that rests on a rigid impermeable base, at time $t = 0$, surface load $p_0$ is applied on the top of the aquifer which will lead to an initial pore pressure field $p^u = \gamma_l p_0$ at $t = 0^+$ where $\gamma_l$ is called loading efficiency [37]. As time goes on, the pressure $p$ will gradually decrease due to the effect of the drainage boundary at $x = 0$, and it satisfies following partial differential equation

$$\frac{\partial p}{\partial t} - c \frac{\partial^2 p}{\partial x^2} = 0,$$

with boundary conditions $p = 0$ at $x = 0$ and $\partial p/\partial x = 0$ at $x = L$. The initial condition is just $p = p^u$ as mentioned before. In above Eq. (4.1), the $c$ is called generalized consolidation coefficient [11] which has the unit of length squared divided by time.

However, the order of magnitudes of different physical quantities $x$ (order of meters), $t$ (order of months) and $p$ (order of kPa) could cast great difficulty on the training of the neural networks [21], and to overcome this problem, we employ a non-dimensionalization technique with the purpose of scaling the input and the output of the neural network in proper scales (e.g., $O(1)$). As a result, we will introduce characteristic values for $x$ and $p$, and for this problem an obvious choice would be characteristic length $L$ and characteristic pressure $p^u$. At this point we define the dimensionless quantities as

$$\hat{x} = \frac{x}{L}, \quad \hat{p} = \frac{p}{p^u}.$$

For the time $t$, the dimensionless time $\hat{t}$ is given as

$$\hat{t} = \frac{c t}{L^2}.$$

Now all the variables take values with $O(1)$ magnitude and we will rewrite Eq. (4.1) using dimensionless variables, the result is shown as follows

$$\frac{\partial \hat{p}}{\partial \hat{t}} - \frac{\partial^2 \hat{p}}{\partial \hat{x}^2} = 0 \quad 0 < \hat{x} < 1 \quad \hat{t} > 0$$

$$\hat{p} = 0 \quad \hat{x} = 0 \quad \hat{t} \geq 0$$

$$\frac{\partial \hat{p}}{\partial \hat{x}} = 0 \quad \hat{x} = 1 \quad \hat{t} > 0$$

$$\hat{p} = 1 \quad 0 < \hat{x} \leq 1 \quad \hat{t} = 0.$$

The analytical solution for Eq. (4.4) is given as

$$\hat{p} = \sum_{m=1,3,5,7,...}^{\infty} \frac{4}{m \pi} \sin \left( \frac{m \pi \hat{x}}{2} \right) e^{-\frac{m^2 \pi^2 \hat{t}}{4}}.$$
For the PINN hyper-parameters, we assume \( \hat{t}_{\text{max}} = 0.5 \), and we draw 500, 250, 125 “residual” points randomly from \((0, 1) \times (0, \hat{t}_{\text{max}})\), boundary of \( \hat{x} \) and boundary of \( \hat{t} \) respectively to form \( \Gamma_f \) and \( \Gamma_b \). The test set \( \Gamma_{\text{test}} \) will have 10011 points. The loss weights are assigned as \( w_f = w_b = 1/4 \) (\( w_d \) is not used in this problem). For the architecture of the neural network, it will contain 5 hidden layers with 50 neurons each, and we will use hyperbolic tangent \( \tanh \) as our activation function in hidden layers, the output activation will be linear activation but we will multiply the network output with \( \hat{x} \) as our final output, because in this way our output will automatically satisfy the second expression in Eq. (4.4). For the optimization procedure, we will run Adam for 50000 epochs with learning rate 0.001 and then run L-BFGS until convergence. For the network initialization, we will use the Glorot normal initializer.

The convergence history, reference solution, PINN solution, and the absolute error is shown in the Fig. 2, from which we can see a perfect match is achieved, and the locations with highest error is close to the origin due to sharp gradient. Note in Fig. 2, we have enforced all the output values are between 0 and 1, and initial condition is strictly satisfied.

![Fig. 2. PINN training results for the one dimensional consolidation problem.](image)

4.2. Steady state heat conduction problem. Our second example is a time-independent problem on 2D. In the steady state heat transfer problem with constant heat conduction coefficient, constant heat source and constant boundary temperature (Dirichlet boundary condition), the \( T \) will satisfy Poisson’s equation [14]. Without loss of generality and to avoid the scaling differences, we will solve following PDE on
\[ \Omega = [-1, 1] \times [-1, 1] \]

\[
\frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial y^2} + 1 = 0 \quad -1 < x < 1 \quad -1 < y < 1,
\]

\[ T = 0 \quad (x, y) \in \partial \Omega. \]

Even though the analytical solution for this problem doesn’t exist, the finite element solution with high accuracy is available in deal.II finite element library tutorial 3, thus it will be interesting to compare these two numerical solutions.

For the PINN hyper-parameters, we choose 1500 and 500 random points to form \( \Gamma_f \) and \( \Gamma_b \), respectively. The test set \( \Gamma_{\text{test}} \) will have 10000 points. The loss weights are assigned as \( w_f = w_b = 1/2 \) (\( w_d \) is not used in this problem). For the architecture of the neural network, it will contain 4 hidden layers with 50 neurons each, and we will not apply any transform for the network output. The learning rate for this problem is 0.0005. For the network initialization, we will use the Glorot uniform initializer. The remaining settings are the same as the first example.

The training results are shown in the Fig. 3, where the two numerical solutions are consistent with each other. Note in Fig. 3, we have enforced all the output values are larger than 0. The largest absolute error happens near the Dirichlet boundary because Eq. (4.7) is not regarded as a hard constrain in the construction of the loss \( \mathcal{L} \). A feasible way to handle this problem would be to introduce a smooth distance function \( \hat{d}(x) \) that gives the distance for \( x \in \Omega \) to \( \partial \Omega \) [3], and then we multiply \( \hat{d} \) with the original network output to get our new improved output, which will automatically satisfy Eq. (4.7).

### 4.3. Inverse problem for simple structural dynamic system.

The last example comes from the dynamic response of the single degree of freedom system [10] and we want to use this example to illustrate the capability of PINN in the inverse modeling or data discovery process. In structural dynamics, the free vibration equation with viscous damping for single mass point is obtained through D’Alembert principle [10]

\[
m \frac{d^2 u}{dt^2} + c \frac{du}{dt} + ku = 0 \quad t > 0
\]

with \( u(0) = u_0 \) and \( \dot{u}(0) = v_0 \). In above equation, \( m > 0 \) is the mass, \( c > 0 \) is the damping coefficient, \( k > 0 \) is the stiffness, \( u \) is the displacement, \( u_0 \) is the initial displacement and \( v_0 \) is the initial velocity. The analytical solution to this ODE is given as

\[
(4.9) \quad u(t) = e^{-\xi \omega_n t} \left[ u_0 \cos (\omega_D t) + \frac{v_0 + \xi \omega_n u_0}{\omega_D} \sin (\omega_D t) \right],
\]

where \( \omega_n = \sqrt{k/m} \) is known as the natural frequency, \( \xi = c/(2m\omega_n) \) is known as the damping ratio, \( \omega_D = \omega_n \sqrt{1 - \xi^2} \). In this problem, we will assume \( \xi < 1 \) which is known as the underdamped system.

Now suppose \( m = u_0 = v_0 = 1 \), \( c \) and \( k \) are the two parameters to be identified from the observations at certain times. The observations are produced with the underlying true parameters \( c = 0.4 \) and \( k = 4 \) using Eq. (4.9). We choose 50 equispaced points to form \( \Gamma_d \), 100 random points to form \( \Gamma_f \), 50 random points to form \( \Gamma_b \) and
10000 points to form $\Gamma_{\text{test}}$. The loss weights are assigned as $w_f = w_b = w_d = \frac{1}{4}$.

For the architecture of the neural network, it will contain 3 hidden layers with 32 neurons each, and we will not apply any transform for the network output. For the optimization procedure, we will only run Adam for 100000 epochs with learning rate 0.0005. For the network initialization, we will use the Glorot uniform initializer.

The evolution trajectories of $c$ and $k$ are presented in Fig. 4, with final values of $\bar{k} = 3.9933197$ and $\bar{c} = 0.40125215$, which agree with their true values.

5. Conclusions. In this project, we have adopted PINN, a new class of universal function approximators that is capable of encoding any underlying conservation laws which can be described by differential equations. Compared to the traditional numerical methods, PINN employs automatic differentiation to handle differential operators, and thus are mesh-free. We use PINN as a data driven algorithm for inferring solutions and parameter identifications of differential equations. A series of promising results for a diverse collection of problems in mechanics are presented and discussed, which opens the path for endowing deep learning with the powerful capacity of mathematical physics to model the world around us. We hope this project could benefit future practitioners across a wide range of scientific domains who want to incorporate deep learning methods in modeling physics-related problems.

For the future work, it is worth to try a group of coupled differential equations such as Eqs. (5.1)(5.2) in poromechanics [40] to see the PINN’s performance, and develop some training points selection strategy similar to the adaptive re-meshing in
the mesh-based methods to achieve better accuracy.

\begin{equation}
\frac{1}{M} \frac{\partial p}{\partial t} - \frac{k}{\mu_f} \nabla^2 p + \alpha \frac{\partial \nabla \cdot u}{\partial t} = 0,
\end{equation}

\begin{equation}
G \nabla^2 u + \frac{G}{1 - 2\nu} \nabla (\nabla \cdot u) - \alpha \nabla p + \rho g = 0.
\end{equation}
REFERENCES


[27] M. Raissi, Deep hidden physics models: Deep learning of nonlinear partial differential equa-


